

# Recovery With Incomplete Knowledge: Fundamental Bounds on Real-Time Quantum Memories

Arshag Danageozian

Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA

The recovery of fragile quantum states from decoherence is the basis of building a quantum memory, with applications ranging from quantum communications to quantum computing. Many recovery techniques, such as quantum error correction, rely on the *a priori* knowledge of the environment noise parameters to achieve their best performance. However, such parameters are likely to drift in time in the context of implementing long-time quantum memories. This necessitates using a “spectator” system, which estimates the noise parameter in real-time, then feedforwards the outcome to the recovery protocol as a classical side-information. The memory qubits and the spectator system hence comprise the building blocks for a real-time (i.e. drift-adapting) quantum memory. In this article, I consider spectator-based (incomplete knowledge) recovery protocols as a real-time parameter estimation problem (generally with nuisance parameters present), followed by the application of the “best-guess” recovery map to the memory qubits, as informed by the estimation outcome. I present information-theoretic and metrological bounds on the performance of this protocol, quantified by the diamond distance between the “best-guess” recovery and optimal recovery outcomes, thereby identifying the cost of adaptation in real-time quantum memories. Finally, I provide fundamental bounds for multi-cycle recovery in the form of recurrence inequalities. The latter suggests that incomplete knowledge of the noise could be an advantage, as errors from various cycles can cohere. These results are illustrated for the approximate [4,1] code of the amplitude-damping channel and relations to various fields are discussed.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>5</b>
2.1	Quantum States and Channels . . . . .	5
2.2	Diamond Distance Between Quantum Channels . . . . .	5
2.3	Quantum Fidelities . . . . .	5
<b>3</b>	<b>Lower-Bounding Diamond Distance Using Entanglement Fidelity</b>	<b>6</b>
<b>4</b>	<b>Fundamental Bounds on Recovery with Incomplete Knowledge</b>	<b>8</b>
4.1	The Regime of Validity . . . . .	8
4.2	A General Framework For Spectator-Based Recovery Protocols . . . . .	9
4.3	Spectator Dynamics With No Nuisance Parameters . . . . .	10
4.3.1	Independent Noise Approximation . . . . .	10
4.3.2	Spectator Qubits As Memory Qubits With Controllable Environment Coupling . . . . .	10
4.3.3	Quantifying The Physical Choice of Spectator Systems . . . . .	11
4.4	Information-Theoretic Bounds . . . . .	11
4.4.1	Fundamental Limitations For All Recovery Protocols . . . . .	11
4.4.2	Metrological Cost of Spectator-Based Recovery Protocols . . . . .	11
4.4.3	Comparison With The Non-Adaptive Case . . . . .	13
<b>5</b>	<b>Application to The [4, 1] Code of The Amplitude-Damping Channel</b>	<b>14</b>
5.1	The Amplitude-Damping Channel . . . . .	14
5.2	The Approximate [4,1] Code . . . . .	14
5.3	Entanglement Fidelity . . . . .	15
5.4	The Effect of Nuisance Parameters . . . . .	15
<b>6</b>	<b>Recovery Bounds in The Multi-Cycle Scenario</b>	<b>16</b>
6.1	The Multi-Cycle Case . . . . .	16
6.2	Recurrence Inequalities for Composite Average Channel Fidelity . . . . .	17
6.3	Contribution of The Spectator System . . . . .	18
6.4	Application to [4,1] Code of The Amplitude-Damping Channel . . . . .	18
<b>7</b>	<b>Comparison With Previous Literature</b>	<b>18</b>
7.1	Relation to Quantum Information-Theoretic Protocols . . . . .	18

Arshag Danageozian: arshag.danageozian@gmail.com

7.2	Relation to Robustness of Channel-adapted QEC . . . . .	19
7.3	Relation to [4,1] AD Code Literature . . . . .	20
7.4	Relation to Time-Dependent QEC . . . . .	20
<b>8</b>	<b>Conclusion and Open Questions</b>	<b>21</b>
<b>9</b>	<b>Acknowledgements</b>	<b>22</b>
	<b>References</b>	<b>22</b>
<b>A</b>	<b>Lower-Bounding Generalized Distinguishability Measures Using Entanglement Fidelity</b>	<b>27</b>
A.1	Generalized Distinguishability and Distance Measures . . . . .	27
A.2	Unitary $t$ -designs . . . . .	28
A.3	Channel Twirlings . . . . .	29
A.4	Lower-Bounding Generalized Distinguishability Measures Using Entanglement Fidelity . . . . .	29
<b>B</b>	<b>Comment on The Chaining Property</b>	<b>31</b>
<b>C</b>	<b>Upper-Bounding Generalized Distance Measures for State Recovery</b>	<b>31</b>
C.1	Single-Cycle Case . . . . .	31
C.2	Multi-Cycle Case . . . . .	32
<b>D</b>	<b>Necessary and Sufficient Condition for Independence of the Reduced Channel From a Mother Channel Parameter</b>	<b>33</b>
<b>E</b>	<b>Quantum Fisher Information Matrix</b>	<b>34</b>
E.1	Useful Definitions . . . . .	34
E.2	Saturation of QCRB . . . . .	34
E.2.1	Parameter-Independent Estimator . . . . .	35
E.2.2	Maximum Likelihood Estimator . . . . .	35
E.3	Partial QFIM . . . . .	35
<b>F</b>	<b>Choi Matrix Methods for Entanglement Fidelity of Composite Parameterized Channels</b>	<b>36</b>
F.1	Zeroth Derivative: Hölder Type Upper Bounds . . . . .	36
F.2	First Derivative: Bound on Robustness of a Recovery Map . . . . .	37
F.3	Proof of Theorem 1: Bounds on Spectator-Based Recovery For Finite Estimation Errors . . . . .	37
<b>G</b>	<b>Sufficient Condition For a Negligible Remainder Term in Theorem 2</b>	<b>37</b>
<b>H</b>	<b>Entanglement Fidelity For The [4,1] Code</b>	<b>38</b>
<b>I</b>	<b><math>\chi</math>-Matrix Representation of Quantum Channels</b>	<b>39</b>
<b>J</b>	<b>Proof of Lemma 3</b>	<b>40</b>
<b>K</b>	<b>Proof of Theorem 3</b>	<b>41</b>

# 1 Introduction

Quantum memories comprise an important component of current and future quantum technologies. Their use ranges from quantum communications and networks [1–3] to sensing [4], and even computation. This wide range of relevance stems from the fact that a quantum memory preserves a quantum system’s (often fragile) state from decoherence, which encodes the desired quantum information.

Depending on the physical implementation of the quantum memory, current coherence times range from milliseconds to minutes [5]. However, various quantum technologies may require even longer coherence times [6]. Two of the most common techniques implemented in a quantum memory are quantum error correction (QEC) [7–9] and dynamical decoupling [10, 11]. These techniques generally benefit from the *a priori* knowledge of the noise surrounding the system of interest. For example, channel-adaptation techniques in QEC [12] have been shown to outperform general QEC codes, as they are given additional knowledge of the environment noise [13, 14]. Such techniques rely on some physical model of the (noisy) implementation medium of the quantum memory. The noise model is partially built upon physical assumptions (e.g. in the choice of the Hamiltonians) and partially upon phenomenology. Hence, the former gives a physically motivated family of quantum dynamics  $\{\mathcal{N}_\theta\}_{\theta \in \Theta}$  for the quantum state of the memory qubits [15], and the latter determines the value of the noise parameter  $\theta$  such that the dynamics  $\mathcal{N}_\theta$  fits the observed data the best.

Although very powerful, a shortcoming of this approach is that the environment noise parameter is generally time-varying. This has been studied most extensively for superconducting qubits [16–20]. Hence, real-time techniques to track the change (drift) of the noise are necessary, assuming we want to operate quantum memories for times larger than the characteristic times of the drift.

Indeed, efforts have been made towards designing “spectator” systems that aid in detecting and tracking such changes [21–30]. Being subject to the same physical environment, the goal of the spectator system is to perform real-time quantum sensing of the noise parameter. The estimate is then used as a classical side-information in various recovery protocols. The physical requirements of spectator systems are two-fold: (1) proximity to the memory qubits, such that the spatial dependence of the noise parameter can be neglected, and (2) exhibiting faster dynamics than the memory qubits. The latter is necessary if the feedback information is to be useful for recovery. We showcase the functionality of the spectator system within a quantum memory using Figs. 1 and 2. Note that, since the memory and spectator systems are generally different physical systems with different couplings to

the same environment, their dynamics will generally be described by different quantum channels with the same noise parameter, i.e.  $\mathcal{N}_\theta$  and  $\mathcal{M}_\theta$ , respectively. More precisely, a spectator-based recovery protocol is comprised of the following characteristic stages, as shown in Fig. 3:

1. *Individual state preparation* of the quantum memory and the spectator system.
2. *Free evolution* of the joint memory-spectator system under the action of the shared environment, with generally unknown noise parameters.
3. *Quantum parameter estimation* of the environment noise parameters, using the spectator system as a real-time quantum sensor (i.e. a probe).
4. *Post-processing* of the measurement outcomes to extract the value of the locally unbiased estimator, and use it to construct the best-guess recovery map.
5. *Recovery* of the original state of the quantum memory by applying the best-guess recovery map.
6. *Recycling of spectator state*, which prepares it for the next recovery cycle.

There exist systems that satisfy the physical properties of a spectator system. For example, nitrogen-vacancy (NV) centers in diamond, which were used to prove the first loophole-free Bell inequality violation [31], provide both a spectator qutrit and a memory qubit. Namely, the nuclear spin degree of freedom of its  $^{14}\text{N}$  or  $^{15}\text{N}$  atom comprises the memory qubit, whereas a nearly-closed three-level  $\Lambda$  system [32, 33], optically selected out of the electronic degrees of freedom of the NV center, comprises the spectator system [27, 34, 35]. The two-time separation between the pure dephasing times of the memory qubit ( $T_\varphi^{\text{memo}} \sim 100\mu\text{s}$ ) and the spectator qutrit ( $T_\varphi^{\text{spec}} \sim 100\text{ns}$  [36]) is necessary to simultaneously yields (i) a metrologically useful spectator state  $\mathcal{M}_\theta(\psi)$  for parameter estimation, and (ii) a relatively small noise parameter value of the memory dynamics  $\mathcal{N}_\theta$  (and hence a generally higher recovery fidelity), for relevant times  $t$  of the spectator dynamics. The latter is seen from the implicit dependence of the noise parameter  $\theta$  on time [37]:  $\theta = 1 - \exp(-t/T_\varphi^{\text{memo}}) = 1 - [\exp(-t/T_\varphi^{\text{spec}})]^{T_\varphi^{\text{spec}}/T_\varphi^{\text{memo}}} \ll 1 - \exp(-t/T_\varphi^{\text{spec}}) \equiv \theta_{\text{eff}}$ , where  $\mathcal{M}_\theta \equiv \mathcal{N}_{\theta_{\text{eff}}}$ .

Although spectator systems are a promising building block for real-time (i.e. drift-adapting) quantum memories, we expect fundamental limitations to manifest nonetheless. This is based on the following physical intuition: In real-time, the spectator system's goal is to perform a quantum estimation of the environment noise parameter  $\theta$ . However, due to the quantum Cramér-Rao bound (QCRB)

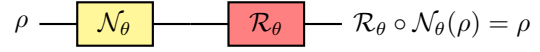


Figure 1: Recovery with perfect knowledge (time flows from left to right). The quantum memory is prepared in the quantum state  $\rho$ . The recovery channel  $\mathcal{R}_\theta$  is implemented using perfect knowledge of the noise parameter  $\theta \in \Theta$  of the environment noise  $\mathcal{N}_\theta$ .

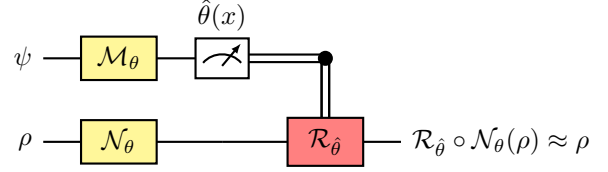


Figure 2: Recovery with limited knowledge (time flows from left to right). The quantum memory (second register) and spectator (first register) systems are prepared in the quantum states  $\rho$  and  $\psi$ , respectively. The recovery channel  $\mathcal{R}_{\hat{\theta}}$  is implemented based on the spectator's best estimate  $\hat{\theta}$  of the noise parameter  $\theta \in \Theta$  of the environment noise  $\mathcal{N}_\theta$ . The estimate is informed by the measurement outcome  $x$  of the spectator observable  $X$ , following the spectator dynamics  $\mathcal{M}_\theta$ .

[38, 39], any locally unbiased estimate  $\hat{\theta}$  of the noise parameter  $\theta$  will have a non-zero variance. Namely,  $\text{Var}(\hat{\theta}) \geq 1/I_{\text{QFI}}(\mathcal{M}_\theta(\psi))$ , where  $I_{\text{QFI}}(\mathcal{M}_\theta(\psi))$  is the quantum Fisher information (QFI) of the family of parametric states  $\{\mathcal{M}_\theta(\psi)\}_{\theta \in \Theta}$  describing the spectator dynamics (see Fig. 2). For a given setup, this fundamental uncertainty in the estimate  $\hat{\theta}$  will propagate within the overall protocol and manifest itself as a fundamental limitation of the specific recovery technique.

In this article, I formalize the above intuition by proving information-theoretic and metrological bounds on the performance of spectator-based recovery protocols for adaptive quantum memories, considering QEC as an example. The main results of the article are summarized below

1. Derivation of a lower bound for the diamond distance between any two quantum channels (Lemma 2). This generalizes a lower bound for the diamond distance between a quantum channel and the identity channel in [40], a qutrit depolarizing channel and the identity channel in [41], as well as the analytic formula in [42] for the diamond distance between two-qubit depolarizing channels. A similar lower bound is shown for generalized distinguishability measures, such as entropy or fidelity-based distinguishability measures (e.g. quantum relative entropy or Bures distance, respectively) is found in Theorem 4 of Appendix A.
2. A general formulation of the spectator-based re-

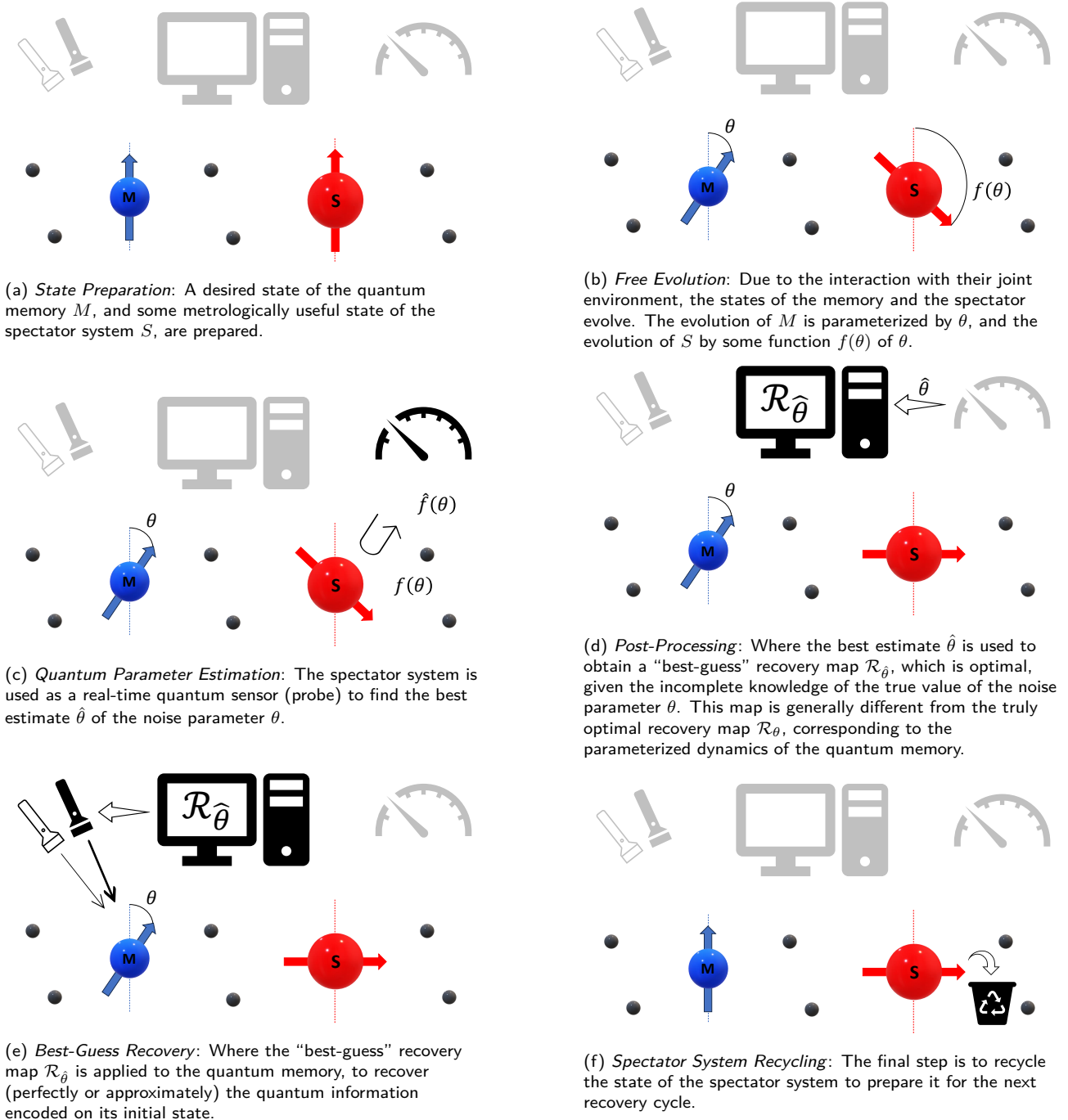


Figure 3: Cartoon description of spectator-based recovery protocols (temporal order corresponds to the alphabetical order of the subfigures). The letters “ $M$ ” and “ $S$ ” stand for quantum memory and spectator system, respectively. The black dots represent the environment spins that contribute to the noise. This protocol combines two important disciplines of quantum information theory: quantum parameter estimation and recovery of quantum information.

covery protocol as two consecutive, but complementary, tasks in a real-time quantum memory (Section 4.2): (i) a multi-parameter quantum estimation in the presence of nuisance parameters, and (ii) recovery of quantum information using the “best-guess” recovery map, as informed by the estimation outcome.

3. Derivation of Information-theoretic costs of adaptation in spectator-based recovery protocols. This is shown both for finite (Theorem 1) and small (Theorem 2) estimation errors of the noise parameters. The latter yields a metrological lower bound, in terms of the quantum Fisher information of the spectator dynamics. The adaptation cost is illustrated for the [4,1] code of the amplitude-damping channel (Figs. 6 and 7), by comparing the performance of the spectator-based recovery protocol with the corresponding optimal recovery protocol [12, 14] when no adaptation is required.
4. Reformulation of an upper bound for the entanglement fidelity of concatenated quantum channels (building upon a theorem of [43]) in the form of recurrence inequalities for multi-cycle recovery protocols (Lemma 4). These bounds are growing in relevance, as multiple QEC rounds have been demonstrated in practice [44]. It is shown that spectator-based recovery protocols, under conditions of varying noise, could outperform optimal recovery protocols with constant noise (Theorem 3). This is exclusively a multi-cycle phenomenon, where errors from different cycle numbers can cohere [43], including errors from incomplete knowledge of the noise parameter. Finally, this is illustrated for the [4,1] code of the amplitude-damping channel (Fig. 9). Similar error coherence effects have been seen recently in the context of quantum state transfer in quantum networks, between gate and readout errors [45].

## 2 Preliminaries

### 2.1 Quantum States and Channels

Let  $\mathcal{H}$  denote a Hilbert space, and  $\mathcal{L}(\mathcal{H})$  be the set of bounded linear operators acting on  $\mathcal{H}$ . Denote by  $\mathcal{L}_+(\mathcal{H})$  the subset of positive semi-definite operators of  $\mathcal{L}(\mathcal{H})$ . We define the Hilbert-Schmidt inner product between two linear operators  $A, B \in \mathcal{L}(\mathcal{H})$  to be  $\langle A, B \rangle := \text{Tr}[A^\dagger B]$ . The state of a physical system is described by a density matrix  $\rho \in \mathcal{D}(\mathcal{H})$ , where  $\mathcal{D}(\mathcal{H})$  is the subset of positive semi-definite linear operators  $\mathcal{L}_+(\mathcal{H})$  that have a unit trace. We denote the dimensions of a Hilbert space  $\mathcal{H}$  by  $d := \dim \mathcal{H}$ .

A linear map from  $\mathcal{L}(\mathcal{H}^A)$  to  $\mathcal{L}(\mathcal{H}^B)$  is denoted by  $\mathcal{Q}^{A \rightarrow B} : \mathcal{L}(\mathcal{H}^A) \rightarrow \mathcal{L}(\mathcal{H}^B)$ . We say that a lin-

ear map is positive if  $\mathcal{Q}^{A \rightarrow B}(L_A) \in \mathcal{L}_+(\mathcal{H}^B)$  for all  $L_A \in \mathcal{L}_+(\mathcal{H}^A)$ , and trace preserving (TP) if  $\text{Tr}[\mathcal{Q}^{A \rightarrow B}(L_A)] = \text{Tr}[L_A]$  for all  $L_A \in \mathcal{L}(\mathcal{H}^A)$ . A positive linear map  $\mathcal{Q}^{A \rightarrow B}$  is called completely positive (CP) if for every Hilbert space  $\mathcal{H}^R$ , the map  $\text{id}^R \otimes \mathcal{Q}^{A \rightarrow B}$  is positive, where  $\text{id}^R$  is the identity map acting on  $\mathcal{L}(\mathcal{H}^R)$ . For any  $\mathcal{Q}^{A \rightarrow B}$  and  $\mathcal{S}^{B \rightarrow C}$  linear maps, their composition is defined to be  $(\mathcal{S} \circ \mathcal{Q})^{A \rightarrow C}(L_A) := \mathcal{S}(\mathcal{Q}(L_A))$ , for all  $L_A \in \mathcal{L}(\mathcal{H}^A)$ . We define the adjoint map  $(\mathcal{Q}^{A \rightarrow B})^\dagger$  of a linear map  $\mathcal{Q}^{A \rightarrow B}$  with respect to the Hilbert-Schmidt inner product as  $\langle \mathcal{Q}^{A \rightarrow B}(N_A), M_B \rangle = \langle N_A, (\mathcal{Q}^{A \rightarrow B})^\dagger(M_B) \rangle$  for all  $N_A \in \mathcal{L}(\mathcal{H}^A)$  and  $M_B \in \mathcal{L}(\mathcal{H}^B)$ . More explicitly, if  $\{Q_i\}_{i=1}^K$  are the Kraus operators of the CP map  $\mathcal{Q}^{A \rightarrow B}$  (see Eq. (2)), then the Kraus operators of the adjoint map  $(\mathcal{Q}^{A \rightarrow B})^\dagger$  are given by  $\{Q_i^\dagger\}_{i=1}^K$ .

The Choi Matrix of any linear map  $\mathcal{Q}^{A \rightarrow B}$  is defined to be

$$\Gamma_{RB}^{\mathcal{Q}} := \text{id}^R \otimes \mathcal{Q}^{A \rightarrow B}(|\Gamma\rangle\langle\Gamma|_{RA}), \quad (1)$$

where  $|\Gamma\rangle_{RA} := \sum_{i=0}^{d-1} |i\rangle_R |i\rangle_A$  is the unnormalized maximally entangled state, with  $d \equiv \dim \mathcal{H}^A = \dim \mathcal{H}^R$ . The corresponding Choi state is defined as  $\Phi_{RB}^{\mathcal{Q}} := \Gamma_{RB}^{\mathcal{Q}}/d$ . The linear map  $\mathcal{Q}^{A \rightarrow B}$  is TP if and only if its Choi matrix satisfies  $\text{Tr}_B[\Gamma_{RB}^{\mathcal{Q}}] = I_R$ , and CP if and only if its Choi matrix is positive, i.e.  $\Gamma_{RB}^{\mathcal{Q}} \geq 0$ .

In what follows, we suppress the system subscript and/or superscript if it does not lead to ambiguities. Every CP map  $\mathcal{Q}^{A \rightarrow B}$  admits a Kraus decomposition

$$\mathcal{Q}(\cdot) = \sum_{i=1}^K Q_i(\cdot)Q_i^\dagger, \quad (2)$$

in terms of Kraus operators  $\{Q_i\}_{i=1}^K$ . If  $\mathcal{Q}^{A \rightarrow B}$  is also TP, then  $\sum_{i=1}^K Q_i^\dagger Q_i = I_A$  holds.

### 2.2 Diamond Distance Between Quantum Channels

The trace norm of any linear operator  $L \in \mathcal{L}(\mathcal{H})$  is given as  $\|L\|_1 := \text{Tr}(|L|)$ , where we have denoted by  $|L| := \sqrt{L^\dagger L}$ . Therefore, we define the trace distance between any two quantum states  $\rho, \sigma \in \mathcal{D}(\mathcal{H})$  to be  $\frac{1}{2}\|\rho - \sigma\|_1$ . More generally, we define the diamond distance  $\frac{1}{2}\|\mathcal{Q}^A - \mathcal{S}^A\|_\diamond$  between two quantum channels  $\mathcal{Q}^A$  and  $\mathcal{S}^A$  as follows

$$\sup_{\rho_{RA}} \frac{1}{2} \left\| \left( \text{id}^R \otimes \mathcal{Q}^A \right) (\rho_{RA}) - \left( \text{id}^R \otimes \mathcal{S}^A \right) (\rho_{RA}) \right\|_1. \quad (3)$$

### 2.3 Quantum Fidelities

Given any two quantum states  $\rho, \sigma \in \mathcal{D}(\mathcal{H})$  with  $\dim \mathcal{H} \equiv d$ , we define the fidelity function as follows

$$F(\rho, \sigma) := \left( \text{Tr} \left[ \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right] \right)^2 = \|\sqrt{\rho} \sqrt{\sigma}\|_2^2. \quad (4)$$

If one of the two state, say  $\sigma \equiv |\psi\rangle\langle\psi|$ , is pure, then we have  $F(\rho, \psi) = \langle\psi|\rho|\psi\rangle$ . Based on this definition, various fidelities that are relevant in QEC and other areas of quantum information have been defined. One such quantity is called entanglement fidelity  $F_e$  of a channel  $\mathcal{Q}^A \equiv \mathcal{Q}^{A \rightarrow A}$  with respect to a state  $\rho \in \mathcal{D}(\mathcal{H}^A)$  [7, 8], which is given by

$$F_e(\mathcal{Q}, \rho) := F(\text{id}^R \otimes \mathcal{Q}^A(\psi_{RA}^\rho), \psi_{RA}^\rho) \quad (5)$$

$$= \langle\psi^\rho|\text{id}^R \otimes \mathcal{Q}^A(\psi_{RA}^\rho)|\psi^\rho\rangle_{RA}, \quad (6)$$

where  $\psi_{RA}^\rho \in \mathcal{H}^R \otimes \mathcal{H}^A$  is a purification of the density matrix  $\rho \in \mathcal{D}(\mathcal{H}^A)$ , i.e.  $\text{Tr}_R \psi_{RA}^\rho = \rho_A$ . It can be shown that the entanglement fidelity is independent of the particular choice of the purification, following from the fact that the former can be expressed in terms of the Kraus operators of  $\mathcal{Q}^A$  as [7, 8]

$$F_e(\mathcal{Q}, \rho) = \sum_{i=1}^K \text{Tr}[\rho Q_i] \text{Tr}[\rho Q_i^\dagger] = \sum_{i=1}^K |\text{Tr}[\rho Q_i]|^2. \quad (7)$$

The entanglement fidelity of a quantum channel  $\mathcal{Q}^A$  is defined to be the entanglement fidelity of  $\mathcal{Q}^A$  with respect to the maximally mixed state  $\rho_A = I_A/d$  [46] (which is purified by the maximally entangled state  $|\Phi\rangle_{RA}$ ). This can also be written in terms of the Choi state of  $\mathcal{Q}^A$ , as follows

$$F_e(\mathcal{Q}) \equiv F_e\left(\mathcal{Q}, \frac{I}{d}\right) \quad (8)$$

$$= \langle\Phi|\text{id}^R \otimes \mathcal{Q}^A(\Phi_{RA})|\Phi\rangle_{RA} \quad (9)$$

$$= \langle\Phi|\Phi_{RA}^\mathcal{Q}|\Phi\rangle_{RA} \quad (10)$$

$$= F(\Phi^\mathcal{Q}, \Phi). \quad (11)$$

Another important fidelity measure of the form  $F(\rho, \psi) = \langle\psi|\rho|\psi\rangle$  is the average channel (gate) fidelity  $F_{\text{avg}}(\mathcal{Q})$ , defined for any  $|\psi\rangle \in \mathcal{H}^A$  and CPTP map  $\mathcal{Q}^A$  as

$$F_{\text{avg}}(\mathcal{Q}) := \int d\psi \langle\psi|\mathcal{Q}(\psi)|\psi\rangle, \quad (12)$$

where the discrete version has appeared in [7, 8]. In [47], the authors have shown that the average and entanglement fidelities are related by

$$F_{\text{avg}}(\mathcal{Q}) = \frac{dF_e(\mathcal{Q}) + 1}{d + 1}. \quad (13)$$

Finally, in what follows, we also use the simplifying notation

$$\delta^\mathcal{Q} := \arccos \sqrt{F_e(\mathcal{Q})}, \quad (14)$$

which can be interpreted in the  $\chi$ -matrix representation of quantum channels (see Appendix I) as the ‘‘error angle’’ by which the Kraus operators of the noisy channel  $\mathcal{Q}$  deviate from the desired ‘‘no error’’ normalized basis element  $B_0 = I/\sqrt{d}$  (where  $\langle B_0, B_0 \rangle = 1$ ) of the vector space  $\mathcal{L}(\mathcal{H})$ . It turns out that the error angle notation is very convenient when expressing the average fidelity of composite channels in terms of the individual average channel fidelities [43].

### 3 Lower-Bounding Diamond Distance Using Entanglement Fidelity

We now consider the diamond distance between any two quantum channels and show that it is lower bounded by the difference between their entanglement fidelities. This is generalized in Theorem 4 of Appendix A, where we show that the lower bound for any generalized distinguishability measure between the two channels is still fully determined by their entanglement fidelities. Besides the diamond distance considered here, fidelity and entropy-based distinguishability measures, such as Bures distance and quantum relative entropy, are also used to quantify the performance of recovery protocols, e.g. in Refs. [48, 49] and Refs. [50, 51], respectively. Therefore, the results of this section (as well as the following sections) could be generalized for other distinguishability measures, in the light of Appendix A.

The diamond distance is especially relevant for two reasons: (1) it has a clear operational meaning in terms of the maximum probability of distinguishing between two channels in a quantum channel discrimination task [52], and (2) it satisfies the triangle inequality and hence also the chaining property, which is useful for bounding errors in fault-tolerant quantum computing (see Appendix B).

We start by proving the following:

**Lemma 1.** *For any two depolarizing channels  $\tilde{\mathcal{Q}}^{A \rightarrow A}$  and  $\tilde{\mathcal{S}}^{A \rightarrow A}$ , the diamond distance between them is equal to the difference between their entanglement fidelities, namely*

$$\frac{1}{2} \|\tilde{\mathcal{Q}} - \tilde{\mathcal{S}}\|_\diamond = |F_e(\tilde{\mathcal{Q}}) - F_e(\tilde{\mathcal{S}})|. \quad (15)$$

*Proof.* Assume that  $\tilde{\mathcal{Q}}$  and  $\tilde{\mathcal{S}}$  are depolarizing channels with depolarizing parameters  $p^\mathcal{Q}$  and  $p^\mathcal{S}$ , respectively. Namely,

$$\tilde{\mathcal{Q}}^A = (1 - p^\mathcal{Q})\text{id}^A + p^\mathcal{Q} \frac{I_A}{d_A} \text{Tr}_A, \quad (16)$$

$$\tilde{\mathcal{S}}^A = (1 - p^\mathcal{S})\text{id}^A + p^\mathcal{S} \frac{I_A}{d_A} \text{Tr}_A, \quad (17)$$

and hence

$$\text{id}^R \otimes \tilde{\mathcal{Q}}^A = (1 - p^\mathcal{Q})\text{id}^{RA} + p^\mathcal{Q} \text{Tr}_A \otimes \frac{I_A}{d_A}, \quad (18)$$

$$\text{id}^R \otimes \tilde{\mathcal{S}}^A = (1 - p^\mathcal{S})\text{id}^{RA} + p^\mathcal{S} \text{Tr}_A \otimes \frac{I_A}{d_A}, \quad (19)$$

yields for the diamond distance

$$\frac{1}{2} \|\tilde{\mathcal{Q}} - \tilde{\mathcal{S}}\|_\diamond = \kappa |p^\mathcal{S} - p^\mathcal{Q}| \quad (20)$$

$$= \frac{d\kappa}{d-1} |F_{\text{avg}}(\tilde{\mathcal{Q}}) - F_{\text{avg}}(\tilde{\mathcal{S}})| \quad (21)$$

$$= \frac{d^2\kappa}{d^2-1} |F_e(\tilde{\mathcal{Q}}) - F_e(\tilde{\mathcal{S}})|, \quad (22)$$

where

$$\kappa(d) \equiv \frac{1}{2} \sup_{\psi_{RA}} \left\| \psi_{RA} - \rho_R \otimes \frac{I_A}{d_A} \right\|_1, \quad (23)$$

and we have used Eq. (13). We can rewrite  $\kappa$  using the diamond distance between the identity and the replacement channel, as follows

$$\kappa(d) = \frac{1}{2} \left\| \text{id}^A - \frac{I_A}{d} \text{Tr}_A \right\|_{\diamond}. \quad (24)$$

Next, we use the semi-definite program for the normalized diamond norm [53]

$$\frac{1}{2} \left\| \text{id}^A - \frac{I_A}{d} \text{Tr}_A \right\|_{\diamond} \quad (25)$$

$$= \sup_{\sigma_{RA}, \rho_R} \text{Tr}_{RA} \left[ \sigma_{RA} \left( \Gamma_{RA}^{\text{id}} - \Gamma_{RA}^{\frac{I_A}{d} \text{Tr}_A} \right) \right] \quad (26)$$

$$= \sup_{\sigma_{RA}, \rho_R} \text{Tr}_{RA} \left[ \sigma_{RA} \left( \Gamma_{RA} - \frac{I_{RA}}{d} \right) \right], \quad (27)$$

where the supremum is taken over all positive matrices  $0 \leq \sigma_{RA} \leq \rho_R \otimes I_A$ , and  $\rho_R \in \mathcal{D}(\mathcal{H}^R)$ . Consider the eigenvalues of the  $d^2 \times d^2$  matrix  $\Gamma_{RA} - I_{RA}/d$ . First, if we denote some fixed eigenvalue of  $\Gamma_{RA}$  by  $\gamma$ , then the corresponding eigenvalue of  $\Gamma_{RA} - I_{RA}/d$  is  $\gamma - 1/d$ . Next, let us show that  $d^2 - 1$  of the  $d^2$  eigenvalues of  $\Gamma_{RA}$  are zero. This follows by considering the kernel space of  $\Gamma_{RA}$  (the zero eigenvalue subspace), denoted by  $\ker(\Gamma_{RA}) \subset \mathcal{H}^R \otimes \mathcal{H}^A$ . Namely, for all  $|\psi\rangle_{RA} \in \ker(\Gamma_{RA})$  we have, by definition,  $\Gamma_{RA}|\psi\rangle_{RA} = 0$ . This is true for all states  $|\psi\rangle_{RA}$  for which  $\langle \Gamma | \psi \rangle_{RA} = 0$ , i.e.  $|\psi\rangle_{RA} \in (\text{span}\{|\Gamma\rangle\})^{\perp}$ , which is the  $(d^2 - 1)$ -dimensional orthogonal complement of the one-dimensional subspace  $\text{span}\{|\Gamma\rangle\} \subset \mathcal{H}^R \otimes \mathcal{H}^A$ . Finally, we note that the only non-zero eigenvalue  $\gamma_0$  of  $\Gamma_{RA}$  is determined by the trace  $\text{Tr}_{RA}[\Gamma_{RA}] = d$ , and hence  $\gamma_0 = d$ . Therefore, the eigenvalues of  $\Gamma_{RA} - I_{RA}/d$  are given by the list

$$\text{eigenval} \left( \Gamma - \frac{I}{d} \right) = \left\{ \frac{d^2 - 1}{d}, -\frac{1}{d}, -\frac{1}{d}, \dots, -\frac{1}{d} \right\}. \quad (28)$$

As we can see, only one of the eigenvalues of  $\Gamma_{RA} - I_{RA}/d$  is positive. Therefore, we write the spectral decomposition of the matrix  $\Gamma_{RA} - I_{RA}/d$  as follows

$$\Gamma_{RA} - \frac{I_{RA}}{d} = \frac{d^2 - 1}{d} |\gamma_0\rangle\langle\gamma_0| - \frac{1}{d} \sum_{i=1}^{d^2-1} |\gamma_i\rangle\langle\gamma_i|, \quad (29)$$

where  $\{|\gamma_i\rangle\}_{i=0}^{d^2-1}$  is its orthonormal eigenbasis.

It follows that, to maximize the argument of Eq. (27), we need to consider the support of the positive semi-definite operator  $\sigma_{RA}$  to be in the (one-dimensional) support of  $\Gamma_{RA}$  (which is orthogonal to  $\ker(\Gamma_{RA})$ ), i.e. we need to search for  $\sigma_{RA}$  in the form  $\sigma_{RA} = z|\gamma_0\rangle\langle\gamma_0|$  for some  $z \geq 0$ . Substituting into the constraint  $\sigma_{RA} \leq \rho_R \otimes I_A$  gives

$$\rho_R \otimes I_A - z|\gamma_0\rangle\langle\gamma_0| \geq 0 \quad (30)$$

$$\Rightarrow \langle \Gamma | \rho_R \otimes I_A | \Gamma \rangle_{RA} - z |\langle \Gamma | \gamma_0 \rangle|^2 \geq 0 \quad (31)$$

$$\Rightarrow \text{Tr}[\rho] - z \langle \gamma_0 | \Gamma | \gamma_0 \rangle_{RA} \geq 0 \quad (32)$$

$$\Rightarrow 1 - zd \geq 0 \Rightarrow z \leq \frac{1}{d}. \quad (33)$$

Consequently, if there exists  $\rho_R \in \mathcal{D}(\mathcal{H}^R)$  such that  $\sigma_{RA}$  is fully in the support of  $\Gamma_{RA}$ , i.e.  $\sigma_{RA} = z|\gamma_0\rangle\langle\gamma_0|_{RA}$  (where  $\Gamma|\gamma_0\rangle = d|\gamma_0\rangle$ ), then it must be the case that the normalization  $z \leq 1/d$ . The resulting maximization in Eq. (27) will thus yield for  $\kappa(d)$

$$\sup_{\sigma_{RA}, \rho_R} \text{Tr}_{RA} \left[ \sigma_{RA} \left( \Gamma_{RA} - \frac{I_{RA}}{d} \right) \right] = \frac{d^2 - 1}{d} z_{\max} \quad (34)$$

$$= \frac{d^2 - 1}{d^2}, \quad (35)$$

or equivalently,

$$\frac{1}{2} \|\tilde{\mathcal{Q}} - \tilde{\mathcal{S}}\|_{\diamond} = \frac{d^2 \kappa(d)}{d^2 - 1} |F_e(\tilde{\mathcal{Q}}) - F_e(\tilde{\mathcal{S}})| \quad (36)$$

$$= |F_e(\tilde{\mathcal{Q}}) - F_e(\tilde{\mathcal{S}})|. \quad (37)$$

In the above analysis, we presumed the existence of a density matrix  $\rho_R$  for which  $\sigma_{RA} = z|\gamma_0\rangle\langle\gamma_0|_{RA} \leq \rho_R \otimes I_A$ . It is easy to see that the pick  $\rho_R = I_R/d$  satisfies the inequality  $z|\gamma_0\rangle\langle\gamma_0|_{RA} \leq \rho_R \otimes I_A$ , as well as allowing the normalization  $z$  to reach its maximum value  $z_{\max} = 1/d$ .

Note that, if we extend the support  $\sigma_{RA} = z|\gamma_0\rangle\langle\gamma_0| + \sum_{i=1}^{d^2-1} z_i |\gamma_i\rangle\langle\gamma_i|$ , then the above argument (starting from Eq. (30)) still yields  $z \leq 1/d$ , while simultaneously leading to a sub-optimal outcome in Eq. (34) due to the contribution of the negative eigenvalues of  $\Gamma_{RA} - I_{RA}/d$ . Taking  $\sigma_{RA}$  to be off-diagonal in the  $\{|\gamma_i\rangle\}_{i=0}^{d^2-1}$  does not change this argument.  $\square$

It is important to note that this lemma has been known previously for special cases, e.g. in [41, 42] between qubit depolarizing maps ( $d = 2$ ) and between a qudit depolarizing map and the identity map ( $p^{\mathcal{S}} = 0$ ), respectively. However, Pirandola *et al.* in [41] used a different technique to compute essentially the same quantity  $\kappa(d)$  appearing in our derivation of Lemma 1, which crucially does not depend on the depolarizing parameters  $p^{\mathcal{Q}}$  and  $p^{\mathcal{S}}$ .

We now prove a lower bound for the diamond distance between any two quantum channels with the same input and output spaces. We frame this as follows

**Lemma 2.** *For any two CPTP maps  $\mathcal{Q}^{A \rightarrow A}$  and  $\mathcal{S}^{A \rightarrow A}$ , the diamond distance between them is lower bounded by the difference in their entanglement fidelities, namely*

$$\frac{1}{2} \|\mathcal{Q} - \mathcal{S}\|_{\diamond} \geq |F_e(\mathcal{Q}) - F_e(\mathcal{S})|. \quad (38)$$

*Proof.* It is known that any quantum supermap (a linear map from one quantum channel to another) that is a convex combination of Pauli unitary supermaps (also known as “twirling”) renders any input channel  $\mathcal{Q}$  into a depolarizing channel  $\tilde{\mathcal{Q}}$  [54], with the same entanglement fidelity. The Lemma is then a direct consequence of applying the data-processing inequality to the diamond distance  $\|\mathcal{Q} - \mathcal{S}\|_\diamond$  with respect to the Pauli twirling supermap [55], which yields

$$\frac{1}{2}\|\mathcal{Q} - \mathcal{S}\|_\diamond \geq \frac{1}{2}\|\tilde{\mathcal{Q}} - \tilde{\mathcal{S}}\|_\diamond, \quad (39)$$

where  $\tilde{\mathcal{Q}}$  and  $\tilde{\mathcal{S}}$  are the resulting depolarizing channels [47] (also see Lemma 6 in Appendix A). Then, we note that the right-hand side is found from Lemma 1. Finally, the proof is completed by the fact that any random unitary supermap preserves the entanglement fidelity [47], hence  $F_e(\tilde{\mathcal{Q}}) = F_e(\mathcal{Q})$  and  $F_e(\tilde{\mathcal{S}}) = F_e(\mathcal{S})$ .  $\square$

**Remark 1.** *If one of the quantum channels is the identity, then a much simpler derivation could be found in the supplementary material of [40] using the Fuchs-van de Graaf inequality for quantum channels, which yields a two-sided bound on the diamond distance.*

## 4 Fundamental Bounds on Recovery with Incomplete Knowledge

Here, we are interested in applying the lower bound derived in Lemma 2 of the previous section to the spectator-based recovery setting, succinctly described in Fig. 2. We will assume that we are given a parametric family of quantum channels  $\{\mathcal{N}_\theta\}_{\theta \in \Theta}$  that is motivated from certain physical assumptions about the memory-environment interaction, where  $\Theta$  is the allowed range of values for the noise parameter  $\theta$ .

### 4.1 The Regime of Validity

Let us now identify four different noise instability regimes and then expand upon the relevant regime for this article. The four cases are described as follows:

1. When neither the noise family  $\{\mathcal{N}_\theta\}_{\theta \in \Theta}$  nor the true noise parameter  $\theta$  change in time. This case is best described by the “perfect” knowledge scenario in Fig. 1, and is the most common in literature.
2. When the noise family  $\{\mathcal{N}_\theta\}_{\theta \in \Theta}$  remains valid, but the true noise parameter  $\theta$  varies *stroboscopically*. Namely, the characteristic timescale  $\tau_\theta$  for the variations of the true value of  $\theta$  is larger compared to the duration of a single recovery cycle  $\Delta t_{\mathcal{R}}$ . Therefore, variations in the noise parameter are on the timescale of multiple recovery cycles (see e.g. [16] for superconducting qubits). In

this regime, performing the real-time quantum estimation of the new value of  $\theta$  for every recovery cycle becomes useful, hence the need for a spectator system. This regime is best described by the “incomplete” knowledge scenario in Fig. 2.

3. When the noise family  $\{\mathcal{N}_\theta\}_{\theta \in \Theta}$  remains valid, but the noise parameter  $\theta$  varies non-negligibly during a single recovery cycle (i.e.  $\tau_\theta \sim \Delta t_{\mathcal{R}}$ ). In this case, the usefulness of the classical side-information in the spectator-based recovery in Fig. 2 is no longer clear. Instead, this noise regime might benefit from continuously applied recovery, e.g. [56]. Alternatively, a robustness approach (as opposed to adaptation) might also be suitable (see below).
4. When the noise family  $\{\mathcal{N}_\theta\}_{\theta \in \Theta}$  is changing within a single recovery cycle. This noise regime will mainly benefit from the design of robust recovery protocols, e.g. in [57–60], rather than the spectator-based recovery presented in Fig. 2.

In the rest of the article, we focus on the stroboscopic noise regime, and consider the advantages and limitations of using the spectator-based recovery protocol in Fig. 2.

To measure the success of the recovery protocol, we recall that the goal is to achieve a complete (or, at least, an approximate) recovery  $\mathcal{R} \circ \mathcal{N}_\theta(\rho) = \rho$  of the noisy channel  $\mathcal{N}_\theta$  for a subset of states  $\rho \in \mathcal{D}(\mathcal{C}) \subset \mathcal{D}(\mathcal{H})$  in the codespace  $\mathcal{C}$ . The “optimality” of the recovery map  $\mathcal{R}$  for a given noisy channel  $\mathcal{N}_\theta$  could be quantified in various ways. Motivated by Lemma 2 for the diamond distance (and more generally Theorem 4 for all distinguishability measures), we choose the entanglement fidelity to be the quantifier of the optimal recovery, i.e.

$$\mathcal{R}_\theta := \operatorname{argmax}_{\mathcal{R}} F_e(\mathcal{R} \circ \mathcal{N}_\theta); \mathcal{R} \in \text{CPTP}(\mathcal{H}). \quad (40)$$

Indeed, entanglement fidelity has been used as a figure of merit for QEC in e.g. [12, 14, 46, 57, 58, 61]. We note that the concatenated form  $\mathcal{R} \circ \mathcal{N}_\theta$  of the memory dynamics presupposes that the recovery map  $\mathcal{R}$  is applied much faster than the noisy dynamics. In what follows, we shall compare two different scenarios:

- *Optimal recovery* scenario (Fig. 1), which corresponds to the optimal choice of the recovery map  $\mathcal{R}_\theta$  (as defined in Eq. (40)) for the noisy channel  $\mathcal{N}_\theta$ , where the value of  $\theta \in \Theta$  is completely known.
- *Best-guess recovery* scenario (Fig. 2), which corresponds to the optimal recovery choice  $\mathcal{R}_{\hat{\theta}}$  (as defined in Eq. (40)) for the estimated noisy channel  $\mathcal{N}_{\hat{\theta}}$ , where  $\hat{\theta}$  is the best estimate of  $\theta$ . The latter is defined to be the minimum variance unbiased estimator (MVUE).



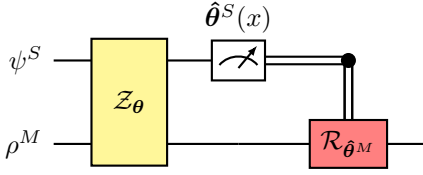


Figure 4: The spectator ( $S$ ) and memory ( $M$ ) systems are subject to a single environment, characterized by the parametric family of quantum channels  $\{\mathcal{Z}_\theta\}_{\theta \in \Theta}$ . The mother channel  $\mathcal{Z}_\theta^{MS}$  generates the two local channels  $\mathcal{N}_{\theta^M}^{MS \rightarrow M} \equiv \text{Tr}_S \circ \mathcal{Z}_\theta$  and  $\mathcal{M}_{\theta^S}^{MS \rightarrow S} \equiv \text{Tr}_M \circ \mathcal{Z}_\theta$  describing the reduced dynamics of the memory and spectator systems, respectively.

**Remark 2.** Not all QEC codes require a recovery channel  $\mathcal{R}_\theta$  that depends on the noise parameter  $\theta$ . Such channels are known to be Pauli channels, where the recovery operation is fully determined by a subset of Pauli operators from the general Pauli group, as is known in the stabilizer formalism of QEC [37, 62]. Therefore, in what follows, we consider non-Pauli channels, of which, the generalized amplitude damping channel is a prime example [37].

## 4.2 A General Framework For Spectator-Based Recovery Protocols

In this section, we formulate the spectator-based recovery protocol (Fig. 2) as the combination of two consecutive processes: (i) parameter estimation of the memory noise, using the spectator system, and (ii) application of the corresponding “best-guess” recovery map (informed by the estimated value of the noise parameter(s), rather than the true value) to the quantum memory. It is important to emphasize that the parameter estimation stage of spectator-based recovery is more resource efficient than a direct process tomography of the noise. This is due to the prior knowledge of the parametric family of noisy quantum channels, either from physical consideration, or even initial process tomography. This efficiency is especially important for operating a real-time quantum memory, which requires such resources to be replenished after each recovery cycle.

To set up a general framework for spectator-based recovery protocols, we start by considering the origins of the reduced memory ( $M$ ) and spectator ( $S$ ) dynamics in the general multi-parameter regime. In what follows,  $MS$  denotes the mother system for both the quantum memory  $M$  and the spectator system  $S$ , and its Hilbert space has the tensor product structure  $\mathcal{H}^{MS} = \mathcal{H}^M \otimes \mathcal{H}^S$ . Since both systems are subject to the same local environment, we assume that there is a mother channel  $\mathcal{Z}_\theta^{MS}$  (with a noise parameter vector  $\theta \in \Theta^p$ , where  $\Theta^p$  is a  $p$ -dimensional parameter space) that acts on the memory and spectator sys-

tems collectively, as shown in Fig. 4. Without loss of generality, we assume that the components of the multiparameter vector  $(\theta_1, \dots, \theta_p) \in \Theta^p$  are treated as independent parameters, namely that none of the components could be expressed as a function of the rest, e.g.  $\theta_i = f(\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$ . Otherwise, for every such constraint, the number of independent parameters is reduced by one. The mother channel yields the reduced dynamics

$$\mathcal{N}_{\theta^M}^{MS \rightarrow M} \equiv \text{Tr}_S \circ \mathcal{Z}_\theta^{MS}, \quad (41)$$

$$\mathcal{M}_{\theta^S}^{MS \rightarrow S} \equiv \text{Tr}_M \circ \mathcal{Z}_\theta^{MS}, \quad (42)$$

where the partial tracing with respect to  $S$  in the first equation defines the partition of the global set of noise parameters  $\theta$  into a subset of relevant parameters  $\theta^M \in \Theta^{p^M}$  for the reduced dynamics of  $M$  (where in general  $p^M \leq p$ ) and a complementary set of parameters  $\theta^{M^\perp} \in \Theta^{p-p^M}$  that is irrelevant for the reduced dynamics (namely  $\Theta^p = \Theta^{p^M} \times \Theta^{p-p^M}$ ). A similar partition of  $\theta = \{\theta^S, \theta^{S^\perp}\}$  to relevant and irrelevant parameters for the spectator dynamics occurs when applying the partial trace with respect to  $M$  to the mother channel (please see Appendix D for details on when this is possible). With these two natural partitions  $\{\theta^M, \theta^{M^\perp}\}$  and  $\{\theta^S, \theta^{S^\perp}\}$  of the parameter vectors  $\theta$ , we can also decompose the latter via a joint partition, as follows

$$\theta = (\theta^S \cap \theta^M) \cup (\theta^S \cap \theta^{M^\perp}) \cup (\theta^{S^\perp} \cap \theta^M) \cup (\theta^{S^\perp} \cap \theta^{M^\perp}). \quad (43)$$

From this joint partition, it is clear that the spectator system  $S$  can only help estimate the subset of memory noise parameters  $\theta_I^S \equiv \theta^S \cap \theta^M \subseteq \theta^M$  (where the subscript “I” stands for “interest” parameters, as opposed to the “nuisance” parameters  $\theta_N^S \equiv \theta^S \cap \theta^{M^\perp} \not\subseteq \theta^M$  of the spectator dynamics [63], which are irrelevant for the memory dynamics). Therefore, we see that it is necessary to have  $\Theta^{p^M} \subseteq \Theta^{p^S}$  for the quantum estimation task via the spectator system to be useful for identifying the relevant noise parameters  $\theta^M$  which affect the quantum memory. If this is not the case, then one would require multiple spectators  $S_1, S_2, \dots, S_l$  (from potentially different physical systems) such that the parameter space  $\Theta^{p^M}$  is contained by the combined parameter subspaces  $\cup_{i=1}^l \Theta^{p^{S_i}}$ . For simplicity, we assume in the rest of the article that  $p_I^S = p^M$ .

In general, the reduced states of  $S$  and  $M$  following the application of Eqs. (41) and (42) will depend on the *global* input state  $MS$ . This observation still holds even if the input state of  $MS$  is of product form  $\rho^{MS} = \rho^M \otimes \psi^S$ . Therefore, to arrive at local noise channels  $\mathcal{N}_{\theta^M}^M$  and  $\mathcal{M}_{\theta^S}^S$  (as shown in Fig. 2) that are independent of the input states of  $S$  and  $M$ , respectively, the mother channel itself has to be separable. Namely

$$\mathcal{Z}_\theta^{MS} = \mathcal{N}_{\theta^M}^M \otimes \mathcal{M}_{\theta^S}^S, \quad (44)$$

similar to the independent noise approximation that is commonly used in QEC literature. The input state  $\rho^{MS} = \rho^M \otimes \psi^S$  is selected such that  $\rho^M$  is the state of the quantum memory that we would like to protect from the noise, whereas  $\psi^S$  is the probe state of the spectator system which we are technically free to choose to achieve the optimal precision in parameter estimation.

In Appendix E, we review relevant aspects of multi-parameter quantum estimation theory. It is known that when there are no nuisance parameters present ( $p_I^S = p^S$ ,  $p_N^S = 0$ ), the quantum estimation limit of the parameters  $\theta_I^S \equiv \theta^M$  is given by the partial symmetric logarithmic derivative (SLD) quantum Fisher information matrix (QFIM), which has dimensions of  $p^S \times p^S$ . However, when  $p_N^S = p^S - p_I^S > 0$  nuisance parameters are present, then the quantum estimation limit will be given by the  $p_I^S \times p_I^S$  partial SLD QFIM, which comprises a tighter lower bound on the estimation variance  $\text{Var}(\hat{\theta}_I^S)$  than the  $p_I^S \times p_I^S$  standard SLD QFIM of the parameters  $\theta_I^S$ . It is known that these two quantities are equal only if the nuisance parameters are informationally orthogonal to the parameters of interest [64], namely when the SLD QFIM of  $\theta^S$  block diagonalizes, with blocks of dimensions  $p_I^S$  and  $p_N^S = p^S - p_I^S$ , respectively. Furthermore, it has been shown in [63] that, in the single parameter regime  $p^M = 1$  ( $p_I^S = 1$ ,  $p_N^S = p^S - 1$ ), the lower bound in the variance of any locally unbiased estimator  $\hat{\theta}_I^S$  is achievable via an optimal measurement constructed from the eigenprojectors of the SLD operators of  $\{\mathcal{M}_{\theta^S}(\psi)\}_{\theta^S}$ . A similar optimal measurement construction achieving the QCRB is not known for  $p^M > 1$  in the presence of nuisance parameters. However, when the nuisance parameters are absent, the above optimal measurement saturates the QCRB for any  $p^M \geq 1$  if and only if the SLD operators of different parameters commute [65]. In the rest of the article, we consider the single parameter case  $p^M = 1$ .

Finally, we would like to point out that the main theorems of this manuscript are generalizable to the multiparameter setting  $p_M > 1$ , however, a full consideration of all its nuances are left for future work. This includes the incompatibility of different parameters [66–68], which is an exclusive problem to the multiparameter regime. Further, it is known that the QCRB is a less tight version of the Holevo Cramér-Rao bound (HCRB) in multi-parameter quantum estimation theory. The latter is known to be efficiently computable [4, 69], but it is generally saturated for collective measurements over different probes. Instead, if one is interested in local measurements (which is more practical for parameter estimation), then the Nagaoka-Hayashi (NH) bound [70] is the relevant bound, which is also efficiently computable [71]. A recent work [72] shows how these Cramér-Rao type bounds are unified under the umbrella of conic linear programming.

### 4.3 Spectator Dynamics With No Nuisance Parameters

In the introduction, as well as Figs. 2 and 6, we have emphasized the fact that the spectator system need not be the same physical system as the computational or memory system. Consequently, the dynamics of the spectator system  $\mathcal{M}_\theta$  is generally different from the dynamics  $\mathcal{N}_\theta = \bigotimes_{i=1}^n \mathcal{N}_\theta^{(1)}$  of the  $n$  memory qubits, though it still depends on the same environment noise parameter  $\theta$ .

#### 4.3.1 Independent Noise Approximation

If the spectator system is made out of  $s$  subsystems (e.g. qubits), then the independent noise model reads

$$\mathcal{M}_\theta = \bigotimes_{i=1}^m \mathcal{M}_\theta^{(1)}, \quad (45)$$

where  $\mathcal{M}_\theta^{(1)}$  is a quantum channel acting on the  $i$ -th subsystem. Note that we assumed negligible spatial variability of the noise parameter  $\theta$ . The noise separability assumption need not mean that the qubit noises are uncorrelated, as classical correlation between the experienced noise parameters by different qubits is still possible in principle. The above separability assumption only means that the noise correlations are classical and hence non-entangling.

#### 4.3.2 Spectator Qubits As Memory Qubits With Controllable Environment Coupling

To perform recovery with incomplete knowledge, we need to hypothesize a relation between the spectator and memory qubit dynamics, i.e.  $\mathcal{M}_\theta^{(1)}$  and  $\mathcal{N}_\theta^{(1)}$ . Since both types of qubits are subject to the same noisy environment with potentially different coupling strengths, we hypothesize

$$\mathcal{M}_\theta^{(1)} = \mathcal{N}_{f(\theta)}^{(1)}, \quad (46)$$

where  $f(\theta) \in [0, 1]$  is a monotone increasing function of its argument. To justify this choice, consider the case where  $\theta$  has the following form

$$\theta = 1 - e^{-t/T_1}, \quad (47)$$

where  $T_1$  is the spin relaxation time [15, 18]. This is the case e.g. for the qubit amplitude-damping channel, which we consider both in Section 4 and Section 6. Then, by expressing  $t/T_1$  in terms of  $\theta$ , we arrive at

$$f_\gamma(\theta) = 1 - (1 - \theta)^\gamma, \quad (48)$$

where  $\gamma = T_1^{\text{memo}}/T_1^{\text{spec}}$ . The requirement that the spectator qubits should exhibit faster dynamics than the memory qubits translates to  $\gamma > 1$ . Eq. (46)

could also be viewed from the point of view of recent progress in quantum control, e.g. via Hamiltonian amplification in bosonic systems [73] or decoherence control in NV centers [74], which yields controllable qubit coupling strengths. If such qubits are used to build a quantum memory, then a portion of these qubits could be reserved as spectators, and the environment coupling could be adjusted to maximize the sensitivity of the spectator qubits (see below).

#### 4.3.3 Quantifying The Physical Choice of Spectator Systems

Given that a particular physical medium (e.g. a spin lattice with multiple spin species A, B, C, ...) is populated with memory qubits (say, spin species A), a natural question is whether the spectator qubits should be chosen from the same or different spin species. This question is especially relevant for QEC in hybrid spin registers e.g. in diamond [75, 76]. To answer this question, we recall that in quantum estimation theory, the sensitivity of the spectator qubit dynamics  $\psi \rightarrow \mathcal{M}_\theta^{(1)}(\psi)$  to the noise parameter  $\theta$  is characterized by the QFI of the output state  $\mathcal{M}_\theta^{(1)}(\psi)$ . The sensing advantage of using a spectator qubit of a different species than the memory qubits is then determined by the ratio  $\text{SM} = I_{\text{QF}}(\mathcal{M}_\theta^{(1)}(\psi))/I_{\text{QF}}(\mathcal{N}_\theta^{(1)}(\psi))$  (which we call the spectator multiplier) following the QCRB (please see Appendix E for a self-contained review of QFI and QCRB)

$$\text{Var}(\hat{\theta}) \geq \frac{1}{I_{\text{QF}}(\mathcal{M}_\theta(\psi^{\otimes m}))} \quad (49)$$

$$= \frac{1}{m I_{\text{QF}}(\mathcal{M}_\theta^{(1)}(\psi))} \quad (50)$$

$$\equiv \frac{1}{n' I_{\text{QF}}(\mathcal{N}_\theta^{(1)}(\psi))}, \quad (51)$$

where  $n' \equiv \text{SM} \times m$  indicates the equivalent number of physical memory qubit species used for sensing. In the case where our hypothesis in Eq. (46) holds, we can use the property of the QFI for the change of parameters [65]

$$I_{\text{QF}}(\mathcal{N}_\theta^{(1)}) = I_{\text{QF}}(\mathcal{N}_{f(\theta)}^{(1)}) \left( \frac{df}{d\theta} \right)^2, \quad (52)$$

which yields  $\text{SM} = 1/(df/d\theta)^2$ . Therefore, the optimal spectator, in this case, is the one that experiences an effective parameter  $\theta_{\text{eff}} \equiv f(\theta)$  with  $f'(\theta) = 0$  at the actual value of the parameter (hence effectively yielding an asymptotic estimation regime  $n' \rightarrow \infty$  for a finite  $m$ ). Realistically, since we do not have prior knowledge of the noise parameter  $\theta$ , the ideal  $f(\theta)$  will be mostly constant, i.e.  $f'(\theta) = 0$  (at least in the relevant variability range of  $\theta$ ), with  $f(0) = 0$  and  $f(1) = 1$ .

Finally, we note that the ratio SM quantifies the relative ‘‘speed’’ of the dynamics between the spectator and the memory qubits, following the relation between QFI and the Bures distance between two consecutive ‘‘instances’’ of a channel [77]. This is important for the spectator-based recovery protocol, as the timely feedforward control of the memory qubits based on the classical side information from the spectator is crucial.

## 4.4 Information-Theoretic Bounds

Here we consider the metrological bounds associated with the spectator-based recovery protocol, following Lemma 2 of the previous section.

### 4.4.1 Fundamental Limitations For All Recovery Protocols

Assume that we are given a noise channel  $\mathcal{N}^{A \rightarrow B}$ . By picking  $\mathcal{Q}^A \equiv \mathcal{R}^{B \rightarrow A} \circ \mathcal{N}^{A \rightarrow B}$  and  $\mathcal{S}^A \equiv \text{id}^A$ , Lemma 2 can be reframed in the context of noise recovery to read

$$\frac{1}{2} \|\mathcal{R} \circ \mathcal{N} - \text{id}\|_\diamond \geq 1 - F_e(\mathcal{R} \circ \mathcal{N}). \quad (53)$$

We note that this is the lower bound of a two-sided bound on the diamond distance between a quantum channel and the identity channel, recently derived in [40]. This lower bound holds for both recovery with perfect and incomplete knowledge scenarios in Figs. 1 and 2, respectively. Next, we consider the incomplete knowledge scenario and analyze the contribution of the spectator system to the lower bound.

### 4.4.2 Metrological Cost of Spectator-Based Recovery Protocols

Consider the scenario described in Fig. 2, which is what we expect for real-time quantum memories. The best estimate  $\hat{\theta}$  of the unknown  $\theta \in \Theta$  is found by the spectator system for each time-interval over which the value of the *stroboscopic* (slowly varying) variable  $\theta$  is approximately constant. This characteristic timescale of the stroboscopic noise parameter  $\theta$  should be larger than the combined characteristic times of the noisy  $\mathcal{N}_\theta$  and the best-guess recovery  $\mathcal{R}_{\hat{\theta}}$  dynamics. Consequently, the relevant total dynamics of the encoded system is given by  $\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_\theta$ . Compared to the ideal case where the noise parameter is known perfectly, the metrological cost associated with the adaptation of spectator-based quantum memories is given by Lemma 2 as

$$\frac{1}{2} \|\mathcal{R}_\theta \circ \mathcal{N}_\theta - \mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_\theta\|_\diamond \geq F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta) - F_e(\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_\theta), \quad (54)$$

where we have substituted for the quantum channels  $\mathcal{Q}^A \equiv \mathcal{R}_\theta^{B \rightarrow A} \circ \mathcal{N}_\theta^{A \rightarrow B}$  and  $\mathcal{S}^A \equiv \mathcal{R}_{\hat{\theta}}^{B \rightarrow A} \circ \mathcal{N}_\theta^{A \rightarrow B}$ . Note that an upper bound to the diamond distance is

given by  $\frac{1}{2} \|\mathcal{R}_\theta - \mathcal{R}_{\hat{\theta}}\|_\diamond$ , which follows from its definition.

We start by providing a lower bound to the right-hand side in Eq. (54) for arbitrary finite estimation errors  $\hat{\theta} - \theta$ , as follows

**Theorem 1.** Consider a parameterized noise channel  $\mathcal{N}_\theta^{A \rightarrow B}$ , and two arbitrary recovery maps  $\mathcal{R}^{B \rightarrow A}$  and  $\tilde{\mathcal{R}}^{B \rightarrow A}$ . Then, the difference between the corresponding entanglement fidelities of recovery is lower bounded by the Choi states of the individual quantum channels  $\mathcal{N}_\theta$ ,  $\mathcal{R}$ , and  $\tilde{\mathcal{R}}$ , as follows

$$\begin{aligned} & |F_e(\mathcal{R} \circ \mathcal{N}_\theta) - F_e(\tilde{\mathcal{R}} \circ \mathcal{N}_\theta)| \\ & \geq \left\| \Phi_{AB}^{\mathcal{R}} - \Phi_{AB}^{\tilde{\mathcal{R}}} \right\|_\alpha \times \left\| \Phi_{AB}^{\mathcal{N}_\theta} \right\|_\beta, \end{aligned} \quad (55)$$

where  $\alpha \in [0, 1)$  and  $1/\alpha + 1/\beta = 1$  defines the  $\beta < 0$  Hölder dual to  $\alpha$ , and  $\|X\|_{\alpha(\beta)} := (\text{Tr}[|X^{\alpha(\beta)}|])^{1/\alpha(\beta)}$ .

*Proof.* Please see Appendix F.3.  $\square$

Applying this theorem particularly to the optimal  $\mathcal{R}_\theta$  and best-guess  $\mathcal{R}_{\hat{\theta}}$  recovery maps, defined via the optimization in Eq. (40), yields a lower bound to the right-hand side in Eq. (54) for arbitrary estimation errors  $\hat{\theta} - \theta$ .

In this article, we will mainly be interested in the small estimation error  $\hat{\theta} - \theta$  case. Again, this is motivated by physical considerations. For example, temporal variations of  $T_1$  and  $T_2$  times in superconducting qubits have been studied extensively, e.g. in [16, 17]. It was observed that such non-negligible temporal variations occur on timescales much longer ( $\sim 1$  second) than the timescale of a single QEC cycle. Therefore, it is sensible to consider the case where the noise parameter does not vary considerably within a single QEC cycle, and hence we expect  $\hat{\theta} - \theta$  to also be small for each QEC cycle, and its effect only accumulates on the timescales of multiple QEC cycles, as observed e.g. in [44]. Hence, in the rest of the manuscript, we provide lower bounds for the spectator-based recovery protocol in this limit, unless stated otherwise.

We now present the main result of the article, which describes a metrological lower bound to the performance of the spectator-based recovery protocol in Fig. 2, compared to the perfect knowledge case in Fig. 1. For simplicity, this result applies when no nuisance parameters are present. A discussion of how the nuisance parameters will impact the lower bound is given later in Section 5.4.

**Theorem 2.** Consider a parameterized noise channel  $\mathcal{N}_\theta^{A \rightarrow B}$ , and the corresponding optimal  $\mathcal{R}_\theta^{B \rightarrow A}$  and best-guess  $\mathcal{R}_{\hat{\theta}}^{B \rightarrow A}$  recovery maps. For small deviations  $\hat{\theta} - \theta$  of the locally unbiased estimate  $\hat{\theta}$  from the true value  $\theta$ , the difference between the corresponding entanglement fidelities of recovery is lower bounded by

$$\mathbb{E} [F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta) - F_e(\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_\theta)]_{p(x|\theta)} \quad \square$$

$$\geq \frac{g(\theta)}{I_{\text{QF}}(\mathcal{M}_\theta(\psi))} - \mathbb{E} [R(\hat{\theta} - \theta)]_{p(x|\theta)}, \quad (56)$$

where  $g(\theta)$  denotes

$$g(\theta) = -\frac{d_B}{2d_A} \text{Tr}_{AB} \left[ \left( \Phi_{AB}^{\mathcal{N}_\theta} \right)^T \partial_\theta^2 \Phi_{BA}^{\mathcal{R}_\theta} \right]. \quad (57)$$

Further,  $I_{\text{QF}}(\mathcal{M}_\theta(\psi))$  denotes the QFI of the spectator dynamics  $\mathcal{M}_\theta(\psi)$  (initialized in state  $\psi$ ) and  $\mathbb{E}(\cdot)_{p(x|\theta)}$  denotes the expectation with respect to the spectator's measurement statistics  $p_X(x|\theta)$ , corresponding to the measurement outcomes  $x \in \mathcal{X}$  of the spectator observable  $X = \sum_{x \in \mathcal{X}} x \Pi_x$ . Finally,  $R(\hat{\theta} - \theta) \equiv \frac{1}{3!} \partial_\nu^3 F_e(\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta)(\hat{\theta} - \theta)^3$  is the Lagrange remainder of the Taylor series expansion of  $F_e(\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta)$  with respect to  $\nu$ , where  $\nu_0 \in [0, \hat{\theta} - \theta]$  is a constant.

*Proof.* The first part of the proof follows directly from Taylor expanding the entanglement fidelity  $F_e(\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta)$  with respect to the difference  $\nu \equiv \hat{\theta} - \theta$  to the second order, and using the Lagrange form for the remainder, as follows

$$F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta) - F_e(\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_\theta) \quad (58)$$

$$= F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta) - F_e(\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta) \quad (59)$$

$$\begin{aligned} & = \frac{1}{1!} \left( -\frac{d}{d\nu} F_e(\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta) \right) \Big|_{\nu=0} \nu \\ & + \frac{1}{2!} \left( -\frac{d^2}{d\nu^2} F_e(\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta) \right) \Big|_{\nu=0} \nu^2 \\ & - \frac{1}{3!} \frac{d^3}{d\nu^3} F_e(\mathcal{R}_{\theta+\nu_0} \circ \mathcal{N}_\theta) \nu^3, \end{aligned} \quad (60)$$

where  $\nu_0 \in [0, \nu]$  is a constant. Taking the expectation  $\mathbb{E}[\cdot]_{p(x|\theta)}$  of both sides with respect to the spectator's measurement statistics  $p_X(x|\theta)$  of the observable  $X$ , and recalling that  $\hat{\theta}$  is a locally unbiased estimate of  $\theta$  (and hence the QCRB applies), we see that

$$\mathbb{E}[\nu]_{p(x|\theta)} = 0 \quad \text{and} \quad \mathbb{E}[\nu^2]_{p(x|\theta)} \geq \frac{1}{I_{\text{QF}}(\mathcal{M}_\theta(\psi))}. \quad (61)$$

This yields Eq. (56) of our theorem. To prove Eq. (57) of this theorem, we first show in Lemma 8 of Appendix F that the entanglement fidelity of the composite dynamics  $\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta$  of the memory qubit is given by the individual Choi states of the noise and recovery maps, as follows

$$F_e(\mathcal{R}_{\theta+\nu} \circ \mathcal{N}_\theta) = \frac{d_B}{d_A} \text{Tr}_{AB} \left[ \left( \Phi_{AB}^{\mathcal{N}_\theta} \right)^T \Phi_{BA}^{\mathcal{R}_{\theta+\nu}} \right]. \quad (62)$$

Then, Eq. (57) of our theorem is a direct consequence of differentiating this entanglement fidelity formula twice with respect to  $\nu$  (assuming the Choi state of the optimal recovery map  $\mathcal{R}_\theta$  is twice differentiable) for a fixed  $\theta$ , which yields

$$\partial_\nu \Phi^{\mathcal{R}_{\theta+\nu}} \Big|_{\nu=0} = \partial_{\theta+\nu} \Phi^{\mathcal{R}_{\theta+\nu}} \Big|_{\theta+\nu=\theta} \quad (63)$$

$$= \partial_{\theta'} \Phi^{\mathcal{R}_{\theta'}} \Big|_{\theta'=\theta} \equiv \partial_\theta \Phi^{\mathcal{R}_\theta}. \quad (64)$$

$\square$

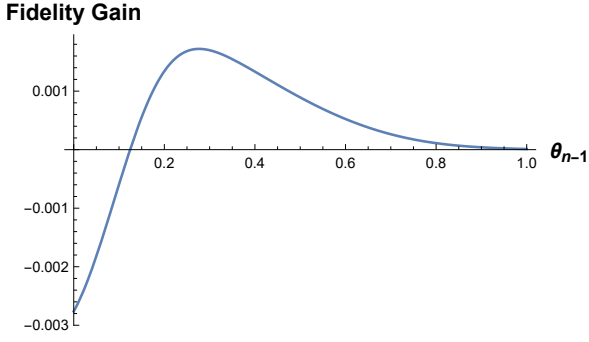


Figure 5: Plot of the fidelity gain Eq. (65) due to using a spectator-based recovery protocol for the [4,1] amplitude damping code (see Section 5). The variation  $\theta_{n-1} \rightarrow \theta_n$  is modeled to be a truncated Gaussian distribution with a maximum at  $\theta_{n-1}$ , and 0.1 standard deviation. The region of outperformance under these conditions is  $\Theta_{\text{spec}} = [0.125, 1] \subset [0, 1]$ .

We note that the saturation of this lower bound is based on the saturation of the QCRB, and is discussed in Appendix E. Moreover, although Theorem 2 is true for any input state  $\psi$  of the spectator system, we would like to use the optimal probe state for the corresponding dynamics  $\mathcal{M}_\theta$ . Furthermore, a sufficient condition for the remainder term in Eq. (56) to be negligible is presented in Appendix G. Further, we note that this theorem could also be interpreted from the information geometric perspective, e.g. in [65, 78].

**Remark 3.** *Due to the tensor product property of the QFI,  $I_{\text{QF}}(\sigma_\theta^{\otimes m}) = mI_{\text{QF}}(\sigma_\theta)$ , the QCRB yields zero variance only in the asymptotic limit. However, in realistic spectator-based recovery protocols (described by Fig. 2), the asymptotic limit (i.e. implementing  $m \rightarrow \infty$  spectator qubits) will necessarily mean that spatial variations of the noise parameter  $\theta$  affecting the spectator qubits cannot be neglected. Therefore, we limit ourselves to the non-zero QCRB variance (finite sample case) and instead attempt to saturate this bound via optimal measurements and initial spectator state (see Appendix E.2).*

Finally, note that the expected dependence of the spectator's contribution separates into the product of two functions: the first,  $g(\theta)$ , depends on the full dynamics of the memory system, and the second,  $\text{Var}(\hat{\theta})$ , depends on the full dynamics of the spectator system. It turns out that the function  $g(\theta)$  could be computed analytically for simple single-qubit channels, such as for the amplitude-damping channel [79].

#### 4.4.3 Comparison With The Non-Adaptive Case

As discussed previously in Section 4.1, the relative advantage of implementing a spectator-based recovery protocol depends on the characteristics of the noise. For example, if the noise is completely static, then implementing a spectator-based recovery will always be

worse than simply characterizing the noise before the experiment (e.g. via process tomography), as there is no advantage to real-time quantum sensing of the noise, where only sparse data is available. Here, we provide a sufficient condition for a spectator-based adaptive protocol to outperform other (non-adaptive) recovery protocols for arbitrary finite estimation error  $\hat{\theta} - \theta$ .

Consider some parameter variation  $\theta_{n-1} \rightarrow \theta_n$  from the  $(n-1)$ -th to the  $n$ -th QEC cycle. This corresponds to the change in the noise  $\mathcal{N}_{\theta_{n-1}} \rightarrow \mathcal{N}_{\theta_n}$ . In a spectator-based recovery protocol, this change is tracked via the spectator system, and the best-guess recovery is updated accordingly  $\mathcal{R}_{\hat{\theta}_{n-1}} \rightarrow \mathcal{R}_{\hat{\theta}_n}$ . The performance of this protocol will hence be quantified by the entanglement fidelity  $F_e(\mathcal{R}_{\hat{\theta}_n} \circ \mathcal{N}_{\theta_n})$ . On the other hand, a non-adaptive protocol will include applying a recovery map  $\mathcal{R}_{\theta_{n-1}}$  that is (in the ideal case) optimal for the previous noise channel, i.e.  $\mathcal{N}_{\theta_{n-1}}$ . The performance of an ideal non-adaptive protocol will hence be quantified by the entanglement fidelity  $F_e(\mathcal{R}_{\theta_{n-1}} \circ \mathcal{N}_{\theta_n})$ . Therefore, to find a sufficient condition for adaptation to yield an advantage, we need to find a non-negative lower bound to the entanglement fidelity difference

$$F_e(\mathcal{R}_{\hat{\theta}_n} \circ \mathcal{N}_{\theta_n}) - F_e(\mathcal{R}_{\theta_{n-1}} \circ \mathcal{N}_{\theta_n}). \quad (65)$$

We accomplish this by using Theorem 1 and Eq. 54, as follows

$$\begin{aligned} & F_e(\mathcal{R}_{\hat{\theta}_n} \circ \mathcal{N}_{\theta_n}) - F_e(\mathcal{R}_{\theta_{n-1}} \circ \mathcal{N}_{\theta_n}) \\ &= [F_e(\mathcal{R}_{\theta_n} \circ \mathcal{N}_{\theta_n}) - F_e(\mathcal{R}_{\theta_{n-1}} \circ \mathcal{N}_{\theta_n})] \\ &\quad - [F_e(\mathcal{R}_{\theta_n} \circ \mathcal{N}_{\theta_n}) - F_e(\mathcal{R}_{\hat{\theta}_n} \circ \mathcal{N}_{\theta_n})] \end{aligned} \quad (66)$$

$$\begin{aligned} &\geq \left\| \Phi^{\mathcal{R}_{\theta_{n-1}}} - \Phi^{\mathcal{R}_{\theta_n}} \right\|_\alpha \times \left\| \Phi^{\mathcal{N}_{\theta_n}} \right\|_\beta \\ &\quad - \frac{1}{2} \left\| \mathcal{R}_{\theta_n} \circ \mathcal{N}_{\theta_n} - \mathcal{R}_{\hat{\theta}_n} \circ \mathcal{N}_{\theta_n} \right\|_\diamond \end{aligned} \quad (67)$$

$$\begin{aligned} &\geq \left\| \Phi^{\mathcal{R}_{\theta_{n-1}}} - \Phi^{\mathcal{R}_{\theta_n}} \right\|_\alpha \times \left\| \Phi^{\mathcal{N}_{\theta_n}} \right\|_\beta \\ &\quad - \frac{1}{2} \left\| \mathcal{R}_{\theta_n} - \mathcal{R}_{\hat{\theta}_n} \right\|_\diamond. \end{aligned} \quad (68)$$

Therefore, a *sufficient* (and initial state independent) condition for the spectator-based recovery protocol to be advantageous for an arbitrary estimation error  $\hat{\theta} - \theta$ , compared to an ideal non-adaptive protocol, is given by

$$\frac{1}{2} \left\| \mathcal{R}_{\theta_n} - \mathcal{R}_{\hat{\theta}_n} \right\|_\diamond \leq c \times \left\| \Phi^{\mathcal{R}_{\theta_{n-1}}} - \Phi^{\mathcal{R}_{\theta_n}} \right\|_\alpha, \quad (69)$$

where  $c \equiv \left\| \Phi^{\mathcal{N}_{\theta_n}} \right\|_\beta$ . It is easy to see that if there is no change, i.e.  $\theta_n = \theta_{n-1}$ , then this condition is not satisfied. On the other hand, for a general (Markovian) stochastic jump model, with a  $\theta_{n-1} \rightarrow \theta_n$  transition probability of  $p(\theta_n|\theta_{n-1})$ , there is a range of values  $\Theta_{\text{spec}} \subseteq \Theta$  of the noise parameter  $\theta_{n-1} \in \Theta$  such that the spectator based recovery exhibits an

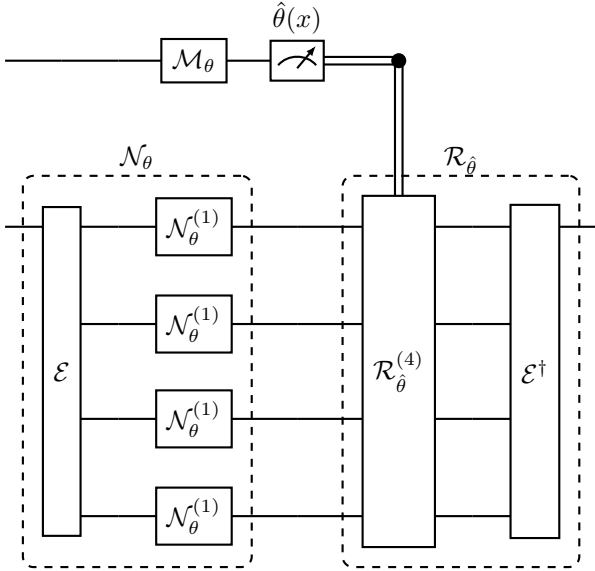


Figure 6: Spectator-based [4, 1] code of the amplitude-damping channel (time flows from left to right). A single logical qubit (second register) is encoded into four physical qubits using the encoding channel  $\mathcal{E} : \mathcal{H}_2 \rightarrow \mathcal{H}_2^{\otimes 4}$  in Eqs. (71) and (72), where  $\mathcal{H}_2$  denotes the two dimensional Hilbert space of a single qubit system. The spectator system (first register) performs an unbiased estimate  $\hat{\theta}(x)$  of the noise parameter  $\theta$  using the POVM  $\{\Pi_x\}$ . The estimated value  $\hat{\theta}$  is fed into the recovery operation described in detail in [12] that is adapted for the amplitude-damping channel.

advantage. Furthermore, this advantage will accumulate over multiple QEC cycles. As an example, in Fig. 5, we plot the fidelity gain in Eq. (65) for the [4,1] amplitude damping code (see the following section), showcasing the region of outperformance of the spectator-based recovery protocol over standard QEC [12] for a stroboscopically varying noise  $\theta_{n-1} \rightarrow \theta_n$ . For more details, please see Appendix H.

## 5 Application to The [4, 1] Code of The Amplitude-Damping Channel

In what follows, we derive the entanglement fidelity  $F_e(\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_{\theta})$  for the [4, 1] code of the amplitude-damping (AD) channel analytically, following the approach developed in [80], and extending the derivation in [79] to the incomplete knowledge recovery scenario. It is worth noting that analytical approaches to the AD channel have also been taken previously e.g. in [81, 82].

Since the AD channel is covariant with respect to the group  $\{I, Z\}$ , the Eastin-Knill theorem [83] guarantees that no perfect QEC codes exist. However, approximate codes for the AD channel have been developed in [13] and later on, channel-adapted codes have been developed [12], where the recovery depends on the value of the noise parameter. The developed

techniques have also been extended beyond the [4, 1] code and towards more general  $[2k+1, k]$  codes [12, 13] (where  $k$  logical qubits are encoded into  $n = 2k + 1$  physical/memory qubits).

### 5.1 The Amplitude-Damping Channel

The single-qubit AD channel is defined as  $\mathcal{N}_{\theta}^{(1)}(\cdot) = N_0(\cdot)N_0^{\dagger} + N_1(\cdot)N_1^{\dagger}$ , where

$$N_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\theta} \end{pmatrix}, \quad N_1 = \begin{pmatrix} 0 & \sqrt{\theta} \\ 0 & 0 \end{pmatrix}. \quad (70)$$

The Kraus operators  $N_0, N_1$  are often called the “no-damping” and “damping” errors, respectively. Here, the noise parameter  $\theta(t) = 1 - \exp(-t/T_1)$  depends on time  $t$  and the relaxation time  $T_1$  [18, 37]. We follow the usual notation in quantum information, where the dependence of the noisy channel (and hence also the noise parameter) on time is suppressed.

### 5.2 The Approximate [4,1] Code

Assuming an independent noise model, we recall the encoding  $\mathcal{E} : \mathcal{D}(\mathcal{H}) \rightarrow \mathcal{D}(\mathcal{C})$  of the [4,1] code [13] from a 1-qubit physical state to a 4-qubit logical state, where  $\mathcal{C} = \text{span}\{|0_L\rangle, |1_L\rangle\} \subset \mathcal{H}^{\otimes 4}$ , as follows

$$|0\rangle \rightarrow |0_L\rangle := \frac{1}{\sqrt{2}} (|0000\rangle + |1111\rangle) \quad (71)$$

$$|1\rangle \rightarrow |1_L\rangle := \frac{1}{\sqrt{2}} (|1100\rangle + |0011\rangle), \quad (72)$$

and hence  $\mathcal{E}(\cdot) := C(\cdot)C^{\dagger}$ , where  $C = |0_L\rangle\langle 0| + |1_L\rangle\langle 1|$ . The encoded Pauli operators  $\sigma_{\text{enc}} = \mathcal{E}(\sigma)$  for  $\sigma \in \{I, X, Y, Z\}$  read

$$I_{\text{enc}} = |0_L\rangle\langle 0_L| + |1_L\rangle\langle 1_L| \quad (73)$$

$$X_{\text{enc}} = |0_L\rangle\langle 1_L| + |1_L\rangle\langle 0_L| \quad (74)$$

$$Y_{\text{enc}} = -i|0_L\rangle\langle 1_L| + i|1_L\rangle\langle 0_L| \quad (75)$$

$$Z_{\text{enc}} = |0_L\rangle\langle 0_L| - |1_L\rangle\langle 1_L|. \quad (76)$$

By definition, the encoded Pauli operators only act on states in the codespace  $\mathcal{C}$ . However, in the stabilizer formalism, the logical Pauli operators  $I_L, X_L, Y_L,$  and  $Z_L$  are defined on the full 4-qubit Hilbert space. For example, the generators of the stabilizer set for the [4, 1] code is given by  $S = \{S_j\}_{j=1}^3 = \{XXXX, ZZII, IIZZ\}$ , along with the logical Pauli operators  $X_L = XXII, Y_L = YXZI,$  and  $Z_L = ZIZI$ . The link between the encoded and logical Pauli operators is found by restricting the action of the latter to the codespace. Namely,  $\sigma_{\text{enc}} = \mathcal{E}(\sigma) = \Pi\sigma_L$ , where  $\Pi = \sum_{j=1}^3 S_j/|S|$  is the projection onto the codespace  $\mathcal{C}$  corresponding to the set of stabilizers [80].

We define the noisy channel  $\mathcal{N}_{\theta}$  to be the physical noise experienced by the four physical qubits in the

[4,1] code post-encoding, as follows

$$\mathcal{N}_\theta = \left( \mathcal{N}_\theta^{(1)} \otimes \mathcal{N}_\theta^{(1)} \otimes \mathcal{N}_\theta^{(1)} \otimes \mathcal{N}_\theta^{(1)} \right) \circ \mathcal{E}. \quad (77)$$

Furthermore, we define the decoding recovery channel  $\mathcal{R}_{\hat{\theta}}$  to be given by

$$\mathcal{R}_{\hat{\theta}} = \mathcal{E}^\dagger \circ \mathcal{R}_{\hat{\theta}}^{(4)}, \quad (78)$$

where  $\mathcal{R}_{\hat{\theta}}^{(4)}$  is taken from Table 1 of [12], which is the channel-adapted recovery of the AD channel (see Fig. 7(b) for the performance of this recovery).

### 5.3 Entanglement Fidelity

In Appendix H, we analytically calculate the numerator  $g(\theta)$  in Theorem 2 for the [4,1] AD code to be

$$g(\theta) = \frac{(1-\theta)^3}{\sqrt{2}(1+(1-\theta)^4)^{3/2}}. \quad (79)$$

It has been shown that the RLD QFI for the AD channel diverges (see the example discussed in [84] for generalized AD channels), however, the SLD QFI is finite and known to be equal to  $\mathsf{l}_{\text{QF}}(\mathcal{N}_\theta^{(1)}) = 1/(\theta(1-\theta)) = \mathsf{l}_{\text{QF}}(\mathcal{N}_\theta^{(1)}(\psi))$  for  $\psi = |1\rangle\langle 1|$  [85]. Consequently, a spectator qubit that satisfies the condition Eq. (46) has an SLD QFI of

$$\mathsf{l}_{\text{QF}}(\mathcal{M}_\theta^{(1)}(\psi)) = \frac{1}{f_\gamma(\theta)(1-f_\gamma(\theta))}, \quad (80)$$

where  $f_\gamma(\theta)$  is given by Eq. (48) for the AD code (or more generally, by Eq. (46)). If the spectator system is made out of  $m$  qubits, then the QFI scales linearly with  $m$  due to the QFI property  $\mathsf{l}_{\text{QF}}(\mathcal{M}_\theta^{(1)}(\psi)^{\otimes m}) = m\mathsf{l}_{\text{QF}}(\mathcal{M}_\theta^{(1)}(\psi))$ , assuming an independent noise model (see Eq. (45)). Note that we can only realistically improve the QCRB to a certain degree by increasing  $m$ , without dropping the negligible spatial variability assumption of the noise parameter  $\theta$  [25].

Combining Eqs. (79) and (80) for the [4,1] AD code with Theorem 2 yields (for small  $\hat{\theta} - \theta$ )

$$\mathbb{E}[\Delta F_e] \geq \frac{g(\theta)}{mf_\gamma(\theta)(1-f_\gamma(\theta))}. \quad (81)$$

Therefore, the contribution of the spectator system to the entanglement fidelity of the [4,1] code of the amplitude-damping channel is determined by two parameters: the number of spectator qubits used ( $m$ ) and their physical nature ( $\gamma = T_1^{\text{memo}}/T_1^{\text{spec}}$ ). When the QCRB is saturated, the resulting entanglement fidelity is illustrated in Fig. 7 for various values of the spectator parameter  $\gamma$ .

### 5.4 The Effect of Nuisance Parameters

In this section, we discuss the effects of nuisance parameters on the lower bound in Theorem 2 for the AD code (i.e. Eq. (81)). We consider three different physical choices of a spectator qubit, which yield one of the following:

1. The presence of an additional constant magnetic field  $B$

$$\mathcal{U}_\phi(\cdot) = U^{-i\phi Z}(\cdot)U^{i\phi Z}, \quad (82)$$

with noise parameter  $\phi = \gamma_{\text{spec}}Bt$ , where  $\gamma_{\text{spec}}$  here is the gyromagnetic ratio of the spectator qubit.

2. The presence of an additional pure dephasing noise

$$\mathcal{P}(\cdot) = P_1(\cdot)P_1^\dagger + P_2(\cdot)P_2^\dagger, \quad (83)$$

where the Kraus operators  $P_1$  and  $P_2$  are given by

$$P_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix}; P_2 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\lambda} \end{pmatrix}, \quad (84)$$

with noise parameter  $\lambda = 1 - \exp(-t/T_\varphi)$ , which yields the dephasing time  $T_2$  with an off-diagonal decay rate of  $1/T_2 = 1/2T_1 + 1/T_\varphi$  [86].

3. The presence of an additional depolarizing noise

$$\mathcal{D}_q(\cdot) = (1-q)(\cdot) + q\frac{I}{2}, \quad (85)$$

with noise parameter  $q \in [0, 1]$ .

When no noise parameters are present, the quantum state of the spectator system

$$\psi = \begin{pmatrix} \psi_{00} & \psi_{01} \\ \psi_{10} & \psi_{11} \end{pmatrix}, \quad (86)$$

is transformed to  $\mathcal{M}_\theta^{(1)}(\psi) = \mathcal{N}_{f(\theta)}^{(1)}(\psi)$ , where

$$\mathcal{N}_{f(\theta)}^{(1)}(\psi) = \begin{pmatrix} \psi_{00} + f(\theta)\psi_{11} & \psi_{01}\sqrt{1-f(\theta)} \\ \psi_{10}\sqrt{1-f(\theta)} & (1-f(\theta))\psi_{11} \end{pmatrix}. \quad (87)$$

However, when a nuisance parameter is present, the output state of the spectator used for quantum estimation limit of  $\theta$  will be modified. In the above three cases, the spectator states are given by, respectively,

$$\mathcal{M}_{\theta,\phi}^{(1)}(\psi) = \begin{pmatrix} \psi_{00} + f(\theta)\psi_{11} & \psi_{01}e^{-i\phi}\sqrt{1-f(\theta)} \\ \psi_{10}e^{i\phi}\sqrt{1-f(\theta)} & (1-f(\theta))\psi_{11} \end{pmatrix}, \quad (88)$$

$$\begin{aligned} \mathcal{M}_{\theta,\lambda}^{(1)}(\psi) &= \\ &= \begin{pmatrix} \psi_{00} + f(\theta)\psi_{11} & \psi_{01}(1-\lambda)\sqrt{1-f(\theta)} \\ \psi_{10}(1-\lambda)\sqrt{1-f(\theta)} & (1-f(\theta))\psi_{11} \end{pmatrix}, \end{aligned} \quad (89)$$

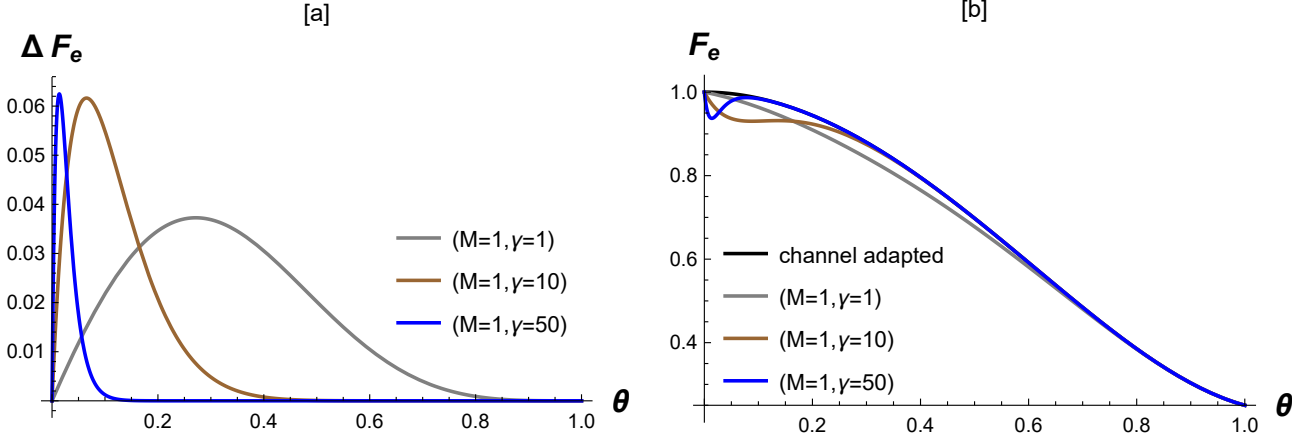


Figure 7: The spectator system is taken to be a single qubit ( $m = 1$ ) with varying values of the physical parameter  $\gamma$  in Eq. (48). Both subfigures consider the channel-adapted approximate [4, 1] code of the amplitude damping channel [12]. (a) Entanglement fidelity difference between the cases of perfect and incomplete knowledge recovery protocols. (b) Comparison between the entanglement fidelities for perfect [12, 79] and incomplete knowledge recovery protocols. In both figures, we assume the best-case scenario where the spectator system saturates the QCRB during parameter estimation.

and

$$\mathcal{M}_{\theta,q}^{(1)}(\psi) = q\mathcal{M}_{\theta}^{(1)}(\psi) + (1-q)\frac{I}{2}. \quad (90)$$

The quantum estimation limit of the parameter of interest  $\theta$ , in the presence of one of the nuisance parameters  $\zeta = \{\phi, \lambda, q\}$  above, is given by the partial QFIM  $l_{\theta|\zeta}$  via  $\text{Var}(\hat{\theta}) \geq 1/l_{\theta|\zeta}$ , where

$$l_{\theta|\zeta} = l_{\theta,\theta} - l_{\theta,\zeta} (l_{\zeta,\zeta})^{-1} l_{\zeta,\theta} \leq l_{\theta,\theta}, \quad (91)$$

and the right hand side are the block matrices of the QFIM (see Appendix E for more details)

$$l_{\text{QF}} \left( \mathcal{M}_{\theta,\zeta}^{(1)}(\psi) \right) = \begin{pmatrix} l_{\theta,\theta} & l_{\theta,\zeta} \\ l_{\zeta,\theta} & l_{\zeta,\zeta} \end{pmatrix}. \quad (92)$$

It is important to emphasize that the matrix element  $l_{\theta,\theta}$  refers to the quantum estimation limit of  $\theta$ , when the noise parameter  $\zeta$  is known. Therefore, we expect to have  $\lim_{\zeta \rightarrow 0} l_{\theta,\theta} = l_{\text{QF}}(\mathcal{M}_{\theta}^{(1)}(\psi))$ , which is easily verified numerically for the above three examples. However, a nuisance parameter  $\zeta$ , similar to the parameter of interest  $\theta$ , is, by nature, *unknown*. Hence, the quantum estimation limit would be reduced by  $l_{\theta,\theta} - l_{\theta|\zeta}$  due to the presence of the unknown noise parameter  $\zeta$ , compared to when it is known.

We now present the impact of the nuisance parameter for the spectator qubit initial state  $\psi = |1\rangle\langle 1|$  that maximizes the QFI  $l_{\text{QF}}(\mathcal{M}_{\theta}^{(1)}(\psi))$  for the AD channel  $\mathcal{M}_{\theta}^{(1)}$  [85]. Simple computation shows that

1. For  $\zeta = \phi$ , we have

$$l_{\theta|\phi} = l_{\theta,\theta} = \frac{1}{f(\theta)(1-f(\theta))} = l_{\text{QF}}(\mathcal{M}_{\theta}^{(1)}(\psi)). \quad (93)$$

2. For  $\zeta = \lambda$ , it still holds that

$$l_{\theta|\lambda} = l_{\theta,\theta} = \frac{1}{f(\theta)(1-f(\theta))} = l_{\text{QF}}(\mathcal{M}_{\theta}^{(1)}(\psi)). \quad (94)$$

This is a direct consequence of the fact that the optimal spectator input state  $\psi = |1\rangle\langle 1|$  has no off-diagonal elements subject to dephasing.

3. For  $\zeta = q$ , we have

$$l_{\theta,\theta} = \frac{(1-q)^2}{(1-q)^2(f(\theta)(1-f(\theta)) - \frac{1}{4}) + \frac{1}{4}}, \quad (95)$$

which also yields  $\lim_{q \rightarrow 0} l_{\theta,\theta} = l_{\text{QF}}(\mathcal{M}_{\theta}^{(1)}(\psi)) = 1/(f(\theta)(1-f(\theta)))$ . However, the partial QFI is computed to be  $l_{\theta|q} = 0$ .

## 6 Recovery Bounds in The Multi-Cycle Scenario

### 6.1 The Multi-Cycle Case

In the article, we have considered a single-cycle recovery, i.e. when the noisy channel  $\mathcal{N}_{\theta}$  is applied only



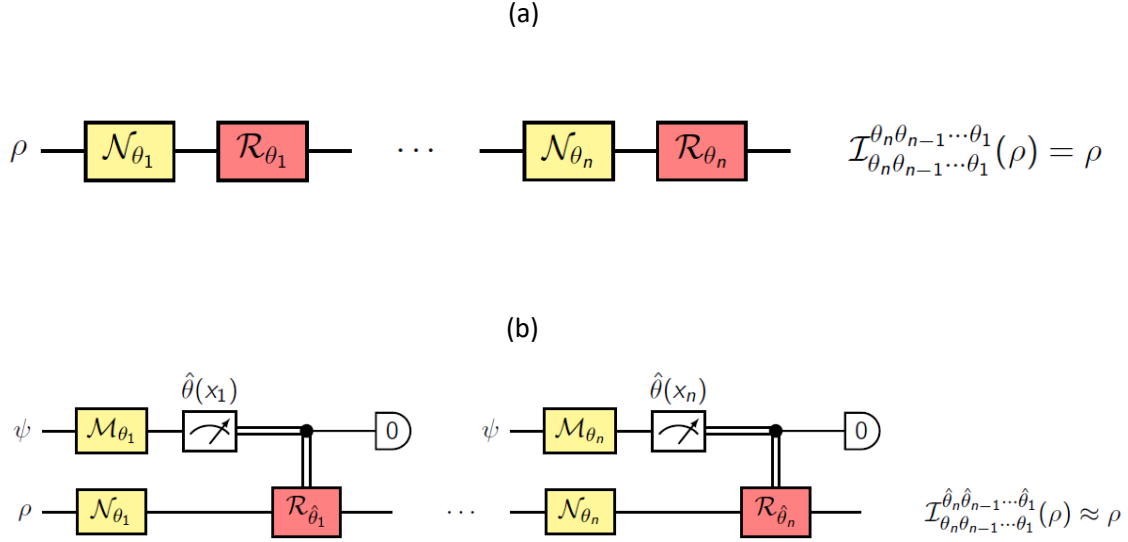


Figure 8: Time flows from left to right. (a) Multi-cycle recovery protocol with perfect knowledge. (b) Multi-cycle recovery protocol with incomplete knowledge. The input state of the quantum memory in both subfigures is given by  $\rho$ , whereas the input state of the spectator system in subfigure (b) is  $\psi$ . In the latter case, the state of the spectator is recycled back to  $\psi$  after every recovery cycle, via a discarding and preparation channel. The final output states  $\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\theta_n \theta_{n-1} \dots \theta_1}(\rho)$  and  $\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)$  of the quantum memory follow the multi-cycle notation in Eq. (98).

once. However, extensions to the multi-cycle regime are also important for real-time applications. A thorough study of the multi-cycle case is beyond the scope of the current article. However, here we present some useful bounds to stimulate future discussions.

To start, consider a stroboscopically varying noise parameter  $\theta$

$$\theta_1 \rightarrow \theta_2 \rightarrow \dots \rightarrow \theta_n, \quad (96)$$

where  $n$  enumerates the recovery cycle in the multi-cycle protocol, executed in the time interval  $[(n-1)\Delta t_{\mathcal{R}}, n\Delta t_{\mathcal{R}}]$ , where  $\Delta t_{\mathcal{R}} \ll \tau_{\theta}$  is the duration of a single recovery cycle, and  $\tau_{\theta}$  is the characteristic time of the noise parameter  $\theta$  (i.e. the expected time in which the value of  $\theta$  will change appreciably, see Section 4.1). The corresponding set of real-time spectator estimates of  $\theta$  for these  $n$  cycles is given by

$$\hat{\theta}_1 \rightarrow \hat{\theta}_2 \rightarrow \dots \rightarrow \hat{\theta}_n. \quad (97)$$

We introduce the following shorthand notation for multi-cycle recovery protocols

$$\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1} := \mathcal{R}_{\hat{\theta}_n} \circ \mathcal{N}_{\theta_n} \circ \mathcal{R}_{\hat{\theta}_{n-1}} \circ \mathcal{N}_{\theta_{n-1}} \dots \circ \mathcal{R}_{\hat{\theta}_1} \circ \mathcal{N}_{\theta_1}, \quad (98)$$

which is an  $n$ -cycle concatenation between the noisy channel (with changing noise parameter values in each timestep) and the corresponding best-guess recovery.

## 6.2 Recurrence Inequalities for Composite Average Channel Fidelity

So far, we have found a lower bound on the desired distinguishability measure in terms of the composite channel entanglement fidelity for the single-cycle case. To extend to the multi-cycle scenario, one option is to consider the entanglement fidelity of  $\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}$ . However, a more insightful approach is to express this entanglement fidelity in terms of individual cycle fidelities. Specifically, this is accomplished by the use of the entanglement fidelities of  $\mathcal{I}_{\theta_{n-1} \dots \theta_1}^{\hat{\theta}_{n-1} \dots \hat{\theta}_1}$  and  $\mathcal{I}_{\theta_n}^{\hat{\theta}_n}$ . Bounding composite channel fidelities using individual channel fidelities has been studied previously in e.g. [43]. The following lemma is largely taken from [43], using the  $\chi$ -matrix representation of quantum dynamics (see Appendix I for a self-contained review).

**Lemma 3.** ([43]) *Given the  $\chi$ -matrix elements  $\chi_{00}^{\mathcal{Q}}$ ,  $\chi_{00}^{\mathcal{S}}$  of the channels  $\mathcal{Q}$ ,  $\mathcal{S}$ , respectively, the composite channel  $\mathcal{S} \circ \mathcal{Q}$   $\chi$ -matrix element  $\chi_{00}^{\mathcal{S} \circ \mathcal{Q}}$  is bounded from above (and hence the corresponding error angle  $\delta^{\mathcal{S} \circ \mathcal{Q}}$  is bounded from below), as follows*

$$\frac{\chi_{00}^{\mathcal{S} \circ \mathcal{Q}}}{d} \leq \cos^2 \left( \arccos \sqrt{\frac{\chi_{00}^{\mathcal{S}}}{d}} - \arccos \sqrt{\frac{\chi_{00}^{\mathcal{Q}}}{d}} \right), \quad (99)$$

or more simply

$$\delta^{\mathcal{S} \circ \mathcal{Q}} \geq |\delta^{\mathcal{S}} - \delta^{\mathcal{Q}}|. \quad (100)$$

The inequality is saturated iff  $v_{ij} = 1$ ,  $\phi_i^{\mathcal{Q}} = \phi_{i'}^{\mathcal{Q}}$ , and  $\phi_j^{\mathcal{S}} = \phi_{j'}^{\mathcal{S}}$  for all  $i, i' = 1, \dots, K(\mathcal{Q})$  and  $j, j' = 1, \dots, K(\mathcal{S})$ . The quantities  $v_{ij}$ ,  $\phi_i^{\mathcal{Q}}$ , and  $\phi_j^{\mathcal{S}}$  are defined in Appendix J in terms of the Kraus operators of  $\mathcal{Q}$ ,  $\mathcal{S}$  and the  $d^2$  matrix basis elements of  $\mathcal{L}(\mathcal{H})$ .

For completeness, the proof of this lemma is found in Appendix J.

Let us denote by  $\mathcal{Q} \equiv \mathcal{I}_{\theta_{n-1}^{\hat{\theta}_1} \dots \theta_1^{\hat{\theta}_1}}$  and  $\mathcal{S} \equiv \mathcal{I}_{\theta_n^{\hat{\theta}_n}}$ , such that  $\mathcal{S} \circ \mathcal{Q} = \mathcal{I}_{\theta_n^{\hat{\theta}_n} \theta_{n-1}^{\hat{\theta}_{n-1}} \dots \theta_1^{\hat{\theta}_1}}$ . We further use the notation  $\chi_{00}^{1 \rightarrow n}$ ,  $\chi_{00}^{1 \rightarrow (n-1)}$ ,  $\chi_{00}^n$ ,  $\delta^{1 \rightarrow n}$ ,  $\delta^{1 \rightarrow (n-1)}$ , and  $\delta^n$  to replace  $\chi_{00}^{\mathcal{S} \circ \mathcal{Q}}$ ,  $\chi_{00}^{\mathcal{Q}}$ ,  $\chi_{00}^{\mathcal{S}}$ ,  $\delta^{\mathcal{S} \circ \mathcal{Q}}$ ,  $\delta^{\mathcal{Q}}$ , and  $\delta^{\mathcal{S}}$ , respectively. We also use the definition in Eq. (14) to write similar notations for the entanglement fidelities  $F_e^{1 \rightarrow n}$ ,  $F_e^{1 \rightarrow (n-1)}$ , and  $F_e^n$ , in terms of  $\delta^{1 \rightarrow n}$ ,  $\delta^{1 \rightarrow (n-1)}$ , and  $\delta^n$ . Therefore, we can reframe Lemma 3 by the authors of [43] as the following set of *recurrence inequalities* in the context of spectator-based recovery:

**Lemma 4.** *Given the entanglement fidelities  $F_e^i$  of the single-cycle recovery protocols at each time step  $\Delta t_{\mathcal{R}} = t_{i+1} - t_i$ , the  $n$ -cycle entanglement fidelity  $F_e^{1 \rightarrow n}$  of the multi-cycle recovery protocol is bounded from above by the  $(n-1)$ -cycle entanglement fidelity  $F_e^{1 \rightarrow (n-1)}$ , as*

$$F_e^{1 \rightarrow n} \leq \cos^2 \left( \arccos \sqrt{F_e^{1 \rightarrow (n-1)}} - \arccos \sqrt{F_e^n} \right), \quad (101)$$

or equivalently,

$$\delta^{1 \rightarrow n} \geq |\delta^{1 \rightarrow (n-1)} - \delta^n|. \quad (102)$$

The necessary and sufficient conditions for the saturation of this inequality are identical to that of Lemma 3.

**Remark 4.** *As noted in [43], the entanglement fidelity of a composite channel exhibits “constructive” and “destructive interference” with respect to the individual channel entanglement fidelities. In our case, we view the  $n$ -cycle recovery as a composite channel, where the individual channels are the  $(n-1)$ -cycle recovery and the  $n$ -th timestep recovery. Therefore, the same phenomenon of constructive and destructive interference applies here. This is purely a multi-cycle recovery phenomenon that is not present in single-cycle recovery case, which has been the main focus of modern literature in QEC.*

### 6.3 Contribution of The Spectator System

To identify the contribution of the lack of complete knowledge of  $\theta$  to the recurrence inequalities, let us consider the error angle

$$\hat{\delta} \equiv \delta^{\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_{\theta}} := \arccos \sqrt{F_e(\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_{\theta})}. \quad (103)$$

**Theorem 3.** *Given the multi-cycle entanglement fidelity  $F_e^{1 \rightarrow (n-1)}$  from the previous  $n-1$  cycles, the*

*contribution of the spectator system to the upper bound of the total  $n$ -cycle entanglement fidelity  $F_e^{1 \rightarrow n}$  is given by*

$$h(\theta_n, F_e^{1 \rightarrow (n-1)}) \text{Var}(\hat{\theta}_n), \quad (104)$$

where

$$h := \frac{g(\theta_n) \sin(2\delta^{1 \rightarrow (n-1)} - 2 \arccos \sqrt{F_e^n})}{2\sqrt{F_e^n(1-F_e^n)}}, \quad (105)$$

with

$$g(\theta_n) := -\frac{1}{2} \left( \frac{d^2}{d\nu^2} F_e(\mathcal{R}_{\theta_n + \nu} \circ \mathcal{N}_{\theta_n}) \right) \Big|_{\nu=0}. \quad (106)$$

The proof of this theorem is found in Appendix K.

## 6.4 Application to [4,1] Code of The Amplitude-Damping Channel

The contribution of the spectator system in multi-cycle bounds can also be computed explicitly for the [4,1] code of the AD channel. For a fixed value of the entanglement fidelity  $F_e^{1 \rightarrow (n-1)}$  (or equivalently,  $\delta^{1 \rightarrow (n-1)}$ ) at the  $(n-1)$ -th step, we can plot the total upper bound in the case of both complete and incomplete knowledge. The simplest case where the spectator system’s parameters are  $\gamma = 1$  and  $m = 1$  is shown in Fig. 9.

Note that, although we expect the incomplete knowledge about the noise parameter to deteriorate the allowed values of the entanglement fidelity (as we have shown for single-cycle QEC of the AD channel in Fig. 7(b)), in the multi-cycle scenario, this can play to our advantage due to the coherence between the accumulated error during the prior  $(n-1)$  cycles and the error due to the limited knowledge about the noise parameter at the  $n$ -th cycle (see Remark 4). This observation further supports the potential superiority of spectator-based recovery techniques in maintaining real-time quantum memories.

## 7 Comparison With Previous Literature

### 7.1 Relation to Quantum Information-Theoretic Protocols

In this article, we focused on the diamond distance due to its operational meaning in terms of a quantum channel discrimination task. In this context, Lemma 2 could be interpreted as a fundamental bound on the success probability of such a task. A similar bound has already been derived by Pirandola *et al.* in [41] using port-based teleportation [87]. In fact, the bound in [41] is valid for general adaptive protocols.

Furthermore, as current techniques of quantum control have matured, the applicability of Lemma 4

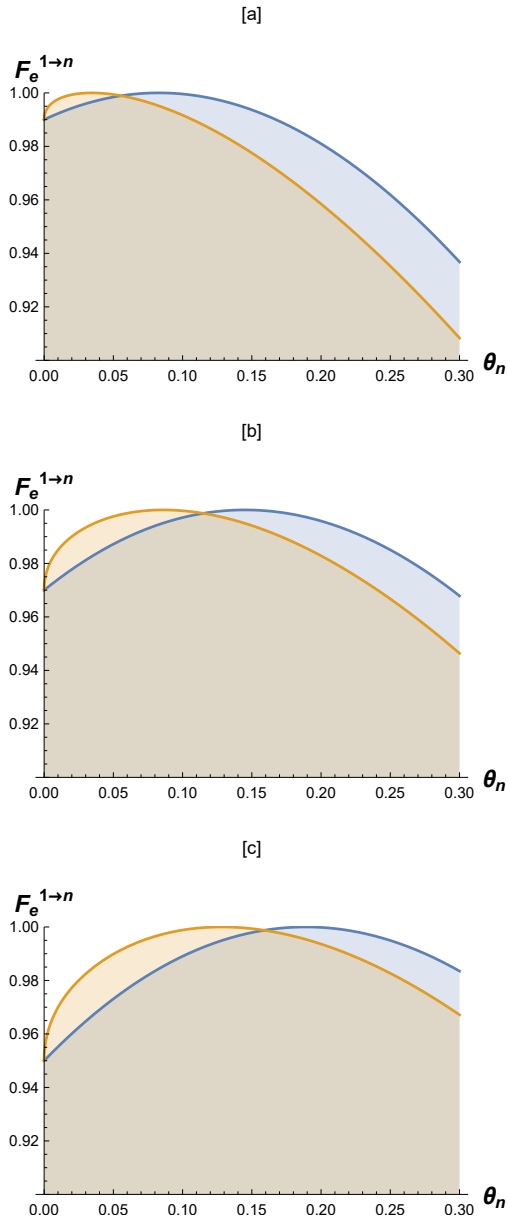


Figure 9: Subplots show the dependence of the accumulated  $n$ -cycle entanglement fidelity  $F_e^{1 \rightarrow n}$  on the value of the noise parameter  $\theta_n$  at the  $n$ -th cycle for the [4,1] code of the amplitude-damping channel. The spectator system is taken to have the simplest characteristic parameters ( $\gamma = 1, m = 1$ ). The colored regions indicate allowed values for the entanglement fidelity. The blue color refers to the case of perfect knowledge of  $\theta_n$  and the orange color to the lack of that knowledge. From top to bottom, the value of the accumulated  $(n - 1)$ -cycle entanglement fidelity  $F_e^{1 \rightarrow (n-1)}$  is picked to be (a) 0.99, (b) 0.97, and (c) 0.95, respectively.

is not only confined to multiple recovery rounds, as demonstrated experimentally in e.g. [44]. It can also be applied in various quantum information-theoretic tasks where multiple calls to the noisy channel and adaptive feedback are allowed, such as quantum channel discrimination with adaptive feedback [41, 88, 89].

## 7.2 Relation to Robustness of Channel-adapted QEC

Our approach to recovery with incomplete knowledge is closely related to the robustness of channel-adapted QEC codes studied previously in literature [57–60]. To elaborate, since QEC codes are designed to correct the most likely errors, an important question to ask is: how resilient (robust) is the designed QEC code with respect to some arbitrary mixing with the next-most likely errors? The authors of [58] have framed the robustness problem such that it applies both for Pauli and non-Pauli channels, as follows: One first finds the optimum recovery  $\mathcal{R}$  of a channel  $\mathcal{N}$  (the most likely noise) by maximizing the entanglement fidelity of  $\mathcal{R} \circ \mathcal{N}$ , and then one mixes the original channel  $\mathcal{N}$  with some other channel  $\mathcal{N}'$  (the next-most likely noise) by taking their convex combination, i.e.  $\mathcal{N}_\mu := (1 - \mu)\mathcal{N} + \mu\mathcal{N}'$  for some mixing parameter  $\mu \in [0, 1]$ . Then, the robustness of the recovery  $\mathcal{R}$  with respect to  $\mu$  is found by considering the entanglement fidelity of  $\mathcal{R} \circ \mathcal{N}_\mu$  and observing if it has major variations as a function of the mixing parameter  $\mu$ . This setup shares some similarities with our approach, however, it has a different quantity of interest, namely the *sensitivity* of entanglement fidelity with respect to changes in the mixing parameter, quantified as the first derivative with respect to  $\mu$  of

$$F_e(\mathcal{R}_\mu \circ \mathcal{N}_\mu) - F_e(\mathcal{R} \circ \mathcal{N}_\mu), \quad (107)$$

where  $\mathcal{R}_\mu$  is the optimum recovery of the mixing  $\mathcal{N}_\mu$  (Also see Appendix F.2 for bounds on a similar quantity). This is to be contrasted with the quantity of interest in this article (using the parameter notation  $\mu$ )

$$F_e(\mathcal{R} \circ \mathcal{N}) - F_e(\mathcal{R}_\mu \circ \mathcal{N}). \quad (108)$$

Here,  $\mu$  plays the role of the uncertainty  $\nu \equiv \hat{\theta} - \theta$  in the noise parameter  $\theta$ , and therefore it has a different interpretation. Namely, there is no next-most likely noise in this description! Instead,  $\mu$  is the random variable describing the uncertainty in the environment noise parameter and has a finite variance, by the QCRB.

The possibility of including the channel uncertainty as a probability distribution  $p(\mu)$  in the optimization problem of entanglement fidelity has been discussed by Fletcher in [90]. The question then, as mentioned in [58], is: how to pick a physical probability distribution  $p(\mu)$ ? In our picture (spectator-based QEC), this question has a relatively simple answer, as one should always pick the probability distribution that maximizes the Shannon entropy with a fixed expectation and variance (larger or equal to the inverse of the quantum Fisher information). Such a probability distribution is called “the truncated normal distribution”.

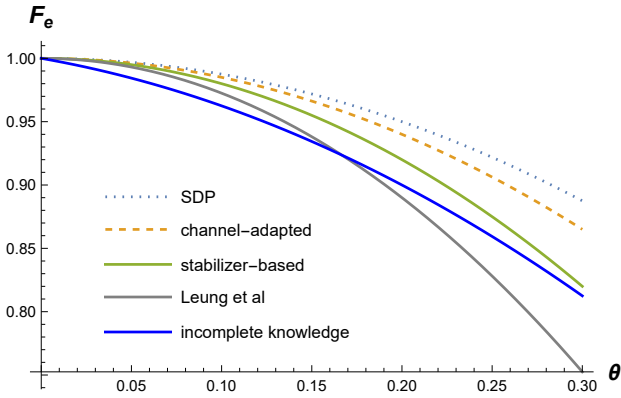


Figure 10: Comparison of the maximum performance of spectator-based recovery (subject to varying noise parameter  $\theta$ ) with various recovery optimization approaches to the  $[4, 1]$  code of the amplitude-damping channel (with fixed noise parameter  $\theta$ ). We consider the “worst case” spectator parameters ( $\gamma = 1, m = 1$ ). Shown are the performances of the well-known approximate QEC code in Leung *et al.* [13], its channel-adapted version by Fletcher *et al.* [12], its SDP optimized version by Fletcher *et al.* [14], its stabilizer-based version [12, 81], and the incomplete knowledge extension of the channel-adapted QEC in [12]. Here, the difference between the “channel-adapted” and “incomplete knowledge” entanglement fidelities showcases the fundamental metrological cost of operating a real-time quantum memory. All other ( $\gamma \geq 1, m \geq 1$ ) spectator-based recoveries lie above the “incomplete knowledge” graph.

### 7.3 Relation to $[4, 1]$ AD Code Literature

As amplitude-damping (qubit decoherence) is one of the most common noises in quantum systems, developing QEC codes for this particular noise has been a major focus of QEC literature since its inception in 1995. The simplest of such QEC codes is the approximate  $[4, 1]$  code [13]. Since then, QEC methods for the AD noise have been developing in sophistication by using various new techniques, such as channel adaptation [12], stabilizer formalism [81], and semi-definite programming [14]. These techniques have been steadily improving upon the entanglement fidelity of the original  $[4, 1]$  code in [13]. However, if we want to implement these techniques for real-time quantum memories (where the decoherence parameter is slowly varying in time), how much of the improvements upon [13] obtained in the last two decades are we likely to retain? The answer to this question, we compare the performance of the  $[4, 1]$  code in the incomplete knowledge scenario with previous literature. As spectator systems are characterized by their physical nature  $\gamma \geq 1$  and the number of independent subsystems  $m \in \mathbb{N}_+$  (see Eq. (45)), the answer will vary from one physical implementation to another. However, we consider the above question in the case ( $\gamma = 1, m = 1$ ). The results of this comparison are summarized in Fig. 10 and the table below

Previous literature	$F_e$ to $\mathcal{O}(\theta^3)$ order
Leung <i>et al.</i> [13]	$1 - 2.75\theta^2$
Stabilizer-Based [81, 90]	$1 - 2\theta^2$
Channel-Adapted [12, 79]	$1 - 1.5\theta^2$
SDP [14]	$1 - 1.25\theta^2$
Incomplete Knowledge	$1 - 0.25\theta - 1.25\theta^2$

Table 1: Comparison between the entanglement fidelities of the  $[4, 1]$  code for small noise parameter value  $\theta$ , for different recovery protocols (here SDP stands for “semi-definite programming”). Note that in the incomplete knowledge scenario, the leading error term in the entanglement fidelity of recovery is linear in  $\theta$ , as opposed to quadratic, which is the optimal result when the noise parameter  $\theta$  is known *a priori*.

We observe that, due to incomplete knowledge of the noise parameter, the  $[4, 1]$  code of the AD channel performs suboptimally to [13] for noise parameter values below a certain threshold  $\theta \leq 0.17$ . However, beyond that point, the improvements introduced by channel-adapted recoveries and semi-definite programming techniques are preserved, as they still outperform [13], even in the presence of incomplete knowledge about  $\theta$ . Furthermore, the range of the values of  $\theta \in [0, 1]$  for which this outperformance is preserved gets larger the larger we pick  $\gamma$  and/or  $m$  (see Fig. 7(a)).

Let us consider one final observation. We noted in Fig. 7 that different values of the spectator parameter  $\gamma$  yield different regions of  $\theta$  where channel-adapted and semi-definite programming techniques in QEC maintain their improvements upon the approximate  $[4, 1]$  code [13], provided that a spectator system is implemented in the incomplete-knowledge recovery protocol. One might observe that, since the value of  $\gamma$  in Eq. (48) generally depends on the couplings of the spectator and memory systems with the environment, the only way to change  $\gamma$  is to change the physical implementation of at least one of these systems. However, recent quantum control techniques, such as Hamiltonian amplification [91], allow for the tuning of the coupling strengths between an environment and any continuous variable quantum system, with a quadratic coupling Hamiltonian. Therefore, provided that the implementation of either the spectator or memory system has continuous degrees of freedom [92], the Hamiltonian amplification technique yields a practical advantage for spectator-based recovery architectures, as the resulting entanglement fidelities can be manipulated in experiments for any desired region of the noise parameter  $\theta$ , as seen in Fig. 7(a).

### 7.4 Relation to Time-Dependent QEC

In [93], the author suggests that knowledge of the error rates for Pauli channels is not the most useful side information in QEC. Indeed, as mentioned pre-

viously (see Remark 2), the assumption that the *optimal* recovery channel (defined by Eq. (40)) depends on the environment noise parameter does not hold for Pauli channels. Nevertheless, it is important to note that optimization-based techniques of QEC for Pauli channels do generally benefit from the knowledge of the noise parameter. This is especially relevant when error identification from syndrome measurements is not unique (e.g. in surface codes [94]). Hence, one can only construct (suboptimal) decoders, rather than the optimal recovery map in Eq. (40). This generally yields decoders that depend on the noise parameters, even for Pauli channels. For example, various types of decoders exist for both repetition codes [95, 96] and surface codes [97], where under the presence of a drifting noise parameter, one can design an adaptive decoder that can track this drift while not interrupting the QEC protocol. Therefore, the results of this article could be expanded to include adaptive decoders for repetition and surface codes, rather than the optimal recovery map defined in Eq. (40). Finally, it is worth mentioning that other approaches to adaptation in QEC literature have been pursued, e.g. in Refs. [98, 99].

## 8 Conclusion and Open Questions

In this article, I consider the problem of building a real-time (drift-adapting) quantum memory and present it as a spectator-based recovery protocol. To counter noise drift, the spectator system performs a real-time parameter estimation (generally in the presence of nuisance parameters) and feeds forward this classical side information to the “best-guess” recovery map. To quantify the single-cycle information-theoretic cost of adaptation in real-time quantum memories, I compute a lower bound for the diamond distance between the optimal (inaccessible) and best-guess (accessible) recovery protocols. This approach is generalized in Appendix A for other relevant distinguishability measures between arbitrary two quantum channels. For slowly drifting noise parameters, I show that a metrological bound exists, determined by the quantum Fisher information of the spectator dynamics. This bound is demonstrated for the [4,1] code of the amplitude-damping channel, and the effects of various physical choices of spectator qubits and nuisance parameters are discussed. Finally, for multi-cycle recovery, I recall a theorem in [43] and use it to derive an upper bound to the fidelity of multi-cycle recovery in terms of recurrence inequalities. The contribution of the lack of knowledge of the noise parameters (i.e. noise-drift adaptation) is also derived. This is also showcased for the [4,1] code of the amplitude-damping channel, and regions of out-performance in the spectator-based recovery protocols are highlighted. The advantages of spectator-based recovery compared to non-adaptive recovery proto-

cols, even in the perfect knowledge scenario, is due to the coherence of errors from different cycle numbers as well as the imperfect knowledge (noise estimation) error. This phenomenon is exclusive to multi-cycle QEC.

The results mentioned above are relevant for various research communities, such as quantum error correction, quantum communication, quantum information, quantum control, and quantum computing. To elaborate, the existence of lower bounds on any channel recovery (Eqs. (53) and (54), or more generally in Theorem 4 in Appendix A) could be valuable in testing the performance of various optimization-based techniques in QEC to determine if optimal performance is reached. As discussed in Section 7, these bounds may also have a broader interest in various domains of quantum information as they hold for any generalized distinguishability measure and between any two quantum channels (see Appendix A). Multi-cycle bounds (Lemma 4) might also be useful in adaptive quantum information-theoretic protocols, where many calls to the noisy channel are made. The analysis made for the [4, 1] code of the amplitude-damping channel sheds light on what to expect when implementing such QEC codes in real-time quantum memories [23, 24], while also providing an excited avenue in terms of out-performance in the incomplete knowledge scenario for multi-cycle recovery, which is quickly starting to become a reality [44]. Finally, implementing novel quantum control techniques, such as Hamiltonian amplification for continuous quantum systems [91], might prove useful in controlling the coupling strength of the spectator system. Therefore, one can experimentally optimize over the selection of all possible spectator system parameters without physically changing the spectator system.

Many questions are left open:

- Extension of the information-theoretic and metrological lower bounds to Pauli channels with suboptimal decoders. This is relevant for current surface codes, as the optimal recovery map Eq. (40) is inaccessible, due to the probabilistic nature of error identification from syndrome measurements for high error rates [94].
- Incorporation of important theoretical techniques of the maximum overlap problem in quantum information theory, e.g. the two-sided bounds by Tyson [100, 101] using directional iterates. This approach has led to various important results previously [49, 102]. An interesting proposal (potentially also in quantum metrology) would be to apply the directional iterate technique to the semi-inner product used to define the QFIM, given in Eq. (196) of Appendix E.
- A deeper analysis of the multi-cycle regime is required. This includes, but is not limited to, the study of optimal conditions for achieving the

coherent error cancellation, as well as incorporating techniques from asymptotic quantum information theory to gain further insight into the multi-cycle case.

- Continuous (dynamical) recovery using Petz recovery maps [56] could also be considered, as well as other adaptive approaches implementing Petz recovery maps [103, 104]. This can potentially extend the temporal range of applicability of spectator-based recovery protocols to faster varying noise. Another interesting dynamical model of real-time quantum memory could be constructed from the open system theory of two subsystems (memory and spectator) with slow and fast dynamics, relative to the environment noise. This has been studied previously in the context of adiabatic elimination in bipartite open quantum systems [105].
- Considerations of spatial variability of the noise parameter are also needed for scalability of the spectator-based recovery protocols [25, 106]. Generalization beyond the independent noise model in Eq. (45) is also of relevance.
- Finally, one would be remiss by not considering the large literature on approximate recoverability of quantum states, see e.g. [50, 51, 107–109] and references within. It is interesting to see whether incomplete knowledge recovery protocols would benefit from a similar approach.

## 9 Acknowledgements

I am grateful to Mark M. Wilde, Hwang Lee, Stav Haldar, Kenneth Brown, Milad Marvian, Todd Brun, and the anonymous referees for helpful comments and stimulating discussions. I am also grateful for the early discussions with Lorenza Viola regarding stroboscopically changing noise parameters, which motivated certain parts of this work. I would like to acknowledge the support of Kenneth R. Brown, Mark M. Wilde, Hwang Lee, and the Coates Research Award by the Department of Physics and Astronomy at Louisiana State University. This work was supported by the U.S. Army Research Office through the U.S. MURI Grant No. W911NF-18-1-0218.

## References

[1] Mustafa Gündoğan, Jasminder S Sidhu, Victoria Henderson, Luca Mazzarella, Janik Wolters, Daniel KL Oi, and Markus Krutzik. “Proposal for space-borne quantum memories for global quantum networking”. *npj Quantum Information* **7**, 128 (2021).

[2] Julius Wallnöfer, Frederik Hahn, Mustafa Gündoğan, Jasminder S Sidhu, Fabian Wiesner, Nathan Walk, Jens Eisert, and Janik Wolters. “Simulating quantum repeater strategies for multiple satellites”. *Communications Physics* **5**, 169 (2022).

[3] Mustafa Gündoğan, Jasminder J Sidhu, Daniel KL Oi, and Markus Krutzik. “Time-delayed single quantum repeater node for global quantum communications with a single satellite” (2023).

[4] Jasminder S Sidhu, Yingkai Ouyang, Earl T Campbell, and Pieter Kok. “Tight bounds on the simultaneous estimation of incompatible parameters”. *Physical Review X* **11**, 011028 (2021).

[5] Ye Wang, Mark Um, Junhua Zhang, Shuoming An, Ming Lyu, Jing-Ning Zhang, L-M Duan, Dahyun Yum, and Kihwan Kim. “Single-qubit quantum memory exceeding ten-minute coherence time”. *Nature Photonics* **11**, 646–650 (2017).

[6] P-J Stas, Yan Qi Huan, Bartholomeus Machiels, Erik N Knall, Aziza Suleymanzade, Benjamin Pingault, Madison Sutula, Sophie W Ding, Can M Knaut, Daniel R Assumpcao, et al. “Robust multi-qubit quantum network node with integrated error detection”. *Science* **378**, 557–560 (2022).

[7] Benjamin Schumacher and Michael A Nielsen. “Quantum data processing and error correction”. *Physical Review A* **54**, 2629 (1996).

[8] Benjamin Schumacher. “Sending entanglement through noisy quantum channels”. *Physical Review A* **54**, 2614 (1996).

[9] Emanuel Knill and Raymond Laflamme. “Theory of quantum error-correcting codes”. *Physical Review A* **55**, 900 (1997).

[10] Lorenza Viola and Seth Lloyd. “Dynamical suppression of decoherence in two-state quantum systems”. *Physical Review A* **58**, 2733 (1998).

[11] Lorenza Viola, Emanuel Knill, and Seth Lloyd. “Dynamical decoupling of open quantum systems”. *Physical Review Letters* **82**, 2417 (1999).

[12] Andrew S Fletcher, Peter W Shor, and Moe Z Win. “Channel-adapted quantum error correction for the amplitude damping channel”. *IEEE Transactions on Information Theory* **54**, 5705–5718 (2008).

[13] Debbie W Leung, Michael A Nielsen, Isaac L Chuang, and Yoshihisa Yamamoto. “Approximate quantum error correction can lead to better codes”. *Physical Review A* **56**, 2567 (1997).

[14] Andrew S Fletcher, Peter W Shor, and Moe Z Win. “Optimum quantum error recovery using semidefinite programming”. *Physical Review A* **75**, 012338 (2007).

[15] Heinz-Peter Breuer, Francesco Petruccione,

- et al. “The theory of open quantum systems”. Oxford University Press on Demand. (2002).
- [16] Clemens Müller, Jürgen Lisenfeld, Alexander Shnirman, and Stefano Poletto. “Interacting two-level defects as sources of fluctuating high-frequency noise in superconducting circuits”. *Physical Review B* **92**, 035442 (2015).
- [17] PV Klimov, Julian Kelly, Z Chen, Matthew Neeley, Anthony Megrant, Brian Burkett, Rami Barends, Kunal Arya, Ben Chiaro, Yu Chen, et al. “Fluctuations of energy-relaxation times in superconducting qubits”. *Physical review letters* **121**, 090502 (2018).
- [18] Josu Etxezarreta Martinez, Patricio Fuentes, Pedro Crespo, and Javier Garcia-Frias. “Time-varying quantum channel models for superconducting qubits”. *npj Quantum Information* **7**, 1–10 (2021).
- [19] Samudra Dasgupta and Travis S Humble. “Characterizing the stability of nisq devices”. In 2020 IEEE International Conference on Quantum Computing and Engineering (QCE). Pages 419–429. IEEE (2020).
- [20] Samudra Dasgupta and Travis S Humble. “Stability of noisy quantum computing devices” (2021).
- [21] Cristian Bonato and Dominic W Berry. “Adaptive tracking of a time-varying field with a quantum sensor”. *Physical Review A* **95**, 052348 (2017).
- [22] Luis Cortez, Areeya Chantasri, Luis Pedro García-Pintos, Justin Dressel, and Andrew N Jordan. “Rapid estimation of drifting parameters in continuously measured quantum systems”. *Physical Review A* **95**, 012314 (2017).
- [23] Timothy Proctor, Melissa Revelle, Erik Nielsen, Kenneth Rudinger, Daniel Lobser, Peter Maunz, Robin Blume-Kohout, and Kevin Young. “Detecting and tracking drift in quantum information processors”. *Nature communications* **11**, 1–9 (2020).
- [24] Swarnadeep Majumder, Leonardo Andreta de Castro, and Kenneth R Brown. “Real-time calibration with spectator qubits”. *npj Quantum Information* **6**, 1–9 (2020).
- [25] Riddhi Swaroop Gupta, Luke CG Govia, and Michael J Biercuk. “Integration of spectator qubits into quantum computer architectures for hardware tune-up and calibration”. *Physical Review A* **102**, 042611 (2020).
- [26] Akram Youssry, Gerardo A Paz-Silva, and Christopher Ferrie. “Noise detection with spectator qubits and quantum feature engineering”. *New Journal of Physics* **25**, 073004 (2023).
- [27] Arshag Danageozian, Ashe Miller, Pratik J Barge, Narayan Bhusal, and Jonathan P Dowling. “Noisy coherent population trapping: applications to noise estimation and qubit state preparation”. *Journal of Physics B: Atomic, Molecular and Optical Physics* **55**, 155503 (2022).
- [28] Hongting Song, Areeya Chantasri, Behnam Tonekaboni, and Howard M Wiseman. “Optimized mitigation of random-telegraph-noise dephasing by spectator-qubit sensing and control”. *Physical Review A* **107**, L030601 (2023).
- [29] Behnam Tonekaboni, Areeya Chantasri, Hongting Song, Yanan Liu, and Howard M Wiseman. “Greedy versus map-based optimized adaptive algorithms for random-telegraph-noise mitigation by spectator qubits”. *Physical Review A* **107**, 032401 (2023).
- [30] K Singh, CE Bradley, S Anand, V Ramesh, R White, and H Bernien. “Mid-circuit correction of correlated phase errors using an array of spectator qubits”. *Science* **380**, eade5337 (2023).
- [31] Bas Hensen, Hannes Bernien, Anaïs E Dréau, Andreas Reiserer, Norbert Kalb, Machiel S Blok, Just Ruitenbergh, Raymond FL Vermeulen, Raymond N Schouten, Carlos Abellán, et al. “Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres”. *Nature* **526**, 682–686 (2015).
- [32] Emre Togan, Yiwen Chu, Alexei S Trifonov, Liang Jiang, Jeronimo Maze, Lilian Childress, MV Gurudev Dutt, Anders Søndberg Sørensen, Phillip R Hemmer, Alexander S Zibrov, et al. “Quantum entanglement between an optical photon and a solid-state spin qubit”. *Nature* **466**, 730–734 (2010).
- [33] D Andrew Golter and Hailin Wang. “Optically driven rabi oscillations and adiabatic passage of single electron spins in diamond”. *Physical review letters* **112**, 116403 (2014).
- [34] Shu-Hao Wu, Ethan Turner, and Hailin Wang. “Continuous real-time sensing with a nitrogen-vacancy center via coherent population trapping”. *Physical Review A* **103**, 042607 (2021).
- [35] Ethan Turner, Shu-Hao Wu, Xinzhu Li, and Hailin Wang. “Real-time magnetometry with coherent population trapping in a nitrogen-vacancy center”. *Physical Review A* **105**, L010601 (2022).
- [36] Ignas Lekavicius, D Andrew Golter, Thein Oo, and Hailin Wang. “Transfer of phase information between microwave and optical fields via an electron spin”. *Physical Review Letters* **119**, 063601 (2017).
- [37] Michael A Nielsen and Isaac Chuang. “Quantum computation and quantum information: 10th anniversary edition” (2002).
- [38] Samuel L Braunstein and Carlton M Caves. “Statistical distance and the geometry of quantum states”. *Physical Review Letters* **72**, 3439 (1994).

- [39] Samuel L Braunstein, Carlton M Caves, and Gerard J Milburn. “Generalized uncertainty relations: theory, examples, and lorentz invariance”. *annals of physics* **247**, 135–173 (1996).
- [40] Yingkai Ouyang, Kaumudibikash Goswami, Jacqueline Romero, Barry C. Sanders, Min-Hsiu Hsieh, and Marco Tomamichel. “Approximate reconstructability of quantum states and noisy quantum secret sharing schemes”. *Phys. Rev. A* **108**, 012425 (2023).
- [41] Stefano Pirandola, Riccardo Laurenza, Cosmo Lupo, and Jason L Pereira. “Fundamental limits to quantum channel discrimination”. *npj Quantum Information* **5**, 1–8 (2019).
- [42] Giuliano Benenti and Giuliano Strini. “Computing the distance between quantum channels: usefulness of the fano representation”. *Journal of Physics B: Atomic, Molecular and Optical Physics* **43**, 215508 (2010).
- [43] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson. “Bounding the average gate fidelity of composite channels using the unitarity”. *New Journal of Physics* **21**, 053016 (2019).
- [44] Julia Cramer, Norbert Kalb, M Adriaan Rol, Bas Hensen, Machiel S Blok, Matthew Markham, Daniel J Twitchen, Ronald Hanson, and Tim H Taminiau. “Repeated quantum error correction on a continuously encoded qubit by real-time feedback”. *Nature communications* **7**, 1–7 (2016).
- [45] Bharat Thotakura and Tzu-Chieh Wei. “Quantum state transfer: interplay between gate and readout errors”. *Quantum Information Processing* **22**, 275 (2023).
- [46] Michael Reimpell and Reinhard F Werner. “Iterative optimization of quantum error correcting codes”. *Physical review letters* **94**, 080501 (2005).
- [47] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. “General teleportation channel, singlet fraction, and quasidistillation”. *Physical Review A* **60**, 1888 (1999).
- [48] Aleksander Kubica and Rafał Demkowicz-Dobrzański. “Using quantum metrological bounds in quantum error correction: a simple proof of the approximate Eastin-Knill theorem”. *Physical Review Letters* **126**, 150503 (2021).
- [49] Cédric Bény and Ognjan Oreshkov. “General conditions for approximate quantum error correction and near-optimal recovery channels”. *Physical review letters* **104**, 120501 (2010).
- [50] Marius Junge, Renato Renner, David Sutter, Mark M Wilde, and Andreas Winter. “Universal recovery maps and approximate sufficiency of quantum relative entropy”. In *Annales Henri Poincaré*. Volume 19, pages 2955–2978. Springer (2018).
- [51] Francesco Buscemi, Siddhartha Das, and Mark M Wilde. “Approximate reversibility in the context of entropy gain, information gain, and complete positivity”. *Physical Review A* **93**, 062314 (2016).
- [52] Antonio Acín. “Statistical distinguishability between unitary operations”. *Physical review letters* **87**, 177901 (2001).
- [53] John Watrous. “Semidefinite programs for completely bounded norms” (2009).
- [54] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. “Exact and approximate unitary 2-designs and their application to fidelity estimation”. *Physical Review A* **80**, 012304 (2009).
- [55] Gilad Gour. “Comparison of quantum channels by superchannels”. *IEEE Transactions on Information Theory* **65**, 5880–5904 (2019).
- [56] Hyukjoon Kwon, Rick Mukherjee, and MS Kim. “Reversing lindblad dynamics via continuous petz recovery map”. *Physical Review Letters* **128**, 020403 (2022).
- [57] Robert L Kosut, Alireza Shabani, and Daniel A Lidar. “Robust quantum error correction via convex optimization”. *Physical review letters* **100**, 020502 (2008).
- [58] Gábor Balló and Péter Gurin. “Robustness of channel-adapted quantum error correction”. *Physical Review A* **80**, 012326 (2009).
- [59] Long Huang, Xiaohua Wu, and Tao Zhou. “Robustness of the concatenated quantum error-correction protocol against noise for channels affected by fluctuation”. *Physical Review A* **100**, 042321 (2019).
- [60] David Layden, Louisa Ruixue Huang, and Paola Cappellaro. “Robustness-optimized quantum error correction”. *Quantum Science and Technology* **5**, 025008 (2020).
- [61] Robert L Kosut and Daniel A Lidar. “Quantum error correction via convex optimization”. *Quantum Information Processing* **8**, 443–459 (2009).
- [62] Daniel Gottesman. “Stabilizer codes and quantum error correction”. *California Institute of Technology*. (1997).
- [63] Jun Suzuki, Yuxiang Yang, and Masahito Hayashi. “Quantum state estimation with nuisance parameters”. *Journal of Physics A: Mathematical and Theoretical* **53**, 453001 (2020).
- [64] Ole E Barndorff-Nielsen and David Roxbee Cox. “Inference and asymptotics”. Volume 13. Springer. (1994).
- [65] Jing Liu, Haidong Yuan, Xiao-Ming Lu, and Xiaoguang Wang. “Quantum fisher information matrix and multiparameter estimation”. *Journal of Physics A: Mathematical and Theoretical* **53**, 023001 (2020).
- [66] Sammy Ragy, Marcin Jarzyna, and Rafał Demkowicz-Dobrzański. “Compatibility in mul-



- tiparameter quantum metrology”. *Physical Review A* **94**, 052108 (2016).
- [67] Federico Belliardo and Vittorio Giovannetti. “Incompatibility in quantum parameter estimation”. *New Journal of Physics* **23**, 063055 (2021).
- [68] Xiao-Ming Lu and Xiaoguang Wang. “Incorporating heisenberg’s uncertainty principle into quantum multiparameter estimation”. *Physical Review Letters* **126**, 120503 (2021).
- [69] Francesco Albarelli, Jamie F Friel, and Animesh Datta. “Evaluating the holevo cramér-rao bound for multiparameter quantum metrology”. *Physical review letters* **123**, 200503 (2019).
- [70] Hiroshi Nagaoka. “A generalization of the simultaneous diagonalization of hermitian matrices and its relation to quantum estimation theory”. In *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*. Pages 133–149. World Scientific (2005).
- [71] Lorcán O Conlon, Jun Suzuki, Ping Koy Lam, and Syed M Assad. “Efficient computation of the nagaoka–hayashi bound for multiparameter estimation with separable measurements”. *npj Quantum Information* **7**, 110 (2021).
- [72] Masahito Hayashi and Yingkai Ouyang. “Tight cramér-rao type bounds for multiparameter quantum metrology through conic programming”. *Quantum* **7**, 1094 (2023).
- [73] Christian Arenz, Daniel Burgarth, and Robin Hillier. “Dynamical decoupling and homogenization of continuous variable systems”. *Journal of Physics A: Mathematical and Theoretical* **50**, 135303 (2017).
- [74] Chao Lei, Shijie Peng, Chenyong Ju, Man-Hong Yung, and Jiangfeng Du. “Decoherence control of nitrogen-vacancy centers”. *Scientific reports* **7**, 11937 (2017).
- [75] Tim Hugo Tamini, Julia Cramer, Toeno van der Sar, Viatcheslav V Dobrovitski, and Ronald Hanson. “Universal control and error correction in multi-qubit spin registers in diamond”. *Nature nanotechnology* **9**, 171–176 (2014).
- [76] Gerald Waldherr, Ya Wang, S Zaiser, M Jamali, T Schulte-Herbrüggen, H Abe, T Ohshima, J Isoya, JF Du, P Neumann, et al. “Quantum error correction in a solid-state hybrid spin register”. *Nature* **506**, 204–207 (2014).
- [77] Haidong Yuan and Chi-Hang Fred Fung. “Fidelity and fisher information on quantum channels”. *New Journal of Physics* **19**, 113039 (2017).
- [78] Jasminder S Sidhu and Pieter Kok. “Geometric perspective on quantum parameter estimation”. *AVS Quantum Science* **2**, 014701 (2020).
- [79] Yun Zhan and Xiao-Yu Chen. “Entanglement fidelity of channel adaptive quantum codes”. *Chinese Physics B* **22**, 010308 (2013).
- [80] Benjamin Rahn, Andrew C Doherty, and Hideo Mabuchi. “Exact performance of concatenated quantum codes”. *Physical Review A* **66**, 032304 (2002).
- [81] Carlo Cafaro and Peter van Loock. “Approximate quantum error correction for generalized amplitude-damping errors”. *Physical Review A* **89**, 022316 (2014).
- [82] Carlo Cafaro and Peter van Loock. “A simple comparative analysis of exact and approximate quantum error correction”. *Open Systems & Information Dynamics* **21**, 1450002 (2014).
- [83] Bryan Eastin and Emanuel Knill. “Restrictions on transversal encoded quantum gate sets”. *Physical review letters* **102**, 110502 (2009).
- [84] Vishal Katariya and Mark M Wilde. “Geometric distinguishability measures limit quantum channel estimation and discrimination”. *Quantum Information Processing* **20**, 1–170 (2021).
- [85] Akio Fujiwara. “Estimation of a generalized amplitude-damping channel”. *Physical Review A* **70**, 012317 (2004).
- [86] Leslie Allen and Joseph H Eberly. “Optical resonance and two-level atoms”. Volume 28. Courier Corporation. (1987).
- [87] Satoshi Ishizaka and Tohya Hiroshima. “Asymptotic teleportation scheme as a universal programmable quantum processor”. *Physical review letters* **101**, 240501 (2008).
- [88] Giulio Chiribella, Giacomo M D’Ariano, and Paolo Perinotti. “Memory effects in quantum channel discrimination”. *Physical review letters* **101**, 180501 (2008).
- [89] Masahito Hayashi. “Discrimination of two channels by adaptive methods and its application to quantum system”. *IEEE Transactions on Information Theory* **55**, 3807–3820 (2009).
- [90] Andrew S Fletcher. “Channel-adapted quantum error correction” (2007).
- [91] Christian Arenz, Denys I Bondar, Daniel Burgarth, Cecilia Cormick, and Herschel Rabitz. “Amplification of quadratic hamiltonians”. *Quantum* **4**, 271 (2020).
- [92] Daniel Gottesman, Alexei Kitaev, and John Preskill. “Encoding a qubit in an oscillator”. *Physical Review A* **64**, 012310 (2001).
- [93] Yuichiro Fujiwara. “Instantaneous quantum channel estimation during quantum information processing” (2014).
- [94] Austin G Fowler, Matteo Mariantoni, John M Martinis, and Andrew N Cleland. “Surface codes: Towards practical large-scale quantum computation”. *Physical Review A* **86**, 032324 (2012).
- [95] Stephen T Spitz, Brian Tarasinski, Carlo WJ Beenakker, and Thomas E O’Brien. “Adaptive

- weight estimator for quantum error correction in a time-dependent environment”. *Advanced Quantum Technologies* **1**, 1800012 (2018).
- [96] J Kelly, R Barends, AG Fowler, A Megrant, E Jeffrey, TC White, D Sank, JY Mutus, B Campbell, Yu Chen, et al. “Scalable in situ qubit calibration during repetitive error detection”. *Physical Review A* **94**, 032321 (2016).
- [97] Ming-Xia Huo and Ying Li. “Learning time-dependent noise to reduce logical errors: real time error rate estimation in quantum error correction”. *New Journal of Physics* **19**, 123032 (2017).
- [98] Soraya Taghavi, Robert L Kosut, and Daniel A Lidar. “Channel-optimized quantum error correction”. *IEEE transactions on information theory* **56**, 1461–1473 (2010).
- [99] Jan Florjanczyk and Todd A Brun. “In-situ adaptive encoding for asymmetric quantum error correcting codes” (2016).
- [100] Jon Tyson. “Two-sided estimates of minimum-error distinguishability of mixed quantum states via generalized holevo–curlander bounds”. *Journal of mathematical physics* **50**, 032106 (2009).
- [101] Jon Tyson. “Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates”. *Journal of mathematical physics* **51**, 092204 (2010).
- [102] Howard Barnum and Emanuel Knill. “Reversing quantum dynamics with near-optimal quantum and classical fidelity”. *Journal of Mathematical Physics* **43**, 2097–2106 (2002).
- [103] Akshaya Jayashankar and Prabha Mandayam. “Quantum error correction: Noise-adapted techniques and applications”. *Journal of the Indian Institute of Science* Pages 1–16 (2022).
- [104] Debjyoti Biswas, Gaurav M Vaidya, and Prabha Mandayam. “Noise-adapted recovery circuits for quantum error correction” (2023).
- [105] Rémi Azouit, Francesca Chittaro, Alain Sarlette, and Pierre Rouchon. “Towards generic adiabatic elimination for bipartite open quantum systems”. *Quantum Science and Technology* **2**, 044011 (2017).
- [106] Rochus Klesse and Sandra Frank. “Quantum error correction in spatially correlated quantum noise”. *Physical review letters* **95**, 230503 (2005).
- [107] Ke Li and Andreas Winter. “Squashed entanglement, k-extendibility, quantum markov chains, and recovery maps”. *Foundations of Physics* **48**, 910–924 (2018).
- [108] Mark M Wilde. “Recoverability in quantum information theory”. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **471**, 20150338 (2015).
- [109] Omar Fawzi and Renato Renner. “Quantum conditional mutual information and approximate markov chains”. *Communications in Mathematical Physics* **340**, 575–611 (2015).
- [110] Richard Kueng, David M Long, Andrew C Doherty, and Steven T Flammia. “Comparing experiments to the fault-tolerance threshold”. *Physical review letters* **117**, 170502 (2016).
- [111] Panos Aliferis. “Level reduction and the quantum threshold theorem” (2007).
- [112] Alexei Gilchrist, Nathan K Langford, and Michael A Nielsen. “Distance measures to compare real and ideal quantum processes”. *Physical Review A* **71**, 062310 (2005).
- [113] Andrea Smirne, Nina Megier, and Bassano Vacchini. “Holevo skew divergence for the characterization of information backflow”. *Phys. Rev. A* **106**, 012205 (2022).
- [114] Felix Leditzky, Eneet Kaur, Nilanjana Datta, and Mark M Wilde. “Approaches for approximate additivity of the holevo information of quantum channels”. *Physical Review A* **97**, 012332 (2018).
- [115] Sumeet Khatri and Mark M Wilde. “Principles of quantum communication theory: A modern approach” (2020).
- [116] Mark M Wilde, Mario Berta, Christoph Hirche, and Eneet Kaur. “Amortized channel divergence for asymptotic quantum channel discrimination”. *Letters in Mathematical Physics* **110**, 2277–2336 (2020).
- [117] David Gross, Koenraad Audenaert, and Jens Eisert. “Evenly distributed unitaries: On the structure of unitary designs”. *Journal of mathematical physics* **48**, 052104 (2007).
- [118] Christoph Dankert. “Efficient simulation of random quantum states and operators” (2005).
- [119] Joseph Emerson, Marcus Silva, Osama Moussa, Colm Ryan, Martin Laforest, Jonathan Baugh, David G Cory, and Raymond Laflamme. “Symmetrized characterization of noisy quantum processes”. *Science* **317**, 1893–1896 (2007).
- [120] Tilo Eggeling and Reinhard F Werner. “Separability properties of tripartite states with  $u \otimes u \otimes u$  symmetry”. *Physical Review A* **63**, 042111 (2001).
- [121] Marcus Silva, Easwar Magesan, David W Kribs, and Joseph Emerson. “Scalable protocol for identification of correctable codes”. *Physical Review A* **78**, 012347 (2008).
- [122] Easwar Magesan. “Gaining information about a quantum channel via twirling”. Master’s thesis. University of Waterloo. (2008).
- [123] Josu Etzezarreta Martinez, Patricio Fuentes, Pedro M Crespo, and Javier Garcia-Frias. “Approximating decoherence processes for the design and simulation of quantum error correction

- codes on classical computers”. *IEEE Access* **8**, 172623–172643 (2020).
- [124] Adam M Meier. “Randomized benchmarking of clifford operators”. PhD thesis. University of Colorado at Boulder. (2013).
- [125] Daniel Gottesman. “The heisenberg representation of quantum computers” (1998).
- [126] Michael A Nielsen. “A simple formula for the average gate fidelity of a quantum dynamical operation”. *Physics Letters A* **303**, 249–252 (2002).
- [127] Giulio Chiribella, G Mauro D’Ariano, and Paolo Perinotti. “Transforming quantum operations: Quantum supermaps”. *EPL (Europhysics Letters)* **83**, 30004 (2008).
- [128] Rajendra Bhatia. “Matrix analysis”. *Volume 169*. Springer Science & Business Media. (2013).
- [129] Dénes Petz and Catalin Ghinea. “Introduction to quantum fisher information”. In *Quantum probability and related topics*. Pages 261–281. World Scientific (2011).
- [130] Masahito Hayashi. “Asymptotic theory of quantum statistical inference: selected papers”. *World Scientific*. (2005).
- [131] Hiroshi Hasegawa. “Exponential and mixture families in quantum statistics: Dual structure and unbiased parameter estimation”. *Reports on Mathematical Physics* **39**, 49–68 (1997).
- [132] Larry Wasserman. “All of statistics: a concise course in statistical inference”. *Volume 26*. Springer. (2004).
- [133] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. “On quantum rényi entropies: A new generalization and some properties”. *Journal of Mathematical Physics* **54**, 122203 (2013).
- [134] Nathaniel Johnston and David W Kribs. “Quantum gate fidelity in terms of choi matrices”. *Journal of Physics A: Mathematical and Theoretical* **44**, 495303 (2011).
- [135] Isaac L Chuang and Michael A Nielsen. “Prescription for experimental determination of the dynamics of a quantum black box”. *Journal of Modern Optics* **44**, 2455–2467 (1997).
- [136] Masoud Mohseni, Ali T Rezakhani, and Daniel A Lidar. “Quantum-process tomography: Resource analysis of different strategies”. *Physical Review A* **77**, 032322 (2008).
- [137] Emanuel Knill, Dietrich Leibfried, Rolf Reichle, Joe Britton, R Brad Blakestad, John D Jost, Chris Langer, Roee Ozeri, Signe Seidelin, and David J Wineland. “Randomized benchmarking of quantum gates”. *Physical Review A* **77**, 012307 (2008).
- [138] Joshua Combes, Christopher Ferrie, Chris Cesare, Markus Tiersch, Gerard J Milburn, Hans J Briegel, and Carlton M Caves. “In-situ characterization of quantum devices with error correction” (2014).
- [139] Shelby Kimmel, Marcus P da Silva, Colm A Ryan, Blake R Johnson, and Thomas Ohki. “Robust extraction of tomographic information via randomized benchmarking”. *Physical Review X* **4**, 011050 (2014).

## A Lower-Bounding Generalized Distinguishability Measures Using Entanglement Fidelity

### A.1 Generalized Distinguishability and Distance Measures

To quantify the success of a recovery protocol (such as QEC), we need to introduce the concepts of generalized distinguishability and distance measures between two states as well as between two channels [110–115].

We say that  $\mathbf{D} : \mathcal{D}(\mathcal{H}) \times \mathcal{L}_+(\mathcal{H}) \rightarrow \mathbb{R}^1$  is a generalized distinguishability measure between two states if it satisfies the *data-processing inequality* (DPI), i.e. for arbitrary  $\mathcal{Q}$  CPTP map and all  $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ , we have <sup>1</sup>

$$\mathbf{D}(\mathcal{Q}(\rho), \mathcal{Q}(\sigma)) \leq \mathbf{D}(\rho, \sigma). \quad (109)$$

An important consequence of DPI is the property of isometric invariance. Namely, for any isometry  $V$ , the following holds [115]

$$\mathbf{D}(\mathcal{V}(\rho), \mathcal{V}(\sigma)) = \mathbf{D}(\rho, \sigma), \quad (110)$$

where  $\mathcal{V}(\cdot) = V^\dagger(\cdot)V$ .

Independently, we say that  $\mathbf{D} : \mathcal{D}(\mathcal{H}) \times \mathcal{D}(\mathcal{H}) \rightarrow \mathbb{R}_+^1$  is a generalized distance measure between two states if it satisfies the following three properties for all  $\rho, \sigma, \tau \in \mathcal{D}(\mathcal{H})$ :

1. *Positivity and faithfulness*:

$$\mathbf{D}(\rho, \sigma) \geq 0, \quad (111)$$

where the equality holds iff  $\rho = \sigma$ .

2. *Symmetry*:  $\mathbf{D}(\rho, \sigma) = \mathbf{D}(\sigma, \rho)$ .

3. *Triangle inequality*:

$$\mathbf{D}(\rho, \sigma) \leq \mathbf{D}(\rho, \tau) + \mathbf{D}(\tau, \sigma). \quad (112)$$

<sup>1</sup>Some previous papers have used the notation  $\mathbf{D}(\rho||\sigma)$ , rather than  $\mathbf{D}(\rho, \sigma)$ , to indicate the generalized distinguishability measure between  $\rho$  and  $\sigma$ . Here, we use the latter notation to emphasize the role that  $\mathbf{D}$  plays also as a distance measure in deriving standard upper bounds in the context of QEC. Please Appendix C for more details.

A common requirement for fault-tolerant QEC and quantum computing is the so-called ‘‘chaining property’’ [110, 112]. However, this property of generalized distinguishability/distance measures is derivative from more elementary properties, such as DPI and the triangle inequality (see Appendix B for a short discussion).

Finally, we say that the map  $\mathbf{D} : \mathcal{D}(\mathcal{H}) \times \mathcal{L}_+(\mathcal{H}) \rightarrow \mathbb{R}^1$  satisfies the *joint convexity* property if for any two ensembles  $\{p_X(x), \rho^x\}_{x \in \mathcal{X}}$  and  $\{p_X(x), \sigma^x\}_{x \in \mathcal{X}}$ , where  $p_X$  is a probability distribution function of the random variable  $X$  over the set  $\mathcal{X}$ , we have

$$\mathbf{D} \left( \sum_{x \in \mathcal{X}} p_X(x) \rho^x, \sum_{x \in \mathcal{X}} p_X(x) \sigma^x \right) \leq \sum_{x \in \mathcal{X}} p_X(x) \mathbf{D}(\rho^x, \sigma^x). \quad (113)$$

For fidelity-based distinguishability measures, such as the Bures and Sine distances, this directly follows from the double concavity of the fidelity function (see e.g. [115]).

Alternatively, it is well-known that the joint convexity property can be derived from the DPI (with respect to the partial trace channel) if we further assume that  $\mathbf{D}$  satisfies the direct sum property for classical-quantum states [115], i.e.

$$\begin{aligned} \mathbf{D} \left( \sum_{x \in \mathcal{X}} p_X(x) |x\rangle\langle x| \otimes \rho^x, \sum_{x \in \mathcal{X}} p_X(x) |x\rangle\langle x| \otimes \sigma^x \right) \\ = \sum_{x \in \mathcal{X}} p_X(x) \mathbf{D}(\rho^x, \sigma^x). \end{aligned} \quad (114)$$

For a summary of various distinguishability and/or distance measures, as well as which properties they satisfy, please see Table 2. All the above properties are satisfied by [112, 115]

1. Trace Distance:  $\mathbf{D}_{\text{Tr}}(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1$ .
2. Bures Distance:  $\mathbf{D}_{\text{B}}(\rho, \sigma) = \sqrt{2 - 2\sqrt{F(\rho, \sigma)}}$ .
3. Sine Distance:  $\mathbf{D}_{\text{S}}(\rho, \sigma) = \sqrt{1 - F(\rho, \sigma)}$ ,

where  $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$  is the fidelity function.

Using the generalized distinguishability (distance) measures between two states, we define the generalized distinguishability (distance) measures between two channels  $\mathcal{Q}^{A \rightarrow B}$  and  $\mathcal{S}^{A \rightarrow B}$ , as follows

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) := \sup_{\rho} \mathbf{D}(\text{id}^R \otimes \mathcal{Q}^{A \rightarrow B}(\rho), \text{id}^R \otimes \mathcal{S}^{A \rightarrow B}(\rho)), \quad (115)$$

where  $\rho \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^R)$ , for arbitrary Hilbert space dimensions of the reference system  $R$ . By using joint convexity and the Schmidt decomposition of pure states, it can be shown that the maximization need only be taken over pure states  $\psi_{RA}$ , with the reference system  $R$  having the same Hilbert space dimensions as  $A$  [115], i.e.

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) := \sup_{\psi} \mathbf{D}(\text{id}^R \otimes \mathcal{Q}^{A \rightarrow B}(\psi), \text{id}^R \otimes \mathcal{S}^{A \rightarrow B}(\psi)). \quad (116)$$

Finally, it is important to note that the joint convexity property of generalized distinguishability measures for states implies the same property for channels. This is seen by considering the two channels  $\mathcal{Q}^{A \rightarrow B} = \sum_{x \in \mathcal{X}} p_X(x) \mathcal{Q}_x^{A \rightarrow B}$  and  $\mathcal{S}^{A \rightarrow B} = \sum_{x \in \mathcal{X}} p_X(x) \mathcal{S}_x^{A \rightarrow B}$ , and then applying the joint convexity property for states, as follows

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) = \sup_{\rho} \mathbf{D}(\text{id}^R \otimes \mathcal{Q}^{A \rightarrow B}(\rho), \text{id}^R \otimes \mathcal{S}^{A \rightarrow B}(\rho)) \quad (117)$$

$$= \mathbf{D}(\text{id}^R \otimes \mathcal{Q}^{A \rightarrow B}(\rho^*), \text{id}^R \otimes \mathcal{S}^{A \rightarrow B}(\rho^*)) \quad (118)$$

$$\leq \sum_{x \in \mathcal{X}} p_X(x) \mathbf{D}(\text{id}^R \otimes \mathcal{Q}_x^{A \rightarrow B}(\rho^*), \text{id}^R \otimes \mathcal{S}_x^{A \rightarrow B}(\rho^*)) \quad (119)$$

$$\leq \sum_{x \in \mathcal{X}} p_X(x) \sup_{\rho} \mathbf{D}(\text{id}^R \otimes \mathcal{Q}_x^{A \rightarrow B}(\rho), \text{id}^R \otimes \mathcal{S}_x^{A \rightarrow B}(\rho)) \quad (120)$$

$$\equiv \sum_{x \in \mathcal{X}} p_X(x) \mathbf{D}(\mathcal{Q}_x, \mathcal{S}_x). \quad (121)$$

Consequently, we have the joint convexity property

$$\begin{aligned} \mathbf{D} \left( \sum_{x \in \mathcal{X}} p_X(x) \mathcal{Q}_x, \sum_{x \in \mathcal{X}} p_X(x) \mathcal{S}_x \right) \\ \leq \sum_{x \in \mathcal{X}} p_X(x) \mathbf{D}(\mathcal{Q}_x, \mathcal{S}_x). \end{aligned} \quad (122)$$

## A.2 Unitary $t$ -designs

We call a function  $P : \mathbb{U}(d) \rightarrow \mathbb{C}$  acting on any unitary  $U$  in  $\mathbb{U}(d)$  to be polynomial of degree  $t$  if its dependence on the  $2d^2$  real entries of  $U$  is a polynomial of degree at most  $t$  in each of its entries. Given a finite set of unitaries  $\{U(x)\}_{x \in \mathcal{X}}$  in  $\mathbb{U}(d)$ , we say that they form a unitary  $t$ -design [54, 117] if the uniform Haar average over  $\mathbb{U}(d)$  of any polynomial  $P$  of degree  $t$  is computed using the uniform average over the finite set  $\{U(x)\}_{x \in \mathcal{X}}$  only, as follows

$$\int_{\mathbb{U}(d)} dU P(U) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} P(U(x)). \quad (123)$$

It has been shown that for unitary 1 and 2 designs, the above averaging condition can be rewritten in a different form. We say that  $\{U(x)\}_{x \in \mathcal{X}}$  forms a unitary 1-design in  $\mathbb{U}(d)$  if

$$\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} U(x) \rho U^\dagger(x) = \pi, \quad (124)$$

for all  $\rho \in \mathcal{D}(\mathcal{H})$ , where  $\pi = I/d$  is the maximally mixed state. An example of unitary 1-designs is given by the Pauli group. Further, we say  $\{U(x)\}_{x \in \mathcal{X}}$  forms a unitary 2-design in  $\mathbb{U}(d)$  if we have the following conditions for twirling of states or channels [54, 118]

$$\int_{\mathbb{U}(d)} dU (U \otimes U) \rho (U \otimes U)^\dagger$$

List of different measures			
Name	Data Processing	Distance Measure	joint convexity
Quantum Relative Entropy	Yes	No	Yes
Generalized $\alpha$ -Relative Entropies (Petz-Renyi, Sandwichead, etc.)	Yes	No	Yes
Trace Distance	Yes	Yes	Yes
Bures Distance	Yes	Yes	Yes
Sine Distance	Yes	Yes	Yes
Amortized Divergence [116]	Yes	No	Yes

Table 2: Summary of properties of various measures used in quantum information theory.

$$= \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} (U(x) \otimes U(x)) \rho (U(x) \otimes U(x))^\dagger, \quad (125)$$

for all  $\rho \in \mathcal{D}(\mathcal{H} \otimes \mathcal{H})$ , or equivalently

$$\int_{\mathbb{U}(d)} dU U^\dagger \mathcal{Q}(U \rho U^\dagger) U \\ = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} U^\dagger(x) \mathcal{Q}(U(x) \rho U^\dagger(x)) U(x), \quad (126)$$

for all  $\rho \in \mathcal{D}(\mathcal{H})$  and quantum channels  $\mathcal{Q}$ . An example of unitary 2-designs is given by the Clifford group [118, 119].

**Remark 5.** Note that, if  $\{U(x)\}_{x \in \mathcal{X}}$  is a unitary  $t$ -design, then it also holds that  $\{U(x)\}_{x \in \mathcal{X}}$  is a unitary  $(t-1)$ -design. For example, the Clifford group forms a unitary 3-design, and hence also a unitary 2-design.

### A.3 Channel Twirlings

Generally, channel twirlings can be defined with respect to both discrete and continuous sets of unitaries. In its most simple form, for a set of unitaries  $\{U_A(x), V_B(x)\}_{x \in \mathcal{X}}$  and a probability distribution function  $p_X$  defined over a finite set  $\mathcal{X}$ , the twirling of a quantum channel  $\mathcal{Q}^{A \rightarrow B}$  (which we denote by a tilde symbol  $\tilde{\mathcal{Q}}^{A \rightarrow B}$ ) is defined as

$$\tilde{\mathcal{Q}}^{A \rightarrow B} := \sum_{x \in \mathcal{X}} p_X(x) \mathcal{V}_x^{B^\dagger} \circ \mathcal{Q}^{A \rightarrow B} \circ \mathcal{U}_x^A, \quad (127)$$

where we have used the notation for the unitary channels  $\mathcal{U}_x^A(\cdot) := U_A^\dagger(x)(\cdot)U_A(x)$  and  $\mathcal{V}_x^B(\cdot) := V_B^\dagger(x)(\cdot)V_B(x)$ , for all  $x \in \mathcal{X}$ . Although most of the results presented in this article are valid for any finite set  $\mathcal{X}$ , the case where it forms a group and  $\{U_A(x), V_B(x)\}_{x \in \mathcal{X}}$  two unitary representations of it are of great interest [114] (see Remark 7).

Twirlings with continuous sets of unitaries have also been studied extensively in the literature. If we have some probability distribution (measure)  $\mu(U)$  over the set of  $d \times d$  unitary matrices  $\mathbb{U}(d)$ , then the continuous twirling of the channel  $\mathcal{Q}^A$  is defined to be

$$\tilde{\mathcal{Q}} := \int_{\mathbb{U}(d)} d\mu(U) U^\dagger \circ \mathcal{Q} \circ U. \quad (128)$$

Twirling of quantum channels plays an important role in QEC and fault-tolerant quantum computing [54, 119–124]. Examples include: (1) similarities between QEC codes for channels and their twirled versions [121], (2) the simulability of twirled quantum channels on a quantum computer, due to the Gottesman-Knill theorem [125], (3) the fact that channels and their twirled versions share the same average and entanglement fidelities [47], (4) various twirlings (Pauli, Clifford, and uniform Haar) rendering channels depolarizing [47, 54, 119, 126], (5) and finally, their close connection to unitary  $t$ -designs. Due to its importance, I recall some relevant properties of unitary  $t$ -designs in Appendix A.2 (also see [118, 122] for a brief review).

### A.4 Lower-Bounding Generalized Distinguishability Measures Using Entanglement Fidelity

We start this section by showing a simple property that all generalized distinguishability measures satisfy with respect to channel twirling if the joint convexity property (or equivalently, if the direct sum property) is further assumed.

**Lemma 5.** Assume that we are given two CPTP maps  $\mathcal{Q}^{A \rightarrow B}$  and  $\mathcal{S}^{A \rightarrow B}$ , a set of unitaries  $\{U_A(x), V_B(x)\}_{x \in \mathcal{X}}$ , and a probability distribution function  $p_X$  defined over the finite set  $\mathcal{X}$ . If the generalized distinguishability measure  $\mathbf{D}$  satisfies the joint convexity property, then  $\mathbf{D}(\mathcal{Q}, \mathcal{S})$  is lower bounded by the generalized distinguishability measure between the corresponding twirled channels  $\tilde{\mathcal{Q}}^{A \rightarrow B}$  and  $\tilde{\mathcal{S}}^{A \rightarrow B}$  with respect to the given weighted set of unitaries above, as follows

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) \geq \mathbf{D}(\tilde{\mathcal{Q}}, \tilde{\mathcal{S}}), \quad (129)$$

where the lower bound is saturated iff the joint convexity of the generalized distinguishability measure between the two quantum channels is saturated with respect to the above set of weighted unitaries.

*Proof.* Consider the isometric invariance property of  $\mathbf{D}(\mathcal{Q}, \mathcal{S})$ , namely for any  $\mathcal{U} := U(\cdot)U^\dagger$ , where  $U \in \mathbb{U}(d)$ , we have

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) = \mathbf{D}(U \circ \mathcal{Q}, U \circ \mathcal{S}) \quad (130)$$

$$= \mathbf{D}(\mathcal{Q} \circ \mathcal{U}, \mathcal{S} \circ \mathcal{U}), \quad (131)$$

where the first equality follows from Eq. (110) and the second equality follows from the definition in Eq. (115). This implies that for all  $U_A \in \mathbf{U}(d_A)$  and for all  $V_B \in \mathbf{U}(d_B)$

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) = \mathbf{D}(\mathcal{V}^\dagger \circ \mathcal{Q} \circ \mathcal{U}, \mathcal{V}^\dagger \circ \mathcal{S} \circ \mathcal{U}). \quad (132)$$

Consequently, by considering the generalized distinguishability measure  $\mathbf{D}(\tilde{\mathcal{Q}}, \tilde{\mathcal{S}})$  between the twirled channels, we arrive at

$$\mathbf{D} \left( \sum_{x \in \mathcal{X}} p_X(x) \mathcal{V}^{x^\dagger} \circ \mathcal{Q} \circ \mathcal{U}^x, \sum_{x \in \mathcal{X}} p_X(x) \mathcal{V}^{x^\dagger} \circ \mathcal{S} \circ \mathcal{U}^x \right) \quad (133)$$

$$\leq \sum_{x \in \mathcal{X}} p_X(x) \mathbf{D}(\mathcal{V}^{x^\dagger} \circ \mathcal{Q} \circ \mathcal{U}^x, \mathcal{V}^{x^\dagger} \circ \mathcal{S} \circ \mathcal{U}^x) \quad (134)$$

$$= \sum_{x \in \mathcal{X}} p_X(x) \mathbf{D}(\mathcal{Q}, \mathcal{S}) = \mathbf{D}(\mathcal{Q}, \mathcal{S}), \quad (135)$$

where the inequality follows from Eq. (122).  $\blacksquare$

**Remark 6.** This lemma can be viewed as a special case of a more general result for quantum supermaps. To elaborate, we recall that a supermap (a linear map from one quantum channel to another) can always be expressed as a pre and post-processing maps concatenated with the input quantum channel, and assisted by a memory [127]. Then, Lemma 5 follows from applying the data-processing inequality for generalized distinguishability measures between two quantum channels [55] with respect to channel twirling, which is a valid quantum supermap.

**Remark 7.** In [114], the authors have shown that for any two covariant channels  $\mathcal{F}^{A \rightarrow B}$  and  $\mathcal{G}^{A \rightarrow B}$  with respect to  $\{U_A(x), V_B(x)\}_{x \in \mathcal{X}}$  (namely that  $\mathcal{V}_x \circ \mathcal{F} = \mathcal{F} \circ \mathcal{U}_x$  for all  $x \in \mathcal{X}$ , and similarly for  $\mathcal{G}$ ), the generalized distinguishability measure

$$\mathbf{D}(\mathcal{F}, \mathcal{G}) = \sup_{\phi} \mathbf{D}((id \otimes \mathcal{F})(\phi_{RA}), (id \otimes \mathcal{G})(\phi_{RA})), \quad (136)$$

can be found by maximizing only over symmetric states  $\phi_{RA}$ , defined as

$$\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} U_A^\dagger(x) \phi_{RA} U_A(x) = \phi_{RA}. \quad (137)$$

However, since the twirlings  $\mathcal{F} \equiv \tilde{\mathcal{Q}}$  and  $\mathcal{G} \equiv \tilde{\mathcal{S}}$  in Lemma 5 are trivially covariant with respect to the unitary representations  $\{U_A(x), V_B(x)\}_{x \in \mathcal{X}}$  of the finite group  $\mathcal{X}$ , this implies that the lower bound in Eq. (129) need only be computed for such symmetric states. Furthermore, if  $\{U_A(x)\}_{x \in \mathcal{X}}$  is a unitary 1-design (i.e. it is an irreducible representation of the group  $\mathcal{X}$  of degree- $d_A$ ), then, using the property Eq. (124) of unitary 1-designs, the maximization is

found by computing the generalized distinguishability measure exactly for the maximally entangled state

$$\mathbf{D}(\tilde{\mathcal{Q}}, \tilde{\mathcal{S}}) = \mathbf{D}((id^R \otimes \tilde{\mathcal{Q}}^A)(\Phi_{RA}), (id^R \otimes \tilde{\mathcal{S}}^A)(\Phi_{RA})). \quad (138)$$

So far, we have shown that the generalized distinguishability measure between  $\mathcal{Q}^{A \rightarrow B}$  and  $\mathcal{S}^{A \rightarrow B}$  is lower bounded by the corresponding distinguishability measure for arbitrary discrete twirlings of these channels. We now show that a similar lower bound can be derived for the uniform Haar twirling. But first, we recall the following important result

**Lemma 6.** ([47]) Given a CPTP map  $\mathcal{Q}^{A \rightarrow A}$  and for all  $\rho \in \mathcal{D}(\mathcal{H}^A)$ , the continuous twirling  $\tilde{\mathcal{Q}} = \int_{\mathbf{U}(d)} dU U^\dagger \circ \mathcal{Q} \circ U$  over the uniform Haar measure on the set of  $d \times d$  unitary matrices  $\mathbf{U}(d)$  is given by the depolarizing channel

$$\tilde{\mathcal{Q}}(\rho) = (1 - p^\mathcal{Q})\rho + p^\mathcal{Q} \frac{I}{d}, \quad (139)$$

where the depolarizing parameter  $p^\mathcal{Q}$  is given by the average fidelity of  $\mathcal{Q}$ , as follows

$$p^\mathcal{Q} = \frac{d}{d-1} (1 - F_{\text{avg}}(\mathcal{Q})). \quad (140)$$

The proof of Eq. (139) is shown in [47, 126] for some parameter value  $p^\mathcal{Q}$ . Eq. (140) is a direct consequence of the fact that the uniform Haar twirled channel  $\tilde{\mathcal{Q}} = \int_{\mathbf{U}(d)} dU U^\dagger \circ \mathcal{Q} \circ U$  has the same average fidelity as the original channel  $\mathcal{Q}$  [47], along with the fact that the average fidelity of the depolarizing channel is given by

$$F_{\text{avg}}(\tilde{\mathcal{Q}}) = 1 - \left( \frac{d-1}{d} \right) p, \quad (141)$$

where we have used the normalization  $\int d\psi = 1$  and the notation  $p$  for the depolarizing parameter.

Using the above Lemmas 5 and 6, we now establish a similar lower bound to that in Lemma 5 for the uniform Haar twirl.

**Theorem 4.** Assume that we are given two CPTP maps  $\mathcal{Q}^{A \rightarrow A}$  and  $\mathcal{S}^{A \rightarrow A}$ . If the generalized distinguishability measure  $\mathbf{D}$  satisfies the joint convexity property, then  $\mathbf{D}(\mathcal{Q}, \mathcal{S})$  is lower bounded by some function  $l_{\mathbf{D}}$  of the channel entanglement fidelities  $F_e(\mathcal{Q})$  and  $F_e(\mathcal{S})$ , as follows

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) \geq l_{\mathbf{D}}(F_e(\mathcal{Q}), F_e(\mathcal{S})), \quad (142)$$

where the specific form of the function  $l_{\mathbf{D}}$  depends on the choice of the generalized distinguishability measure and is determined by the uniform Haar twirls, as follows

$$l_{\mathbf{D}}(F_e(\mathcal{Q}), F_e(\mathcal{S})) \equiv \mathbf{D}(\tilde{\mathcal{Q}}, \tilde{\mathcal{S}}), \quad (143)$$

where  $\tilde{\mathcal{Q}} = \int_{\mathbf{U}(d)} dU U^\dagger \circ \mathcal{Q} \circ U$  and  $\tilde{\mathcal{S}} = \int_{\mathbf{U}(d)} dU U^\dagger \circ \mathcal{S} \circ U$  yield two depolarizing channels. The inequality is saturated if the joint convexity property is saturated for a set of unitary 2-designs and a uniform probability distribution over this set.

*Proof.* This is a direct consequence of applying Lemma 5 to any unitary 2-design  $\{U_A(x)\}_{x \in \mathcal{X}}$ , e.g. the unitary representation of the Clifford group (see Appendix A.2), along with a uniform distribution on  $\mathcal{X}$ , and then using the property of unitary 2-designs in Eq. (126), which finally yields

$$\mathbf{D}(\mathcal{Q}, \mathcal{S}) \geq \mathbf{D}(\tilde{\mathcal{Q}}, \tilde{\mathcal{S}}), \quad (144)$$

where  $\tilde{\mathcal{Q}} = \int_{\mathbf{U}(d)} dU U^\dagger \circ \mathcal{Q} \circ U$  and  $\tilde{\mathcal{S}} = \int_{\mathbf{U}(d)} dU U^\dagger \circ \mathcal{S} \circ U$ . The proof is completed by applying Lemma 6 and plugging in the depolarizing channels into the lower bound in Eq. (144). ■

It directly follows from this proof that the image of the function  $l_{\mathbf{D}}$  coincides with the image of the corresponding generalized distinguishability measure  $\mathbf{D}$ .

**Remark 8.** *The lower bound proof does not require faithfulness, symmetry, nor the triangle inequality, which would also make  $\mathbf{D}$  a generalized distance measure. However, the triangle inequality becomes necessary when deriving an upper bound for the generalized distinguishability measure for concatenated noisy channels (or gates), as it is relevant to fault-tolerant quantum computing (see Appendix C for more details).*

## B Comment on The Chaining Property

In quantum computing literature, one encounters the chaining property for distance measures [110], which is useful for computing upper bounds on error propagation in fault-tolerant quantum computing. This property is framed as follows: Assume we want to apply two maps  $\mathcal{Q}$  and  $\mathcal{S}$  in series, however, we only have access to their noisy versions, which we denote by  $\mathcal{Q}'$  and  $\mathcal{S}'$ , respectively. If the generalized distance measure  $\mathbf{D}$  also satisfies the DPI (i.e.  $\mathbf{D}$  is also a distinguishability measure), then the chaining property reads

$$\mathbf{D}(\mathcal{S} \circ \mathcal{Q}(\rho), \mathcal{S}' \circ \mathcal{Q}'(\rho)) \leq \mathbf{D}(\mathcal{Q}(\rho), \mathcal{Q}'(\rho)) + \mathbf{D}(\mathcal{S}(\rho), \mathcal{S}'(\rho)), \quad (145)$$

for all  $\rho \in \mathcal{D}(\mathcal{H})$ . This is interpreted by saying that the error due to a consecutive application of two faulty channels is no larger than the sum of the errors of applying each of the faulty channels separately. The proof follows by first applying the triangle inequality to the left-hand side of the above inequality, followed up by the data-processing inequality. Therefore, the desirable chaining property is derivative from other, more fundamental, properties of  $\mathbf{D}$ .

## C Upper-Bounding Generalized Distance Measures for State Recovery

Here we present upper bounds on generalized distance and distinguishability measures, showing how they get modified when limited knowledge about the noise parameter  $\theta \in \Theta$  is available, both for the single-cycle and multi-cycle cases. Similar to the chaining property, the derivation of upper bounds on generalized distance and distinguishability measures is important for the analysis of error propagation in noisy quantum processes.

### C.1 Single-Cycle Case

Consider the distance measure  $\mathbf{D}$  and assume that for all  $\rho \in \mathcal{D}(\mathcal{C}) \subseteq \mathcal{D}(\mathcal{H})$ , approximate recovery from the noise  $\mathcal{N}_\theta$  is possible in the presence of perfect information about  $\theta$ , i.e. there exists  $\mathcal{R}_\theta$  such that

$$\mathbf{D}(\mathcal{I}_\theta^\theta(\rho), \rho) \leq \epsilon_\theta \quad \text{where} \quad \mathcal{I}_\alpha^\beta \equiv \mathcal{R}_\beta \circ \mathcal{N}_\alpha. \quad (146)$$

Now consider the distance measure  $\mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \rho)$ , where  $\hat{\theta}$  is the best unbiased estimate of  $\theta \in \Theta$ . Our goal is to bound this quantity from above by two terms: the first depends on how well we can bound the same distance measure when given perfect knowledge about  $\theta$  (see Eq. (146)), and the second should measure our lack of knowledge of the noise parameter  $\theta$ . This intuition is validated by applying the triangle inequality, as follows

$$\mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \rho) \leq \mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho)) + \mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \rho) \quad (147)$$

$$\leq \mathbf{D}(\mathcal{N}_\theta(\rho), \mathcal{N}_{\hat{\theta}}(\rho)) + \mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \rho) \quad (148)$$

$$\leq \mathbf{D}(\mathcal{N}_\theta, \mathcal{N}_{\hat{\theta}}) + \mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \rho) \quad (149)$$

$$\leq \mathbf{D}(\mathcal{N}_\theta, \mathcal{N}_{\hat{\theta}}) + \epsilon_{\hat{\theta}} \equiv \epsilon_{\theta, \hat{\theta}}, \quad (150)$$

where the second inequality follows from the DPI of  $\mathbf{D}$ , the third follows from the definition of the generalized distance for channels, and the fourth from the assumption of Eq. (146). It is worth noting that one can derive a similar upper bound using the recoveries, rather than the noisy channels. The advantage of this approach is that we do not need to assume that  $\mathbf{D}$  satisfies the DPI, i.e. it suffices for  $\mathbf{D}$  to be a distance measure. To see how we simply apply the triangle inequality

$$\mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \rho) \leq \mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho)) + \mathbf{D}(\mathcal{I}_{\hat{\theta}}^{\hat{\theta}}(\rho), \rho) \quad (151)$$

$$\leq \mathbf{D}(\mathcal{R}_\theta, \mathcal{R}_{\hat{\theta}}) + \epsilon_\theta \equiv \epsilon'_{\theta, \hat{\theta}}, \quad (152)$$

where we have used the definition of a distance measure between channels for the second inequality, as well as Eq. (146). We will shortly show that DPI becomes necessary when considering the multi-cycle case.

## C.2 Multi-Cycle Case

Let us now extend the upper bound previously derived in the single-cycle case to adaptive multi-cycle recovery. Using the shorthand notation

$$\mathbf{D}_{\alpha_n \alpha_{n-1} \dots \alpha_1}^{\beta_n \beta_{n-1} \dots \beta_1}(\rho) \equiv \mathbf{D}(\mathcal{I}_{\alpha_n \alpha_{n-1} \dots \alpha_1}^{\beta_n \beta_{n-1} \dots \beta_1}(\rho), \rho), \quad (153)$$

where

$$\mathcal{I}_{\alpha_n \alpha_{n-1} \dots \alpha_1}^{\beta_n \beta_{n-1} \dots \beta_1} \equiv \mathcal{R}_{\beta_n} \circ \mathcal{E}_{\alpha_n} \circ \mathcal{R}_{\beta_{n-1}} \circ \mathcal{E}_{\alpha_{n-1}} \dots \circ \mathcal{R}_{\beta_1} \circ \mathcal{E}_{\alpha_1}, \quad (154)$$

and applying the triangle inequality, we get

$$\begin{aligned} \mathbf{D}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho) &\leq \mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \\ &\quad + \mathbf{D}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho). \end{aligned} \quad (155)$$

The second term could be bounded from above by the individual errors  $\{\epsilon_{\hat{\theta}_i}\}_{i=1}^n$ , using only the triangle inequality, as follows

$$\begin{aligned} \mathbf{D}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho) &\leq \mathbf{D}(\mathcal{I}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \\ &\quad + \mathbf{D}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho). \end{aligned} \quad (156)$$

We assume that

$$\mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho) \in \mathcal{D}(\mathcal{C}), \quad (157)$$

so that the  $n$ -th step approximate recovery with perfect knowledge of  $\theta$  would be possible, in principle. This leads to

$$\mathbf{D}(\mathcal{I}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \quad (158)$$

$$= \mathbf{D}(\mathcal{I}_{\hat{\theta}_n}^{\hat{\theta}_n}(\mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)), \mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \leq \epsilon_{\hat{\theta}_n}. \quad (159)$$

Substituting this result back into Eq. (156), we get

$$\mathbf{D}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho) \leq \mathbf{D}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho) + \epsilon_{\hat{\theta}_n}, \quad (160)$$

and repeating the above two steps yields

$$\mathbf{D}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1} \leq \sum_{i=1}^n \epsilon_{\hat{\theta}_i}. \quad (161)$$

The first term in Eq. (155) is a new error term due to the real-time (drift-adapting) nature of our setup. This term can be bounded from above using the chaining property and the DPI, as follows

$$\begin{aligned} \mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \\ = \mathbf{D}(\mathcal{I}_{\theta_n}^{\hat{\theta}_n} \circ \mathcal{I}_{\theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_n}^{\hat{\theta}_n} \circ \mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \end{aligned} \quad (162)$$

$$\leq \mathbf{D}(\mathcal{I}_{\theta_n}^{\hat{\theta}_n}(\rho), \mathcal{I}_{\hat{\theta}_n}^{\hat{\theta}_n}(\rho)) + \mathbf{D}(\mathcal{I}_{\theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \quad (163)$$

$$\leq \mathbf{D}(\mathcal{N}_{\theta_n}(\rho), \mathcal{N}_{\hat{\theta}_n}(\rho)) + \mathbf{D}(\mathcal{I}_{\theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \quad (164)$$

$$\leq \mathbf{D}(\mathcal{N}_{\theta_n}, \mathcal{N}_{\hat{\theta}_n}) + \mathbf{D}(\mathcal{I}_{\theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)). \quad (165)$$

Repeating the above steps  $n - 1$  times, we arrive at the upper bound

$$\mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho)) \leq \sum_{i=1}^n \mathbf{D}(\mathcal{N}_{\theta_i}, \mathcal{N}_{\hat{\theta}_i}). \quad (166)$$

Combining Eqs. (161) and (166) with Eq. (155), we get

$$\mathbf{D}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1} \leq \sum_{i=1}^n [\mathbf{D}(\mathcal{N}_{\theta_i}, \mathcal{N}_{\hat{\theta}_i}) + \epsilon_{\hat{\theta}_i}] \equiv \sum_{i=1}^n \epsilon_{\theta_i, \hat{\theta}_i}, \quad (167)$$

which generalizes Eq. (150) for real-time approximate recovery. This result says that, if AQEC is possible in principle (see Eq. (157)) when perfect knowledge of  $\theta$  is available, then AQEC is also possible when knowledge about  $\theta$  is limited. As we have shown, this holds for both the single-cycle and multi-cycle regimes.

Alternatively, we can derive an upper bound that is a function of the recoveries, rather than the noisy channels. This is accomplished as follows

$$\begin{aligned} \mathbf{D}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho) &\leq \mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\theta_n \theta_{n-1} \dots \theta_1}(\rho)) \\ &\quad + \mathbf{D}_{\theta_n \theta_{n-1} \dots \theta_1}^{\theta_n \theta_{n-1} \dots \theta_1}(\rho), \end{aligned} \quad (168)$$

where the second term is similarly bounded from above by  $\sum_{i=1}^n \epsilon_{\theta_i}$ , based only on the triangle inequality (see Eq. (161)). We now upper bound the first term in the above inequality as

$$\begin{aligned} \mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\theta_n \theta_{n-1} \dots \theta_1}(\rho)) \\ \leq \mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \theta_{n-1} \dots \theta_1}(\rho)) \\ + \mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \theta_{n-1} \dots \theta_1}(\rho), \mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\theta_n \theta_{n-1} \dots \theta_1}(\rho)) \end{aligned} \quad (169)$$

$$\leq \mathbf{D}(\mathcal{I}_{\theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\theta_{n-1} \dots \theta_1}^{\theta_{n-1} \dots \theta_1}(\rho)) + \mathbf{D}(\mathcal{R}_{\hat{\theta}_n}, \mathcal{R}_{\theta_n}), \quad (170)$$

where we have used the triangle inequality for the first inequality and the DPI and the definition of generalized distance measure between channels for the second inequality. By repeating these two steps  $n - 1$  times, we arrive at

$$\mathbf{D}(\mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1}(\rho), \mathcal{I}_{\theta_n \theta_{n-1} \dots \theta_1}^{\theta_n \theta_{n-1} \dots \theta_1}(\rho)) \leq \sum_{i=1}^n \mathbf{D}(\mathcal{R}_{\hat{\theta}_i}, \mathcal{R}_{\theta_i}). \quad (171)$$

Consequently, Eq. (168) yields

$$\mathbf{D}_{\theta_n \theta_{n-1} \dots \theta_1}^{\hat{\theta}_n \hat{\theta}_{n-1} \dots \hat{\theta}_1} \leq \sum_{i=1}^n [\mathbf{D}(\mathcal{R}_{\theta_i}, \mathcal{R}_{\hat{\theta}_i}) + \epsilon_{\theta_i}] \equiv \sum_{i=1}^n \epsilon_{\theta_i, \hat{\theta}_i}. \quad (172)$$



## D Necessary and Sufficient Condition for Independence of the Reduced Channel From a Mother Channel Parameter

In this appendix, we are interested in proving the following:

**Lemma 7.** *Consider the mother channel  $\mathcal{Z}_\theta^{MS}$  of a bipartite system  $MS$ , where  $\theta \in \Theta^p$  is a  $p$ -dimensional parameter space. The reduced dynamics of subsystem  $S$ , defined by partial tracing over the subsystem  $M$  via  $\mathcal{M}_\theta^S \equiv \text{Tr}_M \circ \mathcal{Z}_\theta^{MS}$ , is independent of the  $\alpha$ -th component of the  $p$ -dimensional vector  $\theta$  if and only if the Choi matrix  $\Gamma^{\mathcal{Z}_\theta}$  of the mother channel satisfies*

$$\text{Tr}_{M''S''} \left[ \left( \frac{\partial}{\partial \theta_\alpha} \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \right) (2P_{S''S'}^{\text{sym}} - I_{S''S'}) \right] = 0, \quad (173)$$

where  $P_{S''S'}^{\text{sym}}$  is the projector onto the symmetric subspace

$$\text{span}(|\mu\rangle_{S''} |\nu\rangle_{S'} + |\nu\rangle_{S''} |\mu\rangle_{S'}) \subset \mathcal{H}^{S''} \otimes \mathcal{H}^{S'}, \quad (174)$$

and  $\{|\mu\rangle\}$  is a basis set of the Hilbert space  $\mathcal{H}^S$ .

*Proof.* We consider the dynamics of the combined memory-spectator ( $MS$ ) system, and express it in terms of the Choi matrix of the mother channel  $\mathcal{Z}_\theta^{MS}$ , as follows [115]

$$\begin{aligned} & \mathcal{Z}_\theta^{MS \rightarrow M'S'}(\rho_{MS}) \\ &= \text{Tr}_{MS} \left[ (\mathbb{T}_{MS}(\rho_{MS}) \otimes I_{M'S'}) \Gamma_{MS, M'S'}^{\mathcal{Z}_\theta} \right], \end{aligned} \quad (175)$$

then the reduced dynamics of the spectator system yields

$$\begin{aligned} & \text{Tr}_{M'} \circ \mathcal{Z}_\theta^{MS \rightarrow M'S'}(\rho_{MS}) \\ &= \text{Tr}_{MS} \left[ (\mathbb{T}_{MS}(\rho_{MS}) \otimes I_{S'}) \Gamma_{MS, S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \right]. \end{aligned} \quad (176)$$

Therefore, for the reduced dynamics to be independent of the noise parameter  $\theta_\alpha$  for some  $\alpha = 1, \dots, p$  and any joint input state  $\rho_{MS}$ , we must have

$$\frac{\partial}{\partial \theta_\alpha} \Gamma_{MS, S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} = 0, \quad (177)$$

for all  $\theta \in \Theta^p$ . We now derive a necessary and sufficient condition for this equality to hold, in terms of the Choi matrix  $\Gamma_{MS, M'S'}^{\mathcal{Z}_\theta}$  of the mother channel. We start by recalling the formula for the Choi matrix of the composite channel in terms of the Choi matrices of the individual channels [115]

$$\begin{aligned} & \Gamma_{MS, S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \\ &= \text{Tr}_{M''S''} \left[ \mathbb{T}_{M''S''} \left( \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \right) \Gamma_{M''S'', S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \right] \end{aligned} \quad (178)$$

$$= \text{Tr}_{M''S''} \left[ \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \mathbb{T}_{M''S''}^\dagger \left( \Gamma_{M''S'', S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \right) \right] \quad (179)$$

$$= \text{Tr}_{M''S''} \left[ \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \mathbb{T}_{M''S''} \left( \Gamma_{M''S'', S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \right) \right]. \quad (180)$$

We now compute the basis dependent matrix  $\mathbb{T}_{M''S''} \left( \Gamma_{M''S'', S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \right)$  in the separable memory-spectator basis  $|i\rangle_{MS} \equiv |i(a, \mu)\rangle_{MS} = |a\rangle_M |\mu\rangle_S$  of the Hilbert space  $\mathcal{H}^M \otimes \mathcal{H}^S$ , as follows

$$\Gamma_{M''S'', S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} = \sum_{ij} |i\rangle\langle j|_{M''S''} \otimes \text{Tr}_{M'}(|i\rangle\langle j|_{M'S'}) \quad (181)$$

$$= I_{M''} \otimes \Gamma_{S''S'}, \quad (182)$$

where  $|\Gamma\rangle_{S''S'} = \sum_\mu |\mu\rangle_{S''} |\mu\rangle_{S'}$  is the maximally entangled state in the special spectator basis  $\{|\mu\rangle\}$ . In the same  $\{|a\rangle \otimes |\mu\rangle\}$  basis, the partial transpose yields

$$\mathbb{T}_{M''S''} \left( \Gamma_{M''S'', S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \right) = \mathbb{T}_{M''S''} (I_{M''} \otimes \Gamma_{S''S'}) \quad (183)$$

$$= I_{M''} \otimes ((T_{S''} \otimes \text{id}^{S'}) (\Gamma_{S''S'})) \quad (184)$$

$$= I_{M''} \otimes ((I^{S''} \otimes \mathbb{T}_{S'}) (\Gamma_{S''S'})) \quad (185)$$

$$= I_{M''} \otimes \Gamma_{S''S'}^\mathbb{T}, \quad (186)$$

where  $\Gamma_{S''S'}^\mathbb{T}$  is the Choi matrix of the partial transpose channel. It has been shown in citejohnston2011quantum that the Choi matrix  $\Gamma_{S''S'}^\mathbb{T}$  is related to the projector  $P_{S''S'}^{\text{sym}} = (I_{S''S'} + \Gamma_{S''S'}^\mathbb{T})/2$  onto the symmetric subspace

$$\text{span}(|\mu\rangle_{S''} |\nu\rangle_{S'} + |\nu\rangle_{S''} |\mu\rangle_{S'}) \subset \mathcal{H}^{S''} \otimes \mathcal{H}^{S'}, \quad (187)$$

where  $d_S(d_S+1)/2$  is the dimensions of the symmetric subspace in the  $d_S^2$  dimensional Hilbert space  $\mathcal{H}^S \otimes \mathcal{H}^S$ . This finally yields

$$\mathbb{T}_{M''S''} \left( \Gamma_{M''S'', S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \right) = I_{M''} \otimes \Gamma_{S''S'}^\mathbb{T} \quad (188)$$

$$= I_{M''} \otimes (2P_{S''S'}^{\text{sym}} - I_{S''S'}). \quad (189)$$

Substituting back into Eq. (180) leads to

$$\begin{aligned} & \Gamma_{MS, S'}^{\text{Tr}_{M'} \circ \mathcal{Z}_\theta} \\ &= \text{Tr}_{M''S''} \left[ \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} (2P_{S''S'}^{\text{sym}} - I_{S''S'}) \right]. \end{aligned} \quad (190)$$

Therefore, Eq. (177) holds if and only if

$$\text{Tr}_{M''S''} \left[ \left( \frac{\partial}{\partial \theta_\alpha} \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \right) (2P_{S''S'}^{\text{sym}} - I_{S''S'}) \right] = 0. \quad (191)$$

The last equation in the proof can be equivalently written as (using the identity  $P_{S''S'}^{\text{sym}} + (P_{S''S'}^{\text{sym}})^\perp = I_{S''S'}$ )

$$\text{Tr}_{M''S''} \left[ \left( \frac{\partial}{\partial \theta_\alpha} \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \right) P_{S''S'}^{\text{sym}} \right]$$

$$= \text{Tr}_{M''S''} \left[ \left( \frac{\partial}{\partial \theta_\alpha} \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \right) (P_{S''S''}^{\text{sym}})^\perp \right] \quad (192)$$

$$= \frac{1}{2} \frac{\partial}{\partial \theta_\alpha} \text{Tr}_{M''S''} \left[ \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} \right] \quad (193)$$

$$= \frac{1}{2} \frac{\partial}{\partial \theta_\alpha} I_{MS} = 0. \quad (194)$$

**Remark 9.** Note that the condition Eq. 191 is weaker than

$$\frac{\partial}{\partial \theta_\alpha} \Gamma_{MS, M''S''}^{\mathcal{Z}_\theta} = 0, \quad (195)$$

for all  $\theta \in \Theta^p$ , which holds when the mother channel  $\mathcal{Z}_\theta^{MS}$  itself does not depend on the noise parameter  $\theta_\alpha$ , and hence trivially also the reduced channel  $\mathcal{M}_\theta^S \equiv \text{Tr}_M \circ \mathcal{Z}_\theta^{MS}$ .

## E Quantum Fisher Information Matrix

We review the relevant definitions and results regarding quantum Fisher Information Matrix (QFIM) and the partial QFIM, following [63, 65].

### E.1 Useful Definitions

For a family of parameterized quantum states  $\{\rho_\theta\}_{\theta \in \Theta}$  with a  $p$ -dimensional parameter space  $\Theta \subseteq \mathbb{R}^p$ , we define the symmetric inner product between two linear operators  $A$  and  $B$ , with respect to the parameterized family of states, as

$$\langle A, B \rangle_{\rho_\theta} := \text{Tr} \left[ \rho_\theta \left( \frac{1}{2} \{A^\dagger, B\} \right) \right], \quad (196)$$

where  $\{a, b\} := ab + ba$  is the anti-commutator. The symmetric logarithmic derivative (SLD) is a Hermitian operator  $L_{\theta, i}$  that is defined by the solution to the Lyapunov type equation [128]

$$\frac{\partial}{\partial \theta_\alpha} \rho_\theta =: \frac{1}{2} \{L_{\theta, \alpha}, \rho_\theta\}, \quad (197)$$

The SLD QFIM corresponding to the parameterized family of states is defined as the  $p \times p$  matrix

$$[I_{\text{QF}}(\theta; \{\rho_\theta\})]_{\alpha\beta} \equiv I_{\alpha\beta}^\theta := \langle L_{\theta, \alpha}, L_{\theta, \beta} \rangle_{\rho_\theta}. \quad (198)$$

Next, assume that a quantum measurement of an observable  $X$  is performed on  $\rho_\theta$ , described by a POVM  $\Pi \equiv \{\Pi_x\}_{x \in \mathcal{X}}$ . This yields the statistics  $p_X(x|\theta) = \text{Tr}[\rho_\theta \Pi_x]$  for the measurement outcomes  $x \in \mathcal{X}$ . We define an estimate  $\hat{\theta} : \mathcal{X} \rightarrow \Theta^p$  as a mapping from the set of measurement outcomes to the parameter space. We say that the pair  $(\Pi, \hat{\theta})$  is an estimator, and call it *unbiased* if

$$\mathbb{E} \left[ \hat{\theta}(X) \right]_{p(x|\theta)} := \sum_{x \in \mathcal{X}} \hat{\theta}(x) \text{Tr}[\rho_\theta \Pi_x] = \theta. \quad (199)$$

In general, such an estimator does not exist for all  $\theta \in \Theta^p$ . Instead, it is customary to use a weaker

condition on our estimator, namely that it is *locally unbiased*. This is defined as follows: at a fixed  $\theta$ , we require that the following two conditions are satisfied

$$\mathbb{E} \left[ \hat{\theta}(X) \right]_{p(x|\theta)} = \sum_{x \in \mathcal{X}} \hat{\theta}(x) \text{Tr}[\rho_\theta \Pi_x] = \theta, \quad (200)$$

and

$$\frac{\partial}{\partial \theta_\beta} \mathbb{E} \left[ \hat{\theta}_\alpha(X) \right]_{p(x|\theta)} = \sum_{x \in \mathcal{X}} \hat{\theta}_\alpha(x) \text{Tr} \left[ \frac{\partial}{\partial \theta_\beta} \rho_\theta \Pi_x \right] = \delta_{\alpha\beta}. \quad (201)$$

Finally, we define the mean-square error (MSE) matrix corresponding to an estimator  $(\Pi, \hat{\theta})$  as follows

$$\text{Var} \left[ \hat{\theta}(X) \right] := \mathbb{E} \left[ \left( \hat{\theta}(X) - \theta \right)^T \left( \hat{\theta}(X) - \theta \right) \right]_{p(x|\theta)}, \quad (202)$$

with the  $(\alpha, \beta)$  entry of this matrix given by

$$\mathbb{E} \left[ \left( \hat{\theta}_\alpha(X) - \theta_\alpha \right) \left( \hat{\theta}_\beta(X) - \theta_\beta \right) \right]_{p(x|\theta)}. \quad (203)$$

### E.2 Saturation of QCRB

The quantum Cramér-Rao bound (QCRB) provides a lower bound to the variance matrix defined in Eq. (202) using the QFIM in Eq. (198) [63, 129]

$$\text{Var} \left[ \hat{\theta} \right] \geq \left( I^\theta \right)^{-1}. \quad (204)$$

The QCRB holds for any locally unbiased estimator, and is a direct consequence to applying the Schwartz inequality for the inner product defined in Eq. (196). Here, we are interested in the saturation condition for this inequality. A necessary and sufficient condition for the saturation of the multi-parameter QCRB is given by (see e.g. [65, 130])

$$\text{Tr}[\rho_\theta [L_{\theta, \alpha}, L_{\theta, \beta}]] = 0 \text{ for all } \alpha, \beta = 1, \dots, p. \quad (205)$$

To design the optimal measurements for the saturation of the multi-parameter QCRB, we conduct the following: (i) find an SLD  $\{L_{\theta, \alpha}\}$  that mutually commute, (ii) using the matrices  $I^\theta$  and  $\{L_{\theta, \alpha}\}$ , construct the (commuting) linear combinations

$$\tilde{L}_{\theta, \alpha} := \sum_{\beta=1}^p \left( I^\theta \right)_{\alpha\beta}^{-1} L_{\theta, \beta}, \quad (206)$$

for all  $\alpha = 1, \dots, p$ , and (iii) write the spectral decomposition of the mutually commuting operators

$$\tilde{L}_{\theta, \alpha} = \sum_i l_{\alpha i} P_i, \quad (207)$$

for  $\alpha = 1, \dots, p$ , where  $\{P_i\}$  are the projectors onto the simultaneous eigenspaces of  $\{\tilde{L}_{\theta, \alpha}\}$  (or equivalently for  $\{L_{\theta, \alpha}\}$ ). Then, the QCRB is saturated if

we pick the locally unbiased estimator  $(\Pi, \hat{\theta})$  to be [63]

$$\Pi_x \equiv P_{i=x}, \quad (208)$$

$$\hat{\theta}_\alpha(x) \equiv \theta_\alpha + l_{\alpha x}. \quad (209)$$

In the case of a single parameter family, this yields the locally unbiased estimator

$$\hat{\theta}(x) = \theta + \left( \mathbf{l}^\theta \right)^{-1} \frac{d}{d\theta} \log p(x|\theta), \quad (210)$$

where  $p(x|\theta) = \text{Tr}[\rho_\theta \Pi_x]$ , which explicitly depend on  $\theta$ . Although the optimal measurements described above saturate the QCRB, they require prior knowledge of the noise parameters, which defeats the point of implementing spectator systems. In the single-parameter case, this can be remedied.

### E.2.1 Parameter-Independent Estimator

Nagaoka has shown that, in the single parameter case, a  $\theta$ -independent locally unbiased estimator exists that saturates the QCRB [130]. This is possible only for an exponential family  $\{\rho_\theta\}_{\theta \in \Theta}$  of parameterized states [131]:

$$\rho_\theta = e^{\frac{1}{2} \int_0^\theta \psi(\theta') d\theta' (O - \theta_\psi)} \rho_0 e^{\frac{1}{2} \int_0^\theta \psi(\theta') d\theta' (O - \theta_\psi)}, \quad (211)$$

for some  $\rho_0$ , where we have assumed for convenience that  $\theta = 0 \in \Theta$ ,  $\psi(\theta)$  is some function of  $\theta$ ,

$$\theta_\psi = \frac{\int_0^\theta \theta' \psi(\theta') d\theta'}{\int_0^\theta \psi(\theta') d\theta'}, \quad (212)$$

and  $O$  is an unbiased observable of  $\theta$ , i.e.  $\text{Tr}[\rho_\theta O] = \theta$ . As such, the SLD of this parametric family is given by  $L_\theta = \psi(\theta)(O - \theta)$ , which guarantees that the Schwartz inequality for the two vectors  $L_\theta$  and  $O - \theta$  is saturated, and hence the saturation of the QCRB [129, 130]. The optimal measurement POVM (as described above) is given by the (parameter-independent) eigenvectors of  $O$ . Therefore, we see that achieving exponential family of states, as defined in Eq. (211), for the output states  $\psi \rightarrow \mathcal{M}_\theta(\psi)$  of the spectator system is generally helpful for our application. Finally, note that for non-full rank parameterized density matrices, the optimal measurements described above are not unique.

### E.2.2 Maximum Likelihood Estimator

The maximum likelihood estimator  $\hat{\theta}_{\text{MLE}}$  corresponding to the choice of POVM  $\Pi \equiv \{\Pi_x\}_{x \in \mathcal{X}}$  is defined as

$$\hat{\theta}_{\text{MLE}}(x) := \arg \max_{\theta} p(x|\theta), \quad (213)$$

where  $p(x|\theta) = \text{Tr}[\rho_\theta \Pi_x]$  for all  $x \in \mathcal{X}$ . Although the above definition seems natural, the MLE is known to

be a biased estimator for a general parametric family  $\{\rho_\theta\}_{\theta \in \Theta}$ . However, the MLE becomes unbiased, and further, saturates the classical CRB in the asymptotic limit [132]. We recall that a necessary condition for the saturation of the QCRB is that the classical and quantum Fisher informations must coincide [65]. Hence, the MLE is also relevant for the asymptotic saturation of the QCRB. In the context of our article, the asymptotic limit necessarily implies that the spatial dependence of the noise parameter cannot be neglected, as we are performing quantum parameter estimation on a large number of spectator qubits that must be spatially distributed within the quantum memory device. Therefore, to retain the spatial homogeneity assumption of the noise parameters used in the main text, we refrain from considering the asymptotic saturation of the QCRB. Hence, the MLE choice is inappropriate within the context of our manuscript, as it is a biased estimator in the non-asymptotic regime.

### E.3 Partial QFIM

Now we consider the bipartition of the parameter space  $\Theta^p$  as  $\theta = (\theta^I, \theta^N)$ , with the number of parameters in each partition is given by  $p_I$  and  $p_N = p - p_I$ , respectively. Here, the subscripts ‘‘I’’ and ‘‘N’’ stand for ‘‘interest’’ and ‘‘nuisance’’, respectively. We can then write the  $p \times p$  SLD QFIM in a block form

$$\mathbf{l}^\theta = \begin{pmatrix} \mathbf{l}_{I,I}^\theta & \mathbf{l}_{I,N}^\theta \\ \mathbf{l}_{N,I}^\theta & \mathbf{l}_{N,N}^\theta \end{pmatrix}, \quad (214)$$

where the upper block diagonal matrix  $\mathbf{l}_{I,I}^\theta$  is  $p_I \times p_I$  and the lower block diagonal matrix  $\mathbf{l}_{N,N}^\theta$  is of size  $p_N \times p_N$ . We also write the inverse of the SLD QFIM in a similar block form

$$\left( \mathbf{l}^\theta \right)^{-1} = \begin{pmatrix} \mathbf{l}^{\theta;I,I} & \mathbf{l}^{\theta;I,N} \\ \mathbf{l}^{\theta;N,I} & \mathbf{l}^{\theta;N,N} \end{pmatrix}. \quad (215)$$

Using these block forms, the partial SLD QFIM is defined as

$$\mathbf{l}_{I|N}^\theta := \left( \mathbf{l}^{\theta;I,I} \right)^{-1} = \mathbf{l}_{I,I}^\theta - \mathbf{l}_{I,N}^\theta \left( \mathbf{l}_{N,N}^\theta \right)^{-1} \mathbf{l}_{N,I}^\theta. \quad (216)$$

Let us further define the  $p_I \times p_I$  MSE matrix for the parameters of interest  $\theta_I$  as

$$\text{Var} \left[ \hat{\theta}_I(X) \right] := \mathbb{E} \left[ \left( \hat{\theta}_\alpha(X) - \theta_\alpha \right) \left( \hat{\theta}_\beta(X) - \theta_\beta \right) \right]_{p(x|\theta)}, \quad (217)$$

for  $\alpha, \beta = 1, \dots, p_I$ . The pair  $(\Pi, \hat{\theta}_I)$  is said to be a locally unbiased estimator for  $\theta_I$  at  $\theta$  when it satisfies the following two conditions [63] (analogous to Eq. (200) and (201))

$$\mathbb{E} \left[ \hat{\theta}_I(X) \right]_{p(x|\theta)} = \sum_{x \in \mathcal{X}} \hat{\theta}_I(x) \text{Tr}[\rho_\theta \Pi_x] = \theta_I, \quad (218)$$

and

$$\frac{\partial}{\partial \theta_\beta} \mathbb{E} \left[ \hat{\theta}_\alpha(X) \right]_{p(x|\theta)} = \sum_{x \in \mathcal{X}} \hat{\theta}_\alpha(x) \text{Tr} \left[ \frac{\partial}{\partial \theta_\beta} \rho_\theta \Pi_x \right] = \delta_{\alpha\beta}, \quad (219)$$

where  $\alpha = 1, \dots, p_I$  and  $\beta = 1, \dots, p$ . For locally unbiased estimators, the following QCRB holds in the presence of nuisance parameters

$$\text{Var} \left[ \hat{\theta}_I(X) \right] \geq \left[ \mathbf{I}_{I|N}^\theta \right]^{-1}, \quad (220)$$

which is a modification of the standard QCRB when nuisance parameters are present.

## F Choi Matrix Methods for Entanglement Fidelity of Composite Parameterized Channels

In what follows, we present a useful lemma for the entanglement fidelity of composite parameterized channels and then dedicate the rest of this appendix to demonstrating its wide range of applicability in the context of the main text.

**Lemma 8.** *The entanglement fidelity of the composite channel  $\mathcal{R}^{B \rightarrow A} \circ \mathcal{N}^{A \rightarrow B}$  is given by the individual Choi states  $\Phi_{AB}^{\mathcal{N}}$  and  $\Phi_{BA}^{\mathcal{R}}$ , as follows*

$$F_e(\mathcal{R} \circ \mathcal{N}) = \frac{d_B}{d_A} \text{Tr} \left[ \Phi^{\mathcal{N}} (\Phi^{\mathcal{R}})^T \right], \quad (221)$$

where  $T$  indicates matrix transposition.

*Proof.* By definition, the entanglement fidelity of the composite channel can be written in terms of its Choi matrix, as follows

$$\begin{aligned} F_e \left( \mathcal{R}^{B \rightarrow A'} \circ \mathcal{N}^{A \rightarrow B} \right) &= \text{Tr}_{AA'} \left[ \Phi_{AA'} \Phi_{AA'}^{\mathcal{R} \circ \mathcal{N}} \right] \\ &= \frac{1}{d_A^2} \text{Tr}_{AA'} \left[ \Gamma_{AA'} \Gamma_{AA'}^{\mathcal{R} \circ \mathcal{N}} \right]. \end{aligned} \quad (222)$$

where  $\mathcal{H}^A$  and  $\mathcal{H}^{A'}$  are isomorphic Hilbert spaces. One can easily verify that the Choi matrix of the composite channel can be written in terms of the Choi matrices of the individual channels, as follows [115]

$$\Gamma_{AC}^{\mathcal{R}^{B \rightarrow C} \circ \mathcal{N}^{A \rightarrow B}} = \text{Tr}_B \left[ \mathbb{T}_B \left( \Gamma_{AB}^{\mathcal{N}} \Gamma_{BC}^{\mathcal{R}} \right) \right], \quad (223)$$

where  $\mathbb{T}_B$  is the partial transpose defined with respect to the same basis as the maximally entangled state  $|\Gamma\rangle$ . By substituting this form into the entanglement fidelity formula, we arrive at

$$F_e = \frac{1}{d_A^2} \text{Tr}_{AA'} \left[ \Gamma_{AA'} \text{Tr}_B \left[ \mathbb{T}_B \left( \Gamma_{AB}^{\mathcal{N}} \Gamma_{BA}^{\mathcal{R}} \right) \right] \right] \quad (224)$$

$$= \frac{1}{d_A^2} \text{Tr}_{A'B} \left[ \left( \text{Tr}_A \left[ \Gamma_{AA'} \mathbb{T}_B \left( \Gamma_{AB}^{\mathcal{N}} \right) \right] \right) \Gamma_{BA}^{\mathcal{R}} \right]. \quad (225)$$

Next, we make standard simplifications for any  $\Lambda_{AB}$ :

$$\text{Tr}_A \left[ \Gamma_{AA'} \Lambda_{AB} \right] \quad (226)$$

$$= \sum_i \langle i|_A \Gamma_{AA'} \Lambda_{AB} |i\rangle_A \quad (227)$$

$$= \sum_{ij} |i\rangle_{A'} \langle j|_{A'} \langle j|_A \Lambda_{AB} |i\rangle_A \quad (228)$$

$$= \sum_{ij} |i\rangle_{A'} \langle j|_{A'} \langle i|_A \mathbb{T}_A \left( \Lambda_{AB} \right) |j\rangle_A \quad (229)$$

$$= \sum_i |i\rangle_{A'} \langle i|_A \mathbb{T}_A \left( \Lambda_{AB} \right) \sum_j |j\rangle_A \langle j|_{A'} \quad (230)$$

$$= \mathbb{T}_{A'} \left( \Lambda_{A'B} \right). \quad (231)$$

This yields for  $\Lambda_{AB} \equiv \mathbb{T}_B \left( \Gamma_{AB}^{\mathcal{N}} \right)$  the following

$$\text{Tr}_A \left[ \Gamma_{AA'} \mathbb{T}_B \left( \Gamma_{AB}^{\mathcal{N}} \right) \right] = \mathbb{T}_{A'} \left( \mathbb{T}_B \left( \Gamma_{A'B}^{\mathcal{N}} \right) \right) \quad (232)$$

$$= \left( \Gamma_{A'B}^{\mathcal{N}} \right)^T. \quad (233)$$

Substituting back into the entanglement fidelity formula completes the proof.  $\blacksquare$

Therefore, according to this lemma, the entanglement fidelity of the memory dynamics is generally written as follows for any recovery map

$$F_e(\mathcal{R} \circ \mathcal{N}_\theta) = \frac{1}{d_A^2} \text{Tr}_{AB} \left[ \left( \Gamma_{AB}^{\mathcal{N}_\theta} \right)^T \Gamma_{BA}^{\mathcal{R}} \right] \quad (234)$$

$$= \frac{1}{d_A^2} \text{Tr}_{AB} \left[ \Gamma_{AB}^{\mathcal{N}_\theta} \left( \Gamma_{BA}^{\mathcal{R}} \right)^T \right]. \quad (235)$$

In particular, this implies that we can search for a recovery map  $\mathcal{R}_\phi$  that is parameterized by some number of parameters  $\phi$  and maximize over them (for a fixed  $\theta$ ) to arrive at an optimal choice  $\phi(\theta)$ , see e.g. [90] in terms of the natural representation of quantum channels (which is related, but not the same as, the Choi representation adopted in this article).

### F.1 Zeroth Derivative: Hölder Type Upper Bounds

A useful upper bound on the entanglement fidelity in Eq. (40) of the quantum memory dynamics can be given in terms of the Choi state of the recovery map, by applying the Hölder inequality [115], as follows

$$F_e(\mathcal{R} \circ \mathcal{N}_\theta) = \frac{d_B}{d_A} \left| \text{Tr}_{AB} \left[ \Phi_{AB}^{\mathcal{N}_\theta} \left( \Phi_{BA}^{\mathcal{R}} \right)^T \right] \right| \quad (236)$$

$$\leq \frac{d_B}{d_A} \text{Tr}_{AB} \left| \Phi_{AB}^{\mathcal{N}_\theta} \left( \Phi_{BA}^{\mathcal{R}} \right)^T \right| \quad (237)$$

$$\leq \frac{d_B}{d_A} \left\| \Phi_{AB}^{\mathcal{N}_\theta} \right\|_\alpha \times \left\| \left( \Phi_{BA}^{\mathcal{R}} \right)^T \right\|_\beta \quad (238)$$

$$= \frac{d_B}{d_A} \left\| \Phi_{AB}^{\mathcal{N}_\theta} \right\|_\alpha \times \left\| \Phi_{BA}^{\mathcal{R}} \right\|_\beta, \quad (239)$$

for  $\alpha \in [1, \infty)$  and its Hölder dual  $\beta$  defined via  $1/\alpha + 1/\beta = 1$ . Here,  $\|X\|_\alpha := (\text{Tr}[|X|^\alpha])^{1/\alpha}$  defines the Schatten norms for  $\alpha \in [1, \infty)$ .

## F.2 First Derivative: Bound on Robustness of a Recovery Map

Consider, for a given recovery map  $\mathcal{R}$ , the partial derivatives of the entanglement of fidelity in Eq. (234) with respect to the components of the noise parameter vector  $\theta$ . Using the inner product definition in Eq. (196) and the definition of SLD operator for the Choi state of  $\mathcal{N}_\theta$ , as given in Eq. (197), we have for the partial derivatives

$$\begin{aligned} & \frac{\partial}{\partial \theta_\alpha} F_e(\mathcal{R} \circ \mathcal{N}_\theta) \\ &= \frac{1}{d_A^2} \text{Tr}_{AB} \left[ \frac{\partial}{\partial \theta_\alpha} \Gamma_{AB}^{\mathcal{N}_\theta} (\Gamma_{BA}^{\mathcal{R}})^T \right] \end{aligned} \quad (240)$$

$$= \frac{d_B}{d_A} \text{Tr}_{AB} \left[ \frac{1}{2} \{ \Phi_{AB}^{\mathcal{N}_\theta}, L_{AB}^{\theta;\alpha} \} (\Phi_{BA}^{\mathcal{R}})^T \right] \quad (241)$$

$$= \frac{d_B}{d_A} \text{Tr}_{AB} \left[ \Phi_{AB}^{\mathcal{N}_\theta} \frac{1}{2} \{ (\Phi_{BA}^{\mathcal{R}})^T, L_{AB}^{\theta;\alpha} \} \right] \quad (242)$$

$$\equiv \frac{d_B}{d_A} \left\langle (\Phi_{BA}^{\mathcal{R}})^T, L_{AB}^{\theta;\alpha} \right\rangle_{\Phi_{AB}^{\mathcal{N}_\theta}}, \quad (243)$$

where  $d_B = 2^n$  and  $d_A = 2^k$  for a generic  $[n, k]$  QEC code, therefore  $d_B/d_A = 2^{n-k}$ . Using the fact that the robustness of  $\mathcal{R}$  (which is the derivative of the entanglement fidelity with respect to the parameters pertaining to  $\mathcal{N}_\theta$ ) is written in terms of the inner product in Eq. (196), we can use the Cauchy-Schwartz inequality to arrive at an upper bound, as follows

$$\begin{aligned} \left| \left\langle (\Phi_{BA}^{\mathcal{R}})^T, L_{AB}^{\theta;\alpha} \right\rangle_{\Phi_{AB}^{\mathcal{N}_\theta}} \right|^2 &\leq \left\langle (\Phi_{BA}^{\mathcal{R}})^T, (\Phi_{BA}^{\mathcal{R}})^T \right\rangle_{\Phi_{AB}^{\mathcal{N}_\theta}} \times \\ &\quad \times \left\langle L_{AB}^{\theta;\alpha}, L_{AB}^{\theta;\alpha} \right\rangle_{\Phi_{AB}^{\mathcal{N}_\theta}} \quad (244) \\ &= \left\langle (\Phi_{BA}^{\mathcal{R}})^T, (\Phi_{BA}^{\mathcal{R}})^T \right\rangle_{\Phi_{AB}^{\mathcal{N}_\theta}} \times \\ &\quad \times [I_{\text{QF}}(\theta; \{ \Phi^{\mathcal{N}_\theta} \})]_{\alpha\alpha}, \end{aligned} \quad (245)$$

which yields the upper bound

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_\alpha} F_e(\mathcal{R} \circ \mathcal{N}_\theta) \right| &\leq \frac{d_B}{d_A} \sqrt{\left\langle (\Phi_{BA}^{\mathcal{R}})^T, (\Phi_{BA}^{\mathcal{R}})^T \right\rangle_{\Phi_{AB}^{\mathcal{N}_\theta}} \times} \\ &\quad \times \sqrt{[I_{\text{QF}}(\theta; \{ \Phi^{\mathcal{N}_\theta} \})]_{\alpha\alpha}}. \end{aligned} \quad (246)$$

## F.3 Proof of Theorem 1: Bounds on Spectator-Based Recovery For Finite Estimation Errors

Schatten norms  $\|X\|_\alpha := (\text{Tr}[|X^\alpha|])^{1/\alpha}$ , where  $\alpha \in [1, \infty)$ , are often used to bound trace quantities, via the well-known Hölder inequality [115]

$$\|XZ\|_1 \leq \|X\|_\alpha \times \|Z\|_\beta, \quad (247)$$

where  $1/\alpha + 1/\beta = 1$  (a pair  $(\alpha, \beta)$  satisfying this equality is called an Hölder pair). It is easy to show that the following statement also applies [115]

$$|\text{Tr}[XZ]| \leq \|X\|_\alpha \times \|Z\|_\beta. \quad (248)$$

For  $\alpha \in [0, 1)$ , the Schatten norm  $\|\cdot\|_\alpha$  is no longer a norm (e.g. it does not satisfy the triangle inequality). However, if  $Z > 0$  (along with  $0 \leq \alpha < 1$ ), then the above inequality is reversed [133] (the Hölder dual  $\beta$  becomes negative)

$$|\text{Tr}[XZ]| \geq \|X\|_\alpha \times \|Z\|_\beta. \quad (249)$$

We can use this inequality to find a lower bound on the difference between the entanglement fidelities of any two recovery maps  $\mathcal{R}^{B \rightarrow A}$ ,  $\tilde{\mathcal{R}}^{B \rightarrow A}$  for the parameterized noise channel  $\mathcal{N}_\theta^{A \rightarrow B}$ , as follows

$$\begin{aligned} & |F_e(\mathcal{R} \circ \mathcal{N}_\theta) - F_e(\tilde{\mathcal{R}} \circ \mathcal{N}_\theta)| \\ &= \left| \text{Tr}_{AA'} \left[ \Phi_{AA'} \left( \text{id}^A \otimes (\mathcal{R}^{B \rightarrow A'} - \tilde{\mathcal{R}}^{B \rightarrow A'}) \right) \left( \Phi_{AB}^{\mathcal{N}_\theta} \right) \right] \right| \end{aligned} \quad (250)$$

$$= \left| \text{Tr}_{AB} \left[ \left( \Phi_{AB}^{\mathcal{R}} - \Phi_{AB}^{\tilde{\mathcal{R}}} \right) \left( \Phi_{AB}^{\mathcal{N}_\theta} \right) \right] \right| \quad (251)$$

$$\geq \left\| \Phi_{AB}^{\mathcal{R}} - \Phi_{AB}^{\tilde{\mathcal{R}}} \right\|_\alpha \times \left\| \Phi_{AB}^{\mathcal{N}_\theta} \right\|_\beta. \quad (252)$$

Next, we use a theorem relating the Choi matrix of a quantum channel  $\mathcal{Q}^A$  to its adjoint [134]

$$\Gamma_{AA'}^{\mathcal{Q}^\dagger} = \Gamma_{AA'}^\mathbb{T} (\Gamma_{AA'}^\mathcal{Q})^T \Gamma_{AA'}^\mathbb{T}, \quad (253)$$

where  $\mathbb{T}$  is the partial transpose channel, and its Choi matrix yields a SWAP unitary [134]. Even though  $\|\cdot\|_\alpha$  is not a norm for  $\alpha \in [0, 1)$ , its definition is still invariant with respect to a unitary transformation and transposition [115]. This yields

$$\left\| \Phi_{AB}^{\mathcal{R}} - \Phi_{AB}^{\tilde{\mathcal{R}}} \right\|_\alpha = \left\| \Phi_{AB}^{\mathcal{R}} - \Phi_{AB}^{\tilde{\mathcal{R}}} \right\|_\alpha. \quad (254)$$

Substituting back into Eq. (252) yields the inequality in Theorem 1.

## G Sufficient Condition For a Negligible Remainder Term in Theorem 2

To quantify the ‘‘smallness’’ of  $\hat{\theta} - \theta$ , for the remainder term in Eq. (56) of Theorem 2 to be negligible (given the noise channel  $\mathcal{N}_\theta$ ), we first use Eq. (62) to arrive at a useful bound, as follows:

$$\begin{aligned} & \left| \frac{1}{3!} \partial_\nu^3 F_e(\mathcal{R}_{\theta+\nu_0} \circ \mathcal{N}_\theta) (\hat{\theta} - \theta)^3 \right| \\ &= \frac{d_B}{3!d_A} \left| \text{Tr}_{AB} \left[ \left( \Phi_{AB}^{\mathcal{N}_\theta} \right)^T \partial_\theta^3 \Phi_{BA}^{\mathcal{R}_\theta} \right] \right| |\hat{\theta} - \theta|^3 \end{aligned} \quad (255)$$

$$\leq \frac{d_B}{3!d_A} \text{Tr}_{AB} \left[ \left| \left( \Phi_{AB}^{\mathcal{N}_\theta} \right)^T \partial_\theta^3 \Phi_{BA}^{\mathcal{R}_\theta} \right| \right] |\hat{\theta} - \theta|^3 \quad (256)$$

$$\leq \frac{d_B}{3!d_A} \left\| \left( \Phi_{AB}^{\mathcal{N}_\theta} \right)^T \right\|_1 \left\| \partial_\theta^3 \Phi_{BA}^{\mathcal{R}_\theta} \right\|_\infty |\hat{\theta} - \theta|^3 \quad (257)$$

$$= \frac{d_B}{3!d_A} \left\| \Phi_{AB}^{\mathcal{N}_\theta} \right\|_1 \left\| \partial_\theta^3 \Phi_{BA}^{\mathcal{R}_\theta} \right\|_\infty |\hat{\theta} - \theta|^3 \quad (258)$$

$$= \frac{d_B}{3!d_A} \left\| \partial_\theta^3 \Phi_{BA}^{\mathcal{R}_\theta} \right\|_\infty |\hat{\theta} - \theta|^3, \quad (259)$$

where we have used the Hölder inequality [115], the invariance of the trace norm  $\|\cdot\|_1$  under transposition, and the unit trace of the Choi state  $\Phi_{AB}^{\mathcal{N}_\theta}$ , in the third, fourth, and fifth lines, respectively. Then, the above bound implies that the remainder term in Eq. 56 of Theorem 2 is negligible if the following sufficient condition holds

$$\frac{d_B}{3!d_A} \left\| \partial_\theta^3 \Phi_{BA}^{\mathcal{R}_\theta} \right\|_\infty \mathbb{E} \left[ |\hat{\theta} - \theta|^3 \right]_{p(x|\theta)} \ll \frac{g(\theta)}{I_{\text{QF}}(\mathcal{M}_\theta)}. \quad (260)$$

We can rewrite this condition as

$$\begin{aligned} \mathbb{E} \left[ |\hat{\theta} - \theta|^3 \right]_{p(x|\theta)} I_{\text{QF}}(\mathcal{M}_\theta(\psi)) &\ll \frac{6d_A g(\theta)}{d_B \left\| \partial_\theta^3 \Phi_{BA}^{\mathcal{R}_\theta} \right\|_\infty} \\ &\equiv g'(\theta), \end{aligned} \quad (261)$$

where the right-hand-side  $g'(\theta)$  is fully determined by the parameterized noise channel  $\mathcal{N}_\theta^{A \rightarrow B}$  (since  $\mathcal{R}_\theta^{B \rightarrow A}$  is found from Eq. (40), given  $\mathcal{N}_\theta^{A \rightarrow B}$ ), similar to  $g(\theta)$  in Theorem 2. The bound in Eq. (261) could be understood by saying that, although  $|\hat{\theta} - \theta|$  cannot be arbitrarily small (due to Eq. (61)), it should not be too large so that the small error expansion that is important for the proof of Theorem 2 will hold.

## H Entanglement Fidelity For The [4,1] Code

The entanglement fidelity for single-qubit ( $k = 1$ )  $[n, k]$  codes is given by [79]

$$F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta) = \frac{1}{4} \text{Tr}[G], \quad (262)$$

where the matrix elements  $G_{\sigma\sigma'} = \text{Tr} \left[ D_{\hat{\theta}, \sigma} \mathcal{N}_\theta[\sigma'_L] \right]$ , with  $D_{\hat{\theta}, \sigma} \equiv 2 \sum_i R_{\hat{\theta}, \sigma}^{(4)} \sigma_L R_{\hat{\theta}, \sigma}^{(4)\dagger}$ , describes the effective dynamics of the Bloch coefficients for the encoded single qubit [80], i.e. if  $\rho_i = \frac{1}{2} \sum_\sigma u_\sigma \sigma$  and  $\rho_f = \frac{1}{2} \sum_\sigma v_\sigma \sigma$ , then  $\vec{v} = G\vec{u}$ .

In [79], the authors derived an analytical formula for the entanglement fidelity  $F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta)$  as

$$F_e = \frac{1}{4} \left[ 1 + \sqrt{2} \text{Re}[\alpha] \tau + 8\tau^2 + (\sqrt{2} \text{Re}[\beta] - 8)\tau^3 + \tau^4 \right], \quad (263)$$

where  $\tau = 1 - \theta$  and  $\alpha, \beta$  (with  $|\alpha|^2 + |\beta|^2 = 1$ ) are the complex parameters that the recovery channel depends on. The optimum recovery channel  $\mathcal{R}_\theta(\alpha, \beta) = \mathcal{R}(\alpha(\theta), \beta(\theta))$  in [12] is the one that maximizes the entanglement fidelity with respect to  $\alpha, \beta$  for the given value of the noise parameter  $\theta$ .

To find the dependence of  $\alpha$  and  $\beta$  on the noise parameter  $\theta$  for the optimum recovery, we first rewrite Eq. (263) using  $\alpha = |\alpha|e^{i\psi}$  and  $\beta = |\beta|e^{i\phi} = \sqrt{1 - |\alpha|^2}e^{i\phi}$ , which yields

$$F_e(|\alpha|, \psi, \phi; \theta) = \frac{1}{4} + \frac{\sqrt{2}}{4} |\alpha| \tau \cos \psi + 2\tau^2$$

$$+ (\sqrt{2(1 - |\alpha|^2)} \cos \phi - 8) \frac{\tau^3}{4} + \frac{\tau^4}{4}. \quad (264)$$

Then we take the partial derivatives of this function with respect to the independent parameters  $|\alpha| \in [0, 1]$  and  $\psi, \phi \in [0, 2\pi)$ , to arrive at

$$|\alpha_{\text{opt}}(\theta)| = \frac{1}{\sqrt{1 + \tau^4}} \quad \text{and} \quad (\psi_{\text{opt}}, \phi_{\text{opt}}) = \{(0, 0), (\pi, \pi)\}. \quad (265)$$

By simple substitution, we can check that  $(\psi, \phi) = (0, 0)$  is the pair that maximizes the entanglement fidelity function.

Now let us find an analytical formula for the incomplete knowledge scenario  $F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta)$ . Note that in Eq. (263), the dependence of the recovery  $\mathcal{R}_\theta(\alpha, \beta) = \mathcal{R}(\alpha(\hat{\theta}), \beta(\hat{\theta}))$  on the estimated noise parameter  $\hat{\theta}$  enters only through  $\alpha$  and  $\beta$  [12]. Therefore, we can use the optimum values  $|\alpha_{\text{opt}}(\hat{\theta})| = 1/\sqrt{1 + \tau(\hat{\theta})}$  and  $\psi_{\text{opt}} = \phi_{\text{opt}} = 0$  and plug it back into Eq. (263), which yields

$$\begin{aligned} F_e(|\alpha_{\text{opt}}(\hat{\theta})|; \theta) &= \frac{1}{4} + \frac{\sqrt{2}}{4} |\alpha_{\text{opt}}(\hat{\theta})| \tau(\theta) + 2\tau^2(\theta) \\ &+ \left( \sqrt{2(1 - |\alpha_{\text{opt}}(\hat{\theta})|^2)} - 8 \right) \frac{\tau^3(\theta)}{4} + \frac{\tau^4(\theta)}{4}. \end{aligned} \quad (266)$$

This yields for arbitrary finite differences  $\theta - \hat{\theta}$  and any estimate  $\hat{\theta}$  the following exact formula

$$\begin{aligned} &F_e(|\alpha_{\text{opt}}(\theta)|; \theta) - F_e(|\alpha_{\text{opt}}(\hat{\theta})|; \theta) \\ &= \frac{\tau(\theta)}{2\sqrt{2}} \left( \frac{1}{\sqrt{1 + \tau^4(\theta)}} - \frac{1}{\sqrt{1 + \tau^4(\hat{\theta})}} \right) \\ &+ \frac{\tau^3(\theta)}{2\sqrt{2}} \left( \frac{\tau^2(\theta)}{\sqrt{1 + \tau^4(\theta)}} - \frac{\tau^2(\hat{\theta})}{\sqrt{1 + \tau^4(\hat{\theta})}} \right). \end{aligned} \quad (267)$$

When adaptation is implemented, the following derivative is relevant

$$\left. \frac{d^2 F_e(|\alpha_{\text{opt}}(\theta + \nu)|; \theta)}{d\nu^2} \right|_{\nu=0} = - \frac{\tau^3(\theta)}{\sqrt{2}(1 + \tau^4(\theta))^{3/2}}. \quad (268)$$

On the other hand, if no adaptation is implemented, the relevant derivative becomes

$$\left. \frac{d^2 F_e(|\alpha_{\text{opt}}(\eta + \nu)|; \theta)}{d\nu^2} \right|_{\nu=0} = - \frac{\tau^3(\eta) q \left( \tau(\eta), \frac{\tau(\theta)}{\tau(\eta)} \right)}{\sqrt{2}(1 + \tau^4(\eta))^{3/2}}, \quad (269)$$

where

$$q(x, y) \equiv \frac{\left(\frac{5}{2}y^3 - \frac{3}{2}y\right)x^4 + \left(\frac{3}{2}y - \frac{1}{2}\right)}{x^4 + 1}, \quad (270)$$

which satisfies  $q(x, 1) = 1$  in the adaptive regime  $\eta = \theta$ . Therefore, further algebraic simplifications yield

$$F_e(|\alpha_{\text{opt}}(\theta)|; \theta) - F_e(|\alpha_{\text{opt}}(\hat{\theta})|; \theta)$$

$$= \frac{\tau^3(\theta)}{\sqrt{2}(1+\tau^4(\theta))^{3/2}}(\theta - \hat{\theta})^2 - R(\hat{\theta} - \theta), \quad (271)$$

where  $R(\hat{\theta} - \theta) \equiv \frac{1}{3!} \partial_\nu^3 F_e(|\alpha_{\text{opt}}(\theta + \nu)|; \theta)(\hat{\theta} - \theta)^3$  is the Lagrange form of the Taylor series expansion remainder of  $F_e(|\alpha_{\text{opt}}(\theta + \nu)|; \theta)$  with respect to  $\nu$ , where  $\nu_0 \in [0, \hat{\theta} - \theta]$  is a constant. Finally, taking the expectation of both sides with respect to the spectator system's probability distribution function  $p_X(x|\theta)$  (where  $x$  is the measurement outcome of the spectator observable  $X = \sum_{x \in \mathcal{X}} x \Pi_x$ ) yields

$$\begin{aligned} & \mathbb{E} \left[ F_e(|\alpha_{\text{opt}}(\theta)|; \theta) - F_e(|\alpha_{\text{opt}}(\hat{\theta})|; \theta) \right] \\ &= g(\theta) \text{Var}(\hat{\theta}) - \mathbb{E} \left[ R(\hat{\theta} - \theta) \right], \end{aligned} \quad (272)$$

where

$$g(\theta) = \frac{(1 - \theta)^3}{\sqrt{2}(1 + (1 - \theta)^4)^{3/2}}, \quad (273)$$

and we have used the fact that  $\hat{\theta}$  is an unbiased estimate of  $\theta$ .

## I $\chi$ -Matrix Representation of Quantum Channels

Besides the well-known Kraus and Stienspring representations of a CPTP map, a lesser-known representation, called the  $\chi$ -matrix representation [135], is also useful in practice. This is most commonly used in quantum state tomography [37] and is extended to quantum process tomography [136] where state tomography of the Choi state of a quantum channel is conducted. This is to be contrasted with other approaches in measuring noise, such as randomized benchmarking [137] and QEC itself [138]. Interestingly, the  $\chi$ -matrix representation can be well motivated in the context of QEC by noting that we can rewrite the ‘‘error operators’’  $\{Q_i\}$  of any noisy map  $\mathcal{Q}(\cdot) = \sum_i Q_i(\cdot)Q_i^\dagger$  in terms of a pre-selected ‘‘error basis’’  $\{B_k\}_{k=0}^{d^2-1}$  in  $\mathcal{L}(\mathcal{H})$ , where  $d \equiv \dim \mathcal{H}$ . It is particularly useful to pick one of the basis elements, e.g.  $B_0$ , as the ‘‘desirable’’ error (such as being proportional to the unit matrix). Consequently, the coefficient associated with this error component indicates how likely it is that the given Kraus operators of the noisy map will change the state of our quantum system in a ‘‘desirable way’’. An additional benefit of the  $\chi$ -matrix representation is that the effects of channel twirling are especially clear [122], as ‘‘diagonalization’’ with respect to the generalized Pauli group. Therefore, the rest of the appendix is devoted to recalling the  $\chi$ -matrix representation in a self-contained way.

Recall that every CP map  $\mathcal{Q}^{A \rightarrow B}$  admits a Kraus decomposition

$$\mathcal{Q}(\cdot) = \sum_{i=1}^K Q_i(\cdot)Q_i^\dagger. \quad (274)$$

in terms of Kraus operators  $\{Q_i\}_{i=1}^K$  satisfying  $\sum_{i=1}^K Q_i^\dagger Q_i \leq I_A$ , where the equality holds for TP maps. Let us consider a CP map  $\mathcal{Q}^{A \rightarrow A} \equiv \mathcal{Q}$ , where by denoting  $d \equiv \dim(\mathcal{H}^A)$ , we can decompose each of the Kraus operators  $\{Q_i\}_{i=1}^K$  as a linear combination of some orthonormal operator basis  $\{B_k\}_{k=0}^{d^2-1}$  in  $\mathcal{L}(\mathcal{H}^A)$ , as follows

$$Q_i = \sum_{k=0}^{d^2-1} \langle B_k, Q_i \rangle B_k, \quad (275)$$

where  $\langle B_k, B_l \rangle = \delta_{kl}$ , and  $\langle \cdot \rangle$  being the Hilbert-Schmidt inner product in  $\mathcal{L}(\mathcal{H}^A)$ . One could take  $B_k \equiv B_{(m,n)} = |m\rangle\langle n|$  for  $m, n = \{1, \dots, d\}$ , which is known as the standard basis in  $\mathcal{L}(\mathcal{H}^A)$ . Substituting Eq. (275) in the Kraus representation of  $\mathcal{Q}$ , we get

$$\mathcal{Q}(\cdot) = \sum_{k=0}^{d^2-1} \sum_{l=0}^{d^2-1} \chi_{kl}^{\mathcal{Q}} B_k(\cdot) B_l^\dagger, \quad (276)$$

where

$$\chi_{kl}^{\mathcal{Q}} := \sum_{i=1}^K \langle B_k, Q_i \rangle \langle Q_i, B_l \rangle, \quad (277)$$

is called the  $\chi$  matrix of the CP map  $\mathcal{Q}$ . It is easy to see that the  $\chi$  matrix is a positive semi-definite matrix. This matrix has  $d^4$  complex entries, corresponding to the matrix entries of the superoperator  $\mathcal{Q}$  in the Liouville representation (see, e.g. [43, 139]), namely

$$\hat{\mathcal{Q}} := \sum_{k=0}^{d^2-1} \sum_{l=0}^{d^2-1} \chi_{kl}^{\mathcal{Q}} |B_k\rangle\rangle \langle\langle B_l|, \quad (278)$$

where  $|B_k\rangle\rangle$  is the  $d^2 \times 1$  vector corresponding to the  $d \times d$  matrix  $B_k$ . The number of independent entries of the  $\chi$  matrix is reduced from  $d^4$  to  $d^4 - d^2$  complex numbers if the CP map  $\mathcal{Q}$  is also TP, since for each of the  $d^2$  standard basis elements  $|n\rangle\langle m|$  in  $\mathcal{L}(\mathcal{H}^A)$ , the map  $\mathcal{Q}$  must also preserve the trace, which leads to  $d^2$  constraints.

In the context of QEC, it is convenient to choose our operator basis in  $\mathcal{L}(\mathcal{H}^A)$  such that  $B_0$  indicates a ‘‘desired effect’’ on a quantum state. Here  $B_0 \equiv I/\sqrt{d}$  is desirable for QEC, but for other applications,  $B_0$  could be chosen differently. Next, we write Eq. (275) for a fixed  $i = 1, \dots, K$ , as

$$Q_i = \langle B_0, Q_i \rangle B_0 + \sum_{k=1}^{d^2-1} \langle B_k, Q_i \rangle B_k. \quad (279)$$

Then, by multiplying both sides on the left by  $Q_i^\dagger$  and taking the trace, we arrive at

$$\langle Q_i, Q_i \rangle = |\langle B_0, Q_i \rangle|^2 + \sum_{k=1}^{d^2-1} |\langle B_k, Q_i \rangle|^2, \quad (280)$$

or equivalently,

$$\frac{|\langle B_0, Q_i \rangle|^2}{\langle Q_i, Q_i \rangle} + \sum_{k=1}^{d^2-1} \frac{|\langle B_k, Q_i \rangle|^2}{\langle Q_i, Q_i \rangle} = 1. \quad (281)$$

By denoting  $q_i^2 := \langle Q_i, Q_i \rangle$ ,  $|\cos(\phi_i)| := |\langle B_0, Q_i \rangle|/q_i$ , and  $v_{i,k} |\sin(\phi_i)| := |\langle B_k, Q_i \rangle|/q_i$  with some real weights  $\{v_{i,k}\}_{k=1}^{d^2-1}$  satisfying  $\sum_{k=1}^{d^2-1} v_{i,k}^2 = 1$ , we can rewrite the previous equation in a simple form

$$\cos^2(\phi_i) + \sin^2(\phi_i) = 1 \quad \text{for for all } i = 1, \dots, K, \quad (282)$$

where the angle  $\phi_i$  indicates how ‘‘close’’ the error  $Q_i$  is the the ‘‘desirable error’’  $B_0$ . Note that, if  $\mathcal{Q}$  is also TP, then  $\sum_{i=1}^K q_i^2 = \sum_{i=1}^K \langle Q_i, Q_i \rangle = \text{Tr}[I] = d$ .

Due to the unitary freedom of choosing the Kraus operators of any fixed quantum channel from its Stein-spring dilation [37], the  $\chi$  matrix is not uniquely determined. Therefore, using the phase freedom  $Q_i \rightarrow Q_i e^{i\omega_i}$  (which is a special case of the unitary freedom mentioned above), we can always choose the phases  $\omega_i$  for  $i = 1, \dots, K$  such that all the inner products  $\langle B_0, Q_i \rangle$  with the basis element  $B_0$  are all non-negative. This means that we can pick  $\phi_i \in [0, \pi/2]$ . Finally, the additional phase in  $\langle B_k, Q_i \rangle$  can always be placed in the vector  $v_k$ , which leads to the final decomposition

$$Q_i = q_i \left( \cos(\phi_i) B_0 + \sin(\phi_i) \sum_{k=1}^{d^2-1} v_{i,k} B_k \right), \quad (283)$$

where  $\{v_{i,k}\}_{k=1}^{d^2-1}$  are now complex numbers with  $\sum_{k=1}^{d^2-1} |v_{i,k}|^2 = 1$ . From here, we can easily compute the matrix element  $\chi_{00}^{\mathcal{Q}}$ , after taking  $B_0 = I/\sqrt{d}$ , as

$$\chi_{00}^{\mathcal{Q}} = \sum_{i=1}^K q_i^2 \cos^2(\phi_i) = \frac{1}{d} \sum_{i=1}^K |\text{Tr}[Q_i]|^2, \quad (284)$$

which is consistent with  $\chi_{00}^{\mathcal{Q}}/d = F_e(\mathcal{Q}, I/d)$  [7, 8]. Using Eq. (283) to compute  $Q_i^\dagger Q_i$ , taking the trace, and using the orthonormality of the operator basis  $\{B_k\}_{k=1}^{d^2-1}$ , we arrive at

$$\sum_{i=1}^K \langle Q_i, Q_i \rangle = \sum_{i=1}^K q_i^2 \leq d, \quad (285)$$

where the inequality follows from  $\sum_i Q_i^\dagger Q_i \leq I$ . Combined with Eq. (284), this implies that  $0 \leq \chi_{00}^{\mathcal{Q}} \leq d$ , or equivalently  $0 \leq F_e(\mathcal{Q}) \leq 1$ .

## J Proof of Lemma 3

Here we derive an upper bound on the matrix element  $\chi_{00}^{\mathcal{S} \circ \mathcal{Q}}$  of the composite channel  $\mathcal{S} \circ \mathcal{Q}$ , given the corresponding  $\chi$  matrix elements  $\chi_{00}^{\mathcal{Q}}$  and  $\chi_{00}^{\mathcal{S}}$  of

the individual channels  $\mathcal{Q}$  and  $\mathcal{S}$ , respectively. The technique used for the following derivation is based on [43].

Given the Kraus operators  $\{Q_i\}_{i=1}^{K(\mathcal{Q})}$ ,  $\{S_j\}_{j=1}^{K(\mathcal{S})}$  of the individual channels  $\mathcal{Q}$  and  $\mathcal{S}$ , the Kraus operators of the composite channel  $\mathcal{S} \circ \mathcal{Q}$  are given by  $\{S_j Q_i\}_{i,j}$  for  $i = 1, \dots, K(\mathcal{Q})$  and  $j = 1, \dots, K(\mathcal{S})$ . Therefore, by using Eq. (283) for the individual Kraus operators and using the notation  $\langle Q_i, Q_i \rangle = q_i$  and  $\langle S_i, S_i \rangle = s_i$ , we find for the Kraus operators of the composite channel

$$\begin{aligned} S_j Q_i &= s_j q_i \cos(\phi_j^{\mathcal{S}}) \cos(\phi_i^{\mathcal{Q}}) B_0^2 \\ &+ s_j q_i \sin(\phi_j^{\mathcal{S}}) \cos(\phi_i^{\mathcal{Q}}) \sum_{k=1}^{d^2-1} v_{j,k}^{\mathcal{S}} B_k B_0 \\ &+ s_j q_i \cos(\phi_j^{\mathcal{S}}) \sin(\phi_i^{\mathcal{Q}}) \sum_{k=1}^{d^2-1} v_{i,k}^{\mathcal{Q}} B_0 B_k \\ &+ s_j q_i \sin(\phi_j^{\mathcal{S}}) \sin(\phi_i^{\mathcal{Q}}) \sum_{k,k'=1}^{d^2-1} v_{j,k}^{\mathcal{S}} v_{i,k'}^{\mathcal{Q}} B_k B_{k'}. \end{aligned} \quad (286)$$

By substituting into Eq. (284) and choosing the operator basis elements to be Hermitian (hence  $\text{Tr}[B_k B_{k'}] = \langle B_k, B_{k'} \rangle = \delta_{kk'}$ , for  $k, k' = 0, 1, \dots, n^2 - 1$ ), we arrive at

$$\begin{aligned} \chi_{00}^{\mathcal{S} \circ \mathcal{Q}} &= \frac{1}{d} \sum_{i,j} s_j^2 q_i^2 |\cos(\phi_j^{\mathcal{S}}) \cos(\phi_i^{\mathcal{Q}}) \\ &+ (v_j^{\mathcal{S}} \bullet v_i^{\mathcal{Q}}) \sin(\phi_j^{\mathcal{S}}) \sin(\phi_i^{\mathcal{Q}})|^2, \end{aligned} \quad (287)$$

where we have denoted by  $v_j^{\mathcal{S}} \bullet v_i^{\mathcal{Q}} \equiv \sum_{k=1}^{d^2-1} v_{j,k}^{\mathcal{S}} v_{i,k}^{\mathcal{Q}}$ , with  $\{v_{i,k}^{\mathcal{Q}}\}_{k=0}^{d^2-1}$  and  $\{v_{j,k}^{\mathcal{S}}\}_{k=0}^{d^2-1} \in \mathbb{C}^{d^2-1}$ . By denoting  $c_{ij} \equiv \cos(\phi_j^{\mathcal{S}}) \cos(\phi_i^{\mathcal{Q}})$ ,  $s_{ij} \equiv \sin(\phi_j^{\mathcal{S}}) \sin(\phi_i^{\mathcal{Q}})$ , and  $v_{ij} \equiv v_j^{\mathcal{S}} \bullet v_i^{\mathcal{Q}}$ , we can use the (forward and reversed) triangle inequality, as follows

$$||c_{ij}| - |v_{ij} s_{ij}|| \leq |c_{ij} + v_{ij} s_{ij}| \leq |c_{ij}| + |v_{ij} s_{ij}|. \quad (288)$$

By recalling that  $c_{ij}, s_{ij} \geq 0$ , since  $\phi_i^{\mathcal{Q}}, \phi_j^{\mathcal{S}} \in [0, \pi/2]$ , the second inequality yields  $|c_{ij} + v_{ij} s_{ij}| \leq |c_{ij}| + |v_{ij} s_{ij}| = c_{ij} + |v_{ij} s_{ij}|$ . If we assume that  $|v_{ij}| \leq 1$ , then the inequality becomes  $|c_{ij} + v_{ij} s_{ij}| \leq c_{ij} + s_{ij}$ . By squaring both sides, we get

$$|c_{ij} + v_{ij} s_{ij}|^2 \leq c_{ij}^2 + s_{ij}^2 + 2c_{ij} s_{ij}, \quad (289)$$

where this inequality is saturated iff  $v_{ij} = 1$  for all  $i = 1, \dots, K(\mathcal{Q})$  and  $j = 1, \dots, K(\mathcal{S})$ . Substituting back into Eq. (287) and using the definitions of  $\chi_{00}^{\mathcal{Q}}$  and  $\chi_{00}^{\mathcal{S}}$  from Eq. (284), as well as the fact that  $\sum_i q_i^2 = \sum_j s_j^2 = d$  for CPTP maps  $\mathcal{Q}$  and  $\mathcal{S}$ , we arrive at

$$\chi_{00}^{\mathcal{S} \circ \mathcal{Q}} = \frac{1}{d} \sum_{i,j} s_j^2 q_i^2 |c_{ij} + v_{ij} s_{ij}|^2 \quad (290)$$



$$\leq \frac{1}{d} \sum_{i,j} s_j^2 q_i^2 (c_{ij}^2 + s_{ij}^2 + 2c_{ij}s_{ij}) \quad (291)$$

$$= \frac{1}{d} [\chi_{00}^S \chi_{00}^Q + (d - \chi_{00}^S)(d - \chi_{00}^Q)] + \frac{2}{d} \sum_{i,j} s_j^2 q_i^2 c_{ij} s_{ij} \quad (292)$$

$$= \frac{1}{d} [\chi_{00}^S \chi_{00}^Q + (d - \chi_{00}^S)(d - \chi_{00}^Q)] + \frac{2}{d} \left( \sum_j s_j^2 \cos(\phi_j^S) \sin(\phi_j^S) \right) \times \left( \sum_i q_i^2 \cos(\phi_i^Q) \sin(\phi_i^Q) \right). \quad (293)$$

Next, we use the Cauchy-Schwartz inequality

$$\sum_i q_i^2 \cos(\phi_i^Q) \sin(\phi_i^Q) \quad (294)$$

$$= \sum_i (q_i \cos(\phi_i^Q)) (q_i \sin(\phi_i^Q)) \quad (295)$$

$$\leq \sqrt{\sum_i q_i^2 \cos^2(\phi_i^Q)} \sqrt{\sum_i q_i^2 \sin^2(\phi_i^Q)} \quad (296)$$

$$= \sqrt{\chi_{00}^Q (d - \chi_{00}^Q)}, \quad (297)$$

where the Cauchy-Schwartz inequality is saturated when the vectors  $\{q_i \cos(\phi_i^Q)\}_{i=1}^{K(Q)}$  and  $\{q_i \sin(\phi_i^Q)\}_{i=1}^{K(Q)}$  are linearly dependent, therefore  $\tan(\phi_1^Q) = \dots = \tan(\phi_{K(Q)}^Q)$ . Since  $\phi_i^Q \in [0, \pi/2]$  for all  $i = 1, \dots, K(Q)$  and the function  $\tan(x)$  is one-to-one in that region, it follows that the above inequality is saturated iff  $\phi_1^Q = \dots = \phi_{K(Q)}^Q$ . The exact same argument for the channel  $\mathcal{S}$  yields

$$\sum_j s_j^2 \cos(\phi_j^S) \sin(\phi_j^S) \leq \sqrt{\chi_{00}^S (d - \chi_{00}^S)}, \quad (298)$$

where the inequality is saturated iff  $\phi_1^S = \dots = \phi_{K(S)}^S$ . Substituting Eqs. (297) and (298) into Eq. (293), we arrive at

$$d\chi_{00}^{S \circ Q} \leq \chi_{00}^S \chi_{00}^Q + (d - \chi_{00}^S)(d - \chi_{00}^Q) + 2\sqrt{\chi_{00}^S \chi_{00}^Q (d - \chi_{00}^S)(d - \chi_{00}^Q)}, \quad (299)$$

or equivalently,

$$\sqrt{d\chi_{00}^{S \circ Q}} \leq \sqrt{\chi_{00}^S} \sqrt{\chi_{00}^Q} + \sqrt{d - \chi_{00}^S} \sqrt{d - \chi_{00}^Q}. \quad (300)$$

Dividing both sides by  $d$  and redefining  $\chi_{00}^Q/d \equiv \cos^2(\delta^Q)$  for  $\delta^Q \in [0, \pi/2]$ , as suggested in Eq. (14), we arrive at

$$\cos(\delta^{S \circ Q}) \leq \cos(\delta^S) \cos(\delta^Q) + \sin(\delta^S) \sin(\delta^Q). \quad (301)$$

Using a triangle identity to simplify the right hand side, we get  $\cos(\delta^{S \circ Q}) \leq \cos(\delta^S - \delta^Q)$ , i.e.

$$\delta^{S \circ Q} \geq |\delta^S - \delta^Q|, \quad (302)$$

which allows  $\delta^{S \circ Q} \in [0, \pi/2]$  given  $\delta^S, \delta^Q \in [0, \pi/2]$ . In other words, given  $\chi_{00}^Q$  and  $\chi_{00}^S$ , the composite channel  $\chi$ -matrix element  $\chi_{00}^{S \circ Q}$  is bounded from above by

$$\frac{\chi_{00}^{S \circ Q}}{d} \leq \cos^2 \left( \arccos \sqrt{\frac{\chi_{00}^S}{d}} - \arccos \sqrt{\frac{\chi_{00}^Q}{d}} \right). \quad (303)$$

## K Proof of Theorem 3

We denote by  $F_e \equiv F_e(\mathcal{R}_\theta \circ \mathcal{N}_\theta)$ ,  $\hat{F}_e \equiv \hat{F}_e(\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_\theta)$ , and  $\Delta F_e \equiv F_e - \hat{F}_e$ , and then do the simple manipulation

$$\arccos \sqrt{F_e(\mathcal{R}_{\hat{\theta}} \circ \mathcal{N}_\theta)} = \arccos \sqrt{F_e - \Delta F_e} \quad (304)$$

$$= \arccos \left\{ \sqrt{F_e} \sqrt{1 - \frac{\Delta F_e}{F_e}} \right\} \quad (305)$$

$$= \arccos \left\{ \sqrt{F_e} \left[ 1 - \frac{\Delta F_e}{2F_e} \right] + \sqrt{F_e} R_1 \left( \frac{\Delta F_e}{F_e} \right) \right\}, \quad (306)$$

where

$$R_1(x) \equiv \frac{1}{3}(1 - x_0)^{-3/2} x^2, \quad (307)$$

is the Lagrange remainder term of the Taylor expansion of  $\sqrt{1 - x}$ , and  $x_0 \in [0, x]$  is a constant. This yields

$$\cos \hat{\delta} = \sqrt{F_e} - \frac{\Delta F_e}{2\sqrt{F_e}} + \sqrt{F_e} R_1 \left( \frac{\Delta F_e}{F_e} \right). \quad (308)$$

By denoting  $\delta \equiv \arccos \sqrt{F_e}$ , we find the Taylor series expansion of  $\hat{\delta}$  with respect to  $x \equiv \cos \delta - \cos \hat{\delta}$ , as follows

$$\hat{\delta} = \arccos(\cos \delta - x) = \delta + \frac{x}{\sin \delta} + R_2(x), \quad (309)$$

where

$$R_2(x) = -\frac{\cos \delta - x_0}{2(1 - (\cos \delta - x_0)^2)^{3/2}} x^2, \quad (310)$$

is the Lagrange form of the remainder term, and  $x_0 \in [0, x]$  is a constant. Substituting for  $x$  using Eq. (308) yields

$$\hat{\delta} - \delta = \frac{x}{\sin \delta} + R_2(x) \quad (311)$$

$$= \frac{\frac{\Delta F_e}{2\sqrt{F_e}} - \sqrt{F_e} R_1 \left( \frac{\Delta F_e}{F_e} \right)}{\sin \delta} + R_2(x) \quad (312)$$

$$= \frac{\Delta F_e}{2\sqrt{F_e}(1 - F_e)} + r \left( \frac{\Delta F_e}{F_e} \right), \quad (313)$$

where

$$r(x) = \sqrt{\frac{F_e}{1 - F_e}} R_1(x) + R_2 \left( \frac{\sqrt{F_e}}{2} x - \sqrt{F_e} R_1(x) \right), \quad (314)$$

is the accumulative remainder term.

Next, we rewrite Theorem 4 as a recurrence inequality, where the contribution of the spectator system in each time step is clearly separated.

$$F_e^{1 \rightarrow n} \leq \cos^2 \left( \arccos \sqrt{F_e^{1 \rightarrow (n-1)}} - \arccos \sqrt{\hat{F}_e^n} \right) \quad (315)$$

$$\equiv \cos^2 \left( \delta^{1 \rightarrow (n-1)} - \hat{\delta}^n \right) \quad (316)$$

$$= \cos^2 \left( \delta^{1 \rightarrow (n-1)} - \delta^n - (\hat{\delta}^n - \delta^n) \right). \quad (317)$$

By denoting  $a = \delta^{1 \rightarrow (n-1)} - \delta^n$  and  $b = \hat{\delta}^n - \delta^n$ , and using the trigonometric identities  $\cos^2(a - b) = \frac{1}{2} + \frac{1}{2} \cos 2(a - b)$  and  $\cos(a - b) = \cos(a) \cos(b) + \sin(a) \sin(b)$ , we arrive at

$$\cos^2(a - b) \quad (318)$$

$$= \frac{1}{2} + \frac{1}{2} (\cos 2a \cos 2b + \sin 2a \sin 2b) \quad (319)$$

$$= \cos^2(a) - \frac{1}{2} \cos 2a (1 - \cos 2b) + \frac{1}{2} \sin 2a \sin 2b \quad (320)$$

$$= \cos^2(a) + b \sin(2a) - \frac{1}{2} (\cos 2a) R_3(2b) + \frac{1}{2} (\sin 2a) R_4(2b) \quad (321)$$

$$\equiv \cos^2(a) + b \sin(2a) + r'(b), \quad (322)$$

where we have used the Taylor expansions of the sine and cosine functions to the first order of  $b$ , and denoted by  $R_3(x)$  and  $R_4(x)$  with their remainders in the Lagrange form, respectively.

Consequently, we have for Theorem 4 the separation

$$F_e^{1 \rightarrow n} \leq \cos^2 \left( \delta^{1 \rightarrow (n-1)} - \delta^n \right) + (\hat{\delta}^n - \delta^n) \sin \left( 2(\delta^{1 \rightarrow (n-1)} - \delta^n) \right) + r'(\hat{\delta}^n - \delta^n), \quad (323)$$

where the first term is the interference between the previous  $n - 1$  cycles and the  $n$ -th cycle error angles with perfect knowledge at the  $n$ -th step. On the other hand, the second term shows the contribution of the lack of knowledge into the recurrence inequality at the  $n$ -th timestep, for a fixed error angle  $\delta^{1 \rightarrow (n-1)}$ , as

$$\frac{\Delta F_e^n}{2\sqrt{F_e^n(1 - F_e^n)}} \sin \left( 2\delta^{1 \rightarrow (n-1)} - 2 \arccos \sqrt{F_e^n} \right). \quad (324)$$

Note that the sign of this contribution depends on the difference between the error angles of the previous  $n - 1$  cycles and the  $n$ -th cycle. Taking the expectation with respect to the probability distribution  $p_X(x_n | \theta_n)$  and using Theorem 2, we arrive at the following: