arXiv:2503.22823v3 [quant-ph] 19 May 2026

Quantum Doeblin Coefficients: Interpretations and Applications

Ian George[1], Christoph Hirche[2], Theshani Nuradha[3,4], and Mark M. Wilde[3]

[1]Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore
[2]Institute for Information Processing (tnt/L3S), Leibniz Universität Hannover, Germany
[3]School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14850, USA
[4]Department of Mathematics and Illinois Quantum Information Science and Technology (IQUIST) Center, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

In classical information theory, the Doeblin coefficient of a classical channel provides an efficiently computable upper bound on the total-variation contraction coefficient of the channel, leading to what is known as a strong data-processing inequality. Here, we investigate quantum Doeblin coefficients as a generalization of the classical concept. In particular, we define various new quantum Doeblin coefficients, one of which has several desirable properties, including concatenation and multiplicativity, in addition to being efficiently computable. We also develop various interpretations of two of the quantum Doeblin coefficients, including representations as minimal singlet fractions, exclusion values, reverse max-mutual and oveloH informations, reverse robustnesses, and hypothesis testing reverse mutual and oveloH informations. Our interpretations of quantum Doeblin coefficients as either entanglement-assisted or unassisted exclusion values are particularly appealing, indicating that they are proportional to the best possible error probabilities one could achieve in state-exclusion tasks by making use of the channel. We also outline various applications of quantum Doeblin coefficients, ranging from limitations on quantum machine learning algorithms that use parameterized quantum circuits (noise-induced barren plateaus), on error mitigation protocols, on the sample complexity of noisy quantum hypothesis testing, on the fairness of noisy quantum models, and on mixing, indistinguishability, and decoupling times of time-varying channels. All of these applications make use of the fact that quantum Doeblin coefficients appear in upper bounds on various trace-distance contraction coefficients of a quantum channel. Furthermore, in all of these applications, our analysis using quantum Doeblin coefficients provides improvements of various kinds over contributions from prior literature, both in terms of generality and being efficiently computable.

Ian George: qit.george@gmail.com
Christoph Hirche: christoph.hirche@gmail.com
Theshani Nuradha: theshani.gallage@gmail.com
Mark M. Wilde: wilde@cornell.edu

Contents

1 Introduction **3**
- 1.1 Background . . . 3
- 1.2 Summary of Contributions . . . 5

2 Preliminaries **7**
- 2.1 Quantum Divergences . . . 9
- 2.2 Conic Programming and the Separable Cone . . . 10
- 2.3 Quantum State Exclusion . . . 12

3 Quantum Doeblin Coefficients **13**
- 3.1 Main Definition of Quantum Doeblin Coefficient . . . 13
- 3.2 Alternative Notions of Quantum Doeblin Coefficients . . . 14
  - 3.2.1 Induced and Cone-Restricted Doeblin Coefficients . . . 16

4 Interpretations of the Quantum Doeblin Coefficient **20**
- 4.1 Minimum Achievable Singlet Fraction . . . 20
- 4.2 Exclusion Value of a Quantum Channel . . . 21
- 4.3 Reverse Max-Mutual Information . . . 24
- 4.4 Reverse Robustness . . . 25
- 4.5 Reverse Hypothesis Testing Mutual Information . . . 26
- 4.6 Interpretations of the Induced Quantum Doeblin Coefficient . . . 27
  - 4.6.1 Minimum Achievable Singlet Fraction . . . 27
  - 4.6.2 Max-oveloH Information . . . 28
  - 4.6.3 Reverse Robustness . . . 29
  - 4.6.4 Hypothesis-testing oveloH information . . . 29

5 Properties of Quantum Doeblin Coefficients **30**
- 5.1 Normalization . . . 30
- 5.2 Concatenation . . . 30
- 5.3 Data Processing under Pre- and Post-Processing . . . 31
- 5.4 Submultiplicativity and Multiplicativity . . . 32
- 5.5 Concavity . . . 36
- 5.6 Properties of Alternative Notions of Quantum Doeblin Coefficient . . . 37

6 Contraction and Expansion Coefficients **40**
- 6.1 Contraction Coefficients and Quantum Doeblin Coefficients . . . 40
- 6.2 Expansion Coefficients and Reverse Doeblin Coefficients . . . 46

7 Quantum Doeblin Coefficients for Example Channels **47**
- 7.1 Classical–Quantum Channels . . . 47
- 7.2 Measurement Channels . . . 48
- 7.3 Entanglement-Breaking Channels . . . 50
- 7.4 Generalized Dephasing Channels . . . 51
- 7.5 Qubit Channels . . . 51
- 7.6 Generalized Amplitude-Damping Channels . . . 59

8 Applications of Quantum Doeblin Coefficients **61**
8.1 Noise-Induced Barren Plateaus . . . 62
8.2 Error Mitigation . . . 68
8.3 Noisy Quantum Hypothesis Testing . . . 72
8.4 Fairness of Noisy Quantum Models . . . 76
8.5 Mixing, Indistinguishability, and Decoupling Times of Quantum Processes . . . 77
8.5.1 Mixing Processes and Mixing Times . . . 78
8.5.2 Indistinguishability Times . . . 80
8.5.3 Decoupling Times . . . 82

9 Conclusion **86**
9.1 Summary . . . 86
9.2 Open Questions . . . 87

A Proof of Theorem 2 **95**

B Proof of Proposition 13 **97**

C Proof of Proposition 28 **99**

D Qubit Channel Parameterizations **101**

E Another Alternative Notion of Quantum Doeblin Coefficient **103**

F Extra Proofs on Mixing and Decoupling Times of Quantum Processes **105**
F.1 Proof of Proposition 45 . . . 105
F.2 Proof of Proposition 50 . . . 106

1 Introduction

1.1 Background

Given a classical channel $\mathcal{W}$ described by a conditional probability distribution $\left(p_{Y|X}(y|x)\right)_{y\in\mathcal{Y},x\in\mathcal{X}}$, its Doeblin coefficient is defined as follows [1]:

$$\alpha(\mathcal{W}) \coloneqq \max_{r\in\mathbb{R}_{\geq 0}^{|\mathcal{Y}|}} \left\{ \sum_{y\in\mathcal{Y}} r(y) : r(y) \leq p_{Y|X}(y|x) \ \forall x \in \mathcal{X}, y \in \mathcal{Y} \right\}, \tag{1.1}$$

where $\mathbb{R}_{\geq 0}^{|\mathcal{Y}|}$ denotes the set of all $|\mathcal{Y}|$-dimensional vectors with non-negative entries. As written above, the Doeblin coefficient $\alpha(\mathcal{W})$ is efficiently computable as a linear program. Its linear programming dual is given by

$$\alpha(\mathcal{W}) = \min_{q(x|y)} \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} q(x|y) p_{Y|X}(y|x), \tag{1.2}$$

where the optimization is over every conditional probability distribution $(q(x|y))_{x\in\mathcal{X},y\in\mathcal{Y}}$. Using either of the expressions in (1.1) or (1.2), one can conclude that

$$\alpha(\mathcal{W}) = \sum_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} \left\{ p_{Y|X}(y|x) \right\}. \tag{1.3}$$

Ref. [2] provides an extensive overview of the Doeblin coefficient of classical channels, which has found various applications, including change detection [3], multi-armed bandits [4], Markov-chain Monte Carlo methods [5], Markov decision processes [6], and mixing models [7].

As observed in [2, Theorem 1, Item 8], the formulations in (1.2)–(1.3) provide the Doeblin coefficient with an operational meaning as being proportional to the optimal error probability in an information-processing task known as exclusion [8]. The goal of this task is as follows: given an input $x \in \mathcal{X}$ chosen uniformly at random and transmitted through the channel $\mathcal{W}$ to produce an output $y \in \mathcal{Y}$, upon receiving y, pick $x' \in \mathcal{X}$ such that $x' \neq x$. The most general strategy the receiver can use involves a conditional probability distribution $q(x|y)$, and so one can optimize over all such strategies to conclude that the optimal error probability is $\frac{1}{|\mathcal{X}|}\alpha(\mathcal{W})$. The equality in (1.3) reflects the fact that an optimal strategy is a minimum likelihood decoding rule [9, Section II-A], in which one guesses a value $x' \in \mathcal{X}$ that is least likely to occur.

Beyond its interpretation given above, the Doeblin coefficient finds possibly its greatest application in bounding the total-variation contraction coefficient of a channel, the latter also known as the Dobrushin coefficient [10]. To motivate this point, let us recall that the total-variation contraction coefficient of a classical channel $\mathcal{W}$ is defined as

$$\eta_{\mathrm{TV}}(\mathcal{W}) \coloneqq \sup_{s\neq t} \frac{\|\mathcal{W}\circ s - \mathcal{W}\circ t\|_1}{\|s-t\|_1} \tag{1.4}$$

$$= \max_{x\neq x'} \frac{1}{2} \left\| p_{Y|X=x} - p_{Y|X=x'} \right\|_1, \tag{1.5}$$

where the optimization is over every pair of probability distributions s and t on the input alphabet $\mathcal{X}$, the probability distribution $\mathcal{W}\circ s$ is defined to be $\left(\sum_{x\in\mathcal{X}} p_{Y|X}(y|x)s(x)\right)_{y\in\mathcal{Y}}$, and $\frac{1}{2}\|s-t\|_1 = \frac{1}{2}\sum_{x\in\mathcal{X}} |s(x)-t(x)|$ is the total variation distance (also called normalized trace distance). The equality in (1.5) was proved in [10]. The contraction coefficient leads to the following strong form of the data-processing inequality for total variation distance:

$$\eta_{\mathrm{TV}}(\mathcal{W}) \|s-t\|_1 \geq \|\mathcal{W}\circ s - \mathcal{W}\circ t\|_1 \tag{1.6}$$

and is considered strong when $\eta_{\mathrm{TV}}(\mathcal{W}) \in [0,1)$.

The following inequality relates the Doeblin coefficient to the contraction coefficient:

$$\eta_{\mathrm{TV}}(\mathcal{W}) \leq 1 - \alpha(\mathcal{W}). \tag{1.7}$$

See, e.g., [11, Theorem 8.17] and [12, Remark III.2]. The significance of this inequality is that one can use the Doeblin coefficient as a proxy for the contraction coefficient in various applications. The Doeblin coefficient additionally satisfies various desirable properties useful in applications, including concatenation [13, Corollary 2.2]

$$1 - \alpha(\mathcal{W}_2\circ\mathcal{W}_1) \leq (1-\alpha(\mathcal{W}_2))\,(1-\alpha(\mathcal{W}_1)), \tag{1.8}$$

and multiplicativity [2, Theorem 1, Item 5]:

$$\alpha(\mathcal{W}_1 \otimes \mathcal{W}_2) = \alpha(\mathcal{W}_1) \cdot \alpha(\mathcal{W}_2) \tag{1.9}$$

for two arbitrary classical channels $\mathcal{W}_1$ and $\mathcal{W}_2$.

1.2 Summary of Contributions

In this paper, our main aim is to explore the notion of Doeblin coefficient in quantum information theory, as well as its applications therein. As a consequence, we significantly expand on the developments of [11, Theorem 8.17 and Eq. (8.86)] and [14]. In particular, we define several notions of a quantum Doeblin coefficient, some of which go beyond those already proposed in [11, Theorem 8.17 and Eq. (8.86)] and [14]. Our primary criterion for a quantity to be called a quantum Doeblin coefficient is that it should reduce to the classical quantity $\alpha(\mathcal{W})$ when evaluated for a classical channel $\mathcal{W}$. Beyond that, we look for connections to the concepts and properties mentioned above: Does the proposed quantum Doeblin coefficient have an operational meaning? Is it efficiently computable? Does it satisfy the concatenation and multiplicativity properties mentioned above?

This line of questioning has led us to winnow down the various proposed quantities and focus primarily on the definition of quantum Doeblin coefficient given in Definition 4, denoted there as $\alpha(\mathcal{N})$. The quantity $\alpha(\mathcal{N})$ reduces to the classical quantity for a classical channel, it is efficiently computable as a semidefinite program, and it has an operational meaning generalizing that in (1.2)–(1.3). However, this quantity $\alpha(\mathcal{N})$ is not multiplicative in general. By modifying the definition just slightly (see Definition 6), we define another quantum Doeblin coefficient, denoted by $\alpha_{\mp}(\mathcal{N})$, that is multiplicative, reduces to the classical case for a classical channel, and is efficiently computable as a semidefinite program. However, the price to pay for this slight modification is that it no longer possesses an operational meaning in general.

A key distinction between the classical and quantum definitions of Doeblin coefficient is the optimization over a Hermitian operator X in Definition 4, Definition 5, and Definition 6, rather than restricting the optimization to positive semidefinite X. In the definition of the Doeblin coefficient of a classical channel in (1.1), we could have instead optimized over $r \in \mathbb{R}^{|\mathcal{Y}|}$; however, it suffices to restrict the optimization in this case to $r \in \mathbb{R}^{|\mathcal{Y}|}_{\geq 0}$. For the quantum case, it is not generally possible to restrict the optimization over Hermitian X to just positive semidefinite X. Let us also advocate that the most prudent definition of quantum Doeblin coefficient involves an optimization over Hermitian X. By doing so, not only do the quantities in Definition 4 and Definition 5 possess operational meanings in terms of exclusion values of the channel (either with entanglement assistance or without), but one also obtains tighter bounds on the trace distance contraction coefficient (see (6.16) and (6.34)), which is one of the main applications of a quantum Doeblin coefficient. Interestingly, let us also note that the exclusion task (and thus the exclusion value) is related to the foundations of quantum mechanics, by means of the PBR theorem from [15], and thus it brings out distinct features of quantum information when compared to classical information.

Let us now summarize the main contributions of our paper:

1. We provide various interpretations of the quantum Doeblin coefficient $\alpha(\mathcal{N})$ from Definition 4 (see Section 4). In particular, we interpret it as a channel's minimum singlet fraction, its entanglement-assisted exclusion value, its reverse max-mutual information,

its reverse robustness, and its reverse hypothesis-testing mutual information. See Table 1 for a summary of these interpretations, which are explained in far greater detail in Section 4. We also prove that it possesses various desirable properties, such as normalization, concatenation, data processing under pre- and post-processing, submultiplicativity, and concavity, in addition to being efficiently computable.

2. We provide various interpretations of the alternative notion of quantum Doeblin coefficient $\alpha_I(\mathcal{N})$ from Definition 5 (Section 4.6). In particular, we interpret it as a channel's minimum singlet fraction under entanglement-breaking decoders, its exclusion value, its reverse max-oveloH information, its reverse robustness under positive, trace-preserving maps, and its hypothesis-testing oveloH information. See Table 2 for a summary of these interpretations, which are explained in far greater detail in Section 4.6. We also prove that it possesses similar desirable properties, as mentioned above.

3. We define an alternative notion of quantum Doeblin coefficient, denoted by $\alpha_{\pm}(\mathcal{N})$ (see Definition 6), and establish that it possesses various desirable properties, including multiplicativity and concatenation. We summarize its properties and how they compare to other quantum Doeblin coefficients in Table 3.

4. We consider complete contraction coefficients for quantum divergences, with trace distance being an important special case; the latter is defined in (6.15) using the notation $\eta_{\mathrm{Tr}}^{c}(\mathcal{N})$. In particular, we prove that $\eta_{\mathrm{Tr}}^{c}(\mathcal{N}) \leq 1 - \alpha(\mathcal{N})$ in Proposition 30, which finds use in applications where there exist noiseless subsystems in tandem with noisy subsystems.

5. We evaluate the quantum Doeblin coefficient $\alpha(\mathcal{N})$ for several examples of channels relevant in applications, such as classical–quantum channels, measurement channels, generalized dephasing channels, qubit channels, and generalized amplitude damping channels (Section 7).

Beyond the contributions listed above, we also provide several applications of quantum Doeblin coefficients, relevant for quantum computation and information theory. These are due to the connection between Doeblin coefficients and contraction coefficients; they include the following:

1. We establish a connection between noise-induced barren plateaus and Doeblin coefficients in Proposition 41 of Section 8.1. In particular, a noise-induced barren plateau [16] occurs when it is difficult to train a parameterized quantum circuit affected by noise. The terminology means that the optimization landscape for training becomes extremely flat, making it too difficult to navigate when conducting an algorithm like gradient descent. In particular, we show that, for a parameterized quantum circuit affected by generic noise, it is quite difficult to train the circuit parameters located in the first few layers of the circuit. To this end, we use quantum Doeblin coefficients to provide efficiently computable bounds on the magnitude of the cost function's gradient, showcasing that the derivatives decay with the depth of the corresponding parameterized gate from the end of the circuit.

Moreover, our results prove that noise-induced barren plateaus occur if the noise channels have non-zero Doeblin coefficients, regardless of the existence of noiseless subsystems, parameter initialization strategy (even with warm starts), or the gradient optimization technique used, further expanding previously known limitations [16–19]. We should also note that this does not contradict the results of [19], where non-unital noise can possibly help in escaping barren plateaus, given that their finding holds only "on average".

2. We establish a connection between limitations on error mitigation and Doeblin coefficients (Section 8.2). In near-term quantum computing, one hope is that error mitigation protocols can reduce or mitigate the effect of noise on quantum computations used to estimate the expectation value of certain observables [20]. However, expanding on [21,22], we prove that all error mitigation protocols have an overhead related to the Doeblin coefficient when the noise channels in the circuit satisfy $\alpha(\mathcal{N}) > 0$ (see Theorem 6). Indeed, we prove that the number of noisy data samples required to achieve a fixed error tolerance scales exponentially with the circuit depth. These results are valid for general error mitigation protocols even with the assistance of noiseless subsystems in the circuit, under mild conditions on the observable and the set of states allowed.

3. We establish limitations on the sample complexity of hypothesis testing when the states are affected by noise (Section 8.3). In particular, our bounds in Proposition 43 feature the Doeblin coefficient of the noisy channel affecting the unknown states. This provides a sense of the cost one needs to pay in the task of hypothesis testing when only noisy samples of the states are available.

4. We finally outline connections between the Doeblin coefficient and the fairness of quantum learners (Section 8.4) and the mixing, indistinguishability, and decoupling times of discrete time homogeneous and inhomogeneous quantum processes (Section 8.5). Notably, we show how complete contraction coefficients can be used to bound the decoupling time and decoupled indistinguishability time defined in Definition 23. We then prove that the Doeblin coefficients provide efficiently computable upper bounds for these times when the corresponding divergence is trace distance (see Corollary 16 and Corollary 17).

In the conclusion of our paper (Section 9), we summarize our main findings and then outline directions for future research.

2 Preliminaries

We begin by establishing some notation used throughout the rest of our paper. For further background on quantum information, see [23–27]. Define the maximally entangled state Φ_d of Schmidt rank $d \in \mathbb{N}$ as

$$\Phi_d \coloneqq \frac{1}{d} \sum_{i,j} |i\rangle\langle j| \otimes |i\rangle\langle j|. \tag{2.1}$$

The Choi operator of a superoperator $\mathcal{M}_{A\to B}$ is defined as

$$\Gamma^{\mathcal{M}}_{AB} \coloneqq \sum_{i,j} |i\rangle\langle j|_A \otimes \mathcal{M}_{A'\to B}(|i\rangle\langle j|_{A'}), \tag{2.2}$$

Interpretation of $\alpha(\mathcal{N})$	Expression
Primal SDP	$\sup_{X_B \in \mathrm{Herm}} \left\{ \mathrm{Tr}[X_B] : I_A \otimes X_B \leq \Gamma^{\mathcal{N}}_{AB} \right\}$
Dual SDP	$\inf_{Y_{AB} \geq 0} \left\{ \mathrm{Tr}[Y_{AB} \Gamma^{\mathcal{N}}_{AB}] : \mathrm{Tr}_A[Y_{AB}] = I_B \right\}$
Minimal singlet fraction	$d^2 \inf_{\mathcal{D} \in \mathrm{CPTP}} F(\Phi^d, (\mathrm{id} \otimes (\mathcal{D} \circ \mathcal{N}))(\Phi^d))$
Entanglement-assisted exclusion value	$\inf_{\mathcal{X}, \Psi, (\mathcal{E}^x, \Lambda^x)_{x \in \mathcal{X}}} \sum_{x \in \mathcal{X}} \mathrm{Tr}[\Lambda^x_{RB} \mathcal{N}_{A \to B}(\mathcal{E}^x_{A' \to A}(\Psi_{RA'}))]$
Reverse max-mutual information	$\exp\left(- \inf_{\tau \in \mathrm{aff}(\mathcal{D})} D_{\max}(\mathcal{R}^\tau \| \mathcal{N}) \right)$
Reverse robustness	$\sup_{\substack{\lambda \in [0,1], \\ \tau \in \mathrm{aff}(\mathcal{D}), \\ \mathcal{M} \in \mathrm{CPTP}}} \{ \lambda : \lambda \mathcal{R}^\tau + (1-\lambda)\mathcal{M} = \mathcal{N} \}$
Reverse hypo.-testing mutual info.	$d^2 \exp\left(- \inf_{\tau \in \mathrm{aff}(\mathcal{D})} D_H^{1 - \frac{1}{d^2}}(\Phi^{\mathcal{R}^\tau} \| \Phi^{\mathcal{N}}) \right)$

Table 1: Various interpretations and expressions for the Doeblin coefficient $\alpha(\mathcal{N})$ of a quantum channel $\mathcal{N}$. The references to these expressions are as follows: primal SDP (3.1), dual SDP (3.3), minimal singlet fraction (4.2), entanglement-assisted exclusion value (4.17), reverse max-mutual information (4.21), reverse robustness (4.31), and reverse hypothesis-testing mutual information (4.40).

Interpretation of $\alpha_I(\mathcal{N})$	Expression
Primal Conic Program	$\sup_{X_B \in \mathrm{Herm}} \left\{ \mathrm{Tr}[X_B] : \mathcal{R}^X \leq_P \mathcal{N} \right\}$
Dual Conic Program	$\inf_{Y_{AB} \in \mathrm{Sep}(A:B)} \left\{ \mathrm{Tr}[Y_{AB} \Gamma^{\mathcal{N}}_{AB}] : \mathrm{Tr}_A[Y_{AB}] = I_B \right\}$
Minimal singlet fraction	$d^2 \inf_{\mathcal{D} \in \mathrm{EB}} F(\Phi^d, (\mathrm{id} \otimes (\mathcal{D} \circ \mathcal{N}))(\Phi^d))$
Exclusion value	$\inf_{\mathcal{X}, (\psi^x, \Lambda^x)_{x \in \mathcal{X}}} \sum_{x \in \mathcal{X}} \mathrm{Tr}[\Lambda^x_B \mathcal{N}_{A \to B}(\psi^x_A)]$
Max-oveloH information	$\exp\left(- \inf_{\tau \in \mathrm{aff}(\mathcal{D})} \sup_{\rho \in \mathcal{D}} D_{\max}(\tau \| \mathcal{N}(\rho)) \right)$
Reverse robustness	$\sup_{\substack{\lambda \in [0,1], \\ \tau \in \mathrm{aff}(\mathcal{D}), \mathcal{M} \in \mathrm{PTP}}} \{ \lambda : \lambda \mathcal{R}^\tau + (1-\lambda)\mathcal{M} = \mathcal{N} \}$
Hypo.-testing oveloH info.	$d^2 \exp\left(- \left[\sup_{(\psi_x)_{x=1}^{d^2}} \inf_{\tau \in \mathrm{aff}(\mathcal{D})} D_H^{1 - \frac{1}{d^2}}(\rho_X \otimes \tau_B \| \mathcal{N}_{A \to B}(\rho_{XA})) \right] \right)$

Table 2: Various interpretations and expressions for the induced Doeblin coefficient $\alpha_I(\mathcal{N})$ of a quantum channel $\mathcal{N}$. The references to these expressions are as follows: primal conic program (3.11), dual conic program (3.45), minimal singlet fraction (4.42), exclusion value (3.29), max-oveloH information (4.47), reverse robustness (4.59), and hypothesis-testing oveloH information (4.60). In the above, $\rho_{XA} = \sum_{x=1}^{d^2} \frac{1}{d^2} |x\rangle\langle x| \otimes \psi_x$, and each ψ_x is a pure state.

Property	$\alpha(\mathcal{N})$ (Def. 4)	$\alpha_I(\mathcal{N})$ (Def. 5)	$\alpha_{\mp}(\mathcal{N})$ (Def. 6)
Recovers Classical Case	✓ (Proposition 9)		
Normalized	✓ (Proposition 18)	✓ (Proposition 23)	✓ (Proposition 24)
Concatenation	✓ (Proposition 19)		✓ (Proposition 28)
Data Processing and Isometric Invariance	✓ (Proposition 20)		✓ (Proposition 26)
Submultiplicative	✓ (Proposition 21)		✓ (Proposition 25)
Multiplicative	× (Example 1)		✓ (Proposition 25)
Concavity	✓ (Proposition 22)		✓ (Proposition 27)

Table 3: Summary of the properties of the main quantum Doeblin coefficients and where to find their proofs. ✓ means the property is satisfied and × means it is not. Note this shows that the main advantage of $\alpha_{\mp}$ is that it is multiplicative at the expense of lacking the operational interpretations that the other two quantities possess.

where system A' is isomorphic to system A.

For two Hermitian operators N and M we use the notation $N \geq M$ to indicate that $N - M$ is a positive semidefinite operator. For two Hermiticity preserving superoperators $\mathcal{N}$ and $\mathcal{M}$, we use the notation $\mathcal{N} \geq \mathcal{M}$ to indicate that $\mathcal{N} - \mathcal{M}$ is a completely positive map, and we use the notation $\mathcal{N} \geq_P \mathcal{M}$ to indicate that $\mathcal{N} - \mathcal{M}$ is a positive map. The latter condition is equivalent to $\mathcal{N}(\rho) \geq \mathcal{M}(\rho)$ for every density operator ρ.

2.1 Quantum Divergences

Here we define some quantum divergences, which we take to be functions of a quasi-state (Hermitian operator with trace equal to one) and a positive semidefinite operator. In particular, we define the hockey-stick divergence and the hypothesis testing divergence (also called smoothed min-divergence). These were originally defined for the first argument being a state in [28, Eq. (9)] and [29–31], respectively, and the hypothesis-testing divergence was recently extended to the case of a quasi-state first argument in [32, Eq. (62)].

Let $\tau \in \operatorname{aff}(\mathcal{D})$ be a quasi-state, where $\operatorname{aff}(\mathcal{D})$ denotes the affine hull of the set of density operators (i.e., Hermitian operators with unit trace), and let σ be a positive semidefinite operator. For $\gamma \geq 0$, the hockey-stick divergence is defined as

$$E_\gamma(\tau\|\sigma) \coloneqq \operatorname{Tr}[(\tau - \gamma\sigma)_+] - (1-\gamma)_+, \tag{2.3}$$

where $(A)_+$ denotes the positive part of a Hermitian operator A. To define this, first define $(x)_+ \coloneqq \max\{0, x\}$ for scalars. Then $(A)_+ \coloneqq \sum_i (a_i)_+ |i\rangle\langle i|$, with $A = \sum_i a_i |i\rangle\langle i|$ a spectral

decomposition of A. We also have the following equivalent formulation:

$$E_\gamma(\tau\|\sigma) = \sup_{0\leq M\leq I} \mathrm{Tr}[M(\tau-\gamma\sigma)] - (1-\gamma)_+. \tag{2.4}$$

At $\gamma = 1$, the hockey-stick divergence reduces to the normalized trace distance:

$$E_1(\tau\|\sigma) = T(\tau,\sigma) \coloneqq \frac{1}{2}\left\|\tau-\sigma\right\|_1. \tag{2.5}$$

For $\varepsilon \in [0,1]$, the hypothesis testing relative entropy is defined as

$$D_H^\varepsilon(\tau\|\sigma) \coloneqq -\ln \inf_{\Lambda\geq 0} \{\mathrm{Tr}[\Lambda\sigma] : \mathrm{Tr}[\Lambda\tau] \geq 1-\varepsilon,\ \Lambda \leq I\} \tag{2.6}$$

$$= -\ln \sup_{\mu,Z\geq 0} \{\mu(1-\varepsilon) - \mathrm{Tr}[Z] : \mu\tau \leq \sigma + Z\}. \tag{2.7}$$

The dual expression in (2.7) was given in [33, Eq. (B2)] for the case of τ being a state, and the derivation easily extends to the case of τ being a quasi-state.

We have two different definitions of max-relative entropy of $\tau \in \mathrm{aff}(\mathcal{D})$ and σ positive semidefinite, which was originally defined for states in [34]. These are as follows:

$$D_{\max}(\tau\|\sigma) \coloneqq \ln \inf_{\lambda\geq 0} \{\lambda : \tau \leq \lambda\sigma\} \tag{2.8}$$

$$\widetilde{D}_\infty(\tau\|\sigma) \coloneqq \ln \inf_{\lambda\geq 0} \{\lambda : -\lambda\sigma \leq \tau \leq \lambda\sigma\} \tag{2.9}$$

$$= \ln \left\|\sigma^{-1/2}\tau\sigma^{-1/2}\right\|_\infty. \tag{2.10}$$

These quantities are related as follows:

$$D_{\max}(\tau\|\sigma) \leq \widetilde{D}_\infty(\tau\|\sigma). \tag{2.11}$$

We also define two different definitions of max-relative entropy of Hermiticity preserving maps $\mathcal{T}$ and $\mathcal{S}$ as

$$D_{\max}(\mathcal{T}\|\mathcal{S}) \coloneqq \ln \inf_{\lambda\geq 0} \{\lambda : \mathcal{T} \leq \lambda\mathcal{S}\} \tag{2.12}$$

$$\widetilde{D}_\infty(\mathcal{T}\|\mathcal{S}) \coloneqq \ln \inf_{\lambda\geq 0} \{\lambda : -\lambda\mathcal{S} \leq \mathcal{T} \leq \lambda\mathcal{S}\}, \tag{2.13}$$

where the superoperator inequality $\mathcal{T} \leq \lambda\mathcal{S}$ holds if and only if $\lambda\mathcal{S}-\mathcal{T}$ is a completely positive map. These quantities are likewise ordered as

$$D_{\max}(\mathcal{T}\|\mathcal{S}) \leq \widetilde{D}_\infty(\mathcal{T}\|\mathcal{S}). \tag{2.14}$$

2.2 Conic Programming and the Separable Cone

Throughout this work we use conic programs, which include the more commonly used semidefinite programs as a special case. Of particular importance are cone programs over the separable cone. We provide a brief review of these concepts as used in this work. We refer the reader to [35, 36] for further details on cone programs.

Cone Programs We begin by defining a cone program generically in a way that aligns with the rest of this work's notation. Let $(V, \langle\cdot,\cdot\rangle_V)$, $(W, \langle\cdot,\cdot\rangle_W)$ be real-valued inner product spaces. Let $\mathcal{K} \subset V$ be a closed, convex cone, and let $\mathcal{K}^*$ denote its dual cone. Let $a \in V$ and $b \in W$. Let $\Phi : V \to W$ be a linear map. The primal and dual of the cone program may then be written as[1]

$$\begin{aligned} &\textit{Primal problem} \\ \text{maximize:} \quad & \langle b, x\rangle_W \\ \text{subject to:} \quad & \Phi^*(x) - a \in \mathcal{K}^* \\ & x \in W \end{aligned} \tag{2.15}$$

$$\begin{aligned} &\textit{Dual problem} \\ \text{minimize:} \quad & \langle a, y\rangle_V \\ \text{subject to:} \quad & \Phi(y) = b \\ & y \in \mathcal{K}\,, \end{aligned} \tag{2.16}$$

where $\Phi^* : W \to V$ is the unique linear map satisfying $\langle w, \Phi(v)\rangle_W = \langle \Phi^*(w), v\rangle_V$ for all $v \in V, w \in W$.[2] Such cone programs always satisfy weak duality; i.e., defining the feasible sets

$$F \coloneqq \{x \in W : \Phi^*(x) - a \in \mathcal{K}^*\}, \tag{2.17}$$

$$G \coloneqq \{y \in \mathcal{K} : \Phi(y) = b)\}, \tag{2.18}$$

and the optimal values

$$\alpha \coloneqq \sup_{x \in F} \langle b, x\rangle_W, \tag{2.19}$$

$$\beta \coloneqq \inf_{y \in G} \langle a, y\rangle_V, \tag{2.20}$$

which are taken over the extended real line, we have $\alpha \leq \beta$. A sufficient condition for strong duality ($\beta = \alpha$) is Slater's criterion, which we state here (see also [35, Theorem 1.3]).

Proposition 1. *Let* $\mathrm{Relint}(C)$ *denote the relative interior of a convex set* C*. With respect to the notation given,*

1. *If* $F \neq \emptyset$ *and* $\mathrm{Relint}(\mathcal{K}) \cap G \neq \emptyset$*, then there exists* $x \in F$ *such that* $\langle b, x\rangle_W = \beta$*; i.e.,* $\alpha = \beta$*.*

2. *If* $G \neq \emptyset$ *and there exists* $x \in W$ *such that* $\Phi^*(x) - a \in \mathrm{Relint}(\mathcal{K}^*)$*, then there exists* $y \in G$ *such that* $\langle a, y\rangle_V = \alpha$*; i.e.,* $\alpha = \beta$*.*

Remark 1. *While we have presented the above more generally, we note throughout this work that we will always use the inner product to be the Hilbert–Schmidt product,* $\langle X, Y\rangle = \mathrm{Tr}[X^\dagger Y]$ *for all linear operators* $X, Y \in \mathrm{Lin}(A, B)$*, which will simplify some of the above notation.*

Separable Cone The cone that we primarily work with concretely is the separable cone. We summarize key properties to which we will appeal.

[1] Note that this is a relabeling of what is the primal and dual in [35, Section 1.4]. Thus, everything stated in this subsection is a relabeling of statements in [35].

[2] We reserve the notation $\mathcal{E}^\dagger$ for the adjoint map with respect to the Hilbert–Schmidt inner product as is standard in quantum information theory.

Definition 1. *Given Hilbert spaces A and B, the separable cone (equivalently, set of separable operators) are positive semidefinite operators $P_{AB} \geq 0$ such that there exists a finite alphabet $\mathcal{X}$ and tuples of positive operators $(Q_x)_{x\in\mathcal{X}} \subset \mathrm{Pos}(A)$, $(R_x)_{x\in\mathcal{X}} \subset \mathrm{Pos}(B)$ such that*

$$P = \sum_{x\in\mathcal{X}} Q_x \otimes R_x \ . \tag{2.21}$$

It is straightforward to verify that the separable cone is indeed a cone, and a proof that it is closed is established in [25, Proposition 6.8]. Using the definition of a dual cone and standard properties of separable operators, one may verify the relation

$$X \in \mathrm{Sep}^*(A:B) \Leftrightarrow \langle\alpha|_A\langle\beta|_B X|\alpha\rangle_A|\beta\rangle_B \geq 0 \quad \forall|\alpha\rangle_A|\beta\rangle_B \ . \tag{2.22}$$

The set of operators in $\mathrm{Sep}^*(A:B)$ are sometimes referred to as being 'block-positive.'

Lastly, we state the following well-known relation between a linear map being positive and the Choi operator being block-positive (see [37] for an elementary proof).

Proposition 2. *A linear map $\Phi_{A\to B}$ is positive if and only if $\Gamma^{\Phi}_{AB} \in \mathrm{Sep}^*(A:B)$.*

Beyond the separable cone itself, we also use the following well-known SDP-representable cones that contain the separable cone. First, let us recall the positive partial transpose (PPT) cone:

$$\mathrm{PPT}(A:B) \coloneqq \{P_{AB} \geq 0 : P^{\Gamma}_{AB} \geq 0\} \ , \tag{2.23}$$

where $X^{\Gamma} \coloneqq (\mathrm{id}_A \otimes T_B)(X)$ is the partial transpose as T_B is the transpose map. Second, let us recall the family of cones that interpolate between the separable and positive cone [38].

Definition 2. *For $k \in \mathbb{N}$, the k-symmetrically extendible cone is defined as*

$$\mathrm{Sym}_k(A:B) \coloneqq \left\{ P_{AB} \geq 0 : \exists P_{AB_1^k} \geq 0 : \begin{matrix} P_{AB_1^k} = (I_A \otimes W_\pi) P_{AB_1^k} (I_A \otimes W_\pi^\dagger) \ \forall \pi \in \mathcal{S}([k]) \\ \mathrm{Tr}_{B_1^{k-1}}[P_{AB_1^k}] = P_{AB} \end{matrix} \right\} \ ,$$

where $\mathcal{S}_k$ is the set of permutations on k elements, and for every $\pi \in \mathcal{S}_k$, W_π is the unitary that permutes the k copies of B according to π; i.e., for every unit vector $\{|v_i\rangle \in B_i\}_{i\in\{1,\ldots,k\}}$,

$$W_\pi |v_1\rangle_{B_1}|v_2\rangle_{B_2}\cdots|v_k\rangle_{B_k} = |v_1\rangle_{B_{\pi(1)}}|v_2\rangle_{B_{\pi(2)}}\cdots|v_k\rangle_{B_{\pi(k)}} \ .$$

This means that $\mathrm{Sym}_k(A:B)$ is the set of positive semidefinite operators, labeled by P_{AB}, that admit extensions where $P_{AB_i} = P_{AB}$ for all $i \in \{1,\ldots,k\}$. It is well-known that $\mathrm{Pos}(A\otimes B) = \mathrm{Sym}_1(A:B) \supsetneq \mathrm{Sym}_2(A:B) \supsetneq \cdots$ and $\mathrm{Sep}(A:B) = \lim_{k\to\infty} \mathrm{Sym}_k(A:B)$.

2.3 Quantum State Exclusion

The task of quantum state exclusion [8] appears in various forms throughout our paper. Quantum state exclusion was considered in [15] to study whether quantum states are ψ-ontic or ψ-epistemic, and it was studied directly subsequently, e.g., [8, 39]. More recently quantum state exclusion has been studied in a number of papers [9, 40–44].

Let us recall the task for a given ensemble of states.

Definition 3. *Consider a finite alphabet $\mathcal{X}$, a probability distribution over $\mathcal{X}$, p, and a tuple of quantum states $(\rho_A^x)_x$. For a given POVM with $|\mathcal{X}|$ outcomes, $(\Lambda_x)_x$, the error probability of state exclusion is $\sum_x p(x)\operatorname{Tr}[\Lambda_x\rho^x]$. We then define the minimal error probability of state exclusion as follows:*

$$P_{\text{err}}((p_x,\rho^x)_x) \coloneqq \min_{(\Lambda_x)_x\in\text{POVM}} \sum_x p(x)\operatorname{Tr}[\Lambda_x\rho^x]\ . \tag{2.24}$$

Note that this is quantum state exclusion, as the agent 'wins' whenever they do not guess the true value of the state; thus they win whenever they successfully exclude the state they were sent.

3 Quantum Doeblin Coefficients

3.1 Main Definition of Quantum Doeblin Coefficient

The main definition of a quantum channel's Doeblin coefficient that we use in this paper was essentially introduced in [11, Eq. (8.86)] and revisited in [14, Corollary III.7].

Definition 4 (Quantum Doeblin Coefficient)**.** *The quantum Doeblin coefficient of a channel $\mathcal{N}_{A\to B}$ is defined as*

$$\alpha(\mathcal{N}) \coloneqq \sup_{X_B\in\text{Herm}} \left\{\operatorname{Tr}[X_B] : I_A\otimes X_B \le \Gamma_{AB}^{\mathcal{N}}\right\}, \tag{3.1}$$

where the Choi operator of $\mathcal{N}_{A\to B}$ is defined in (2.2).

We note here that Definition 4 is equivalent to

$$\alpha(\mathcal{N}) = \sup_{X\in\text{Herm}} \left\{\operatorname{Tr}[X] : \mathcal{R}^X \le \mathcal{N}\right\}, \tag{3.2}$$

where $\mathcal{R}^X(\cdot) \coloneqq \operatorname{Tr}[\cdot]X$ is a replacer map.

By observing that (3.1) is a semidefinite program, we obtain the following dual expression for the Doeblin coefficient, which is helpful in our analysis carried out later.

Proposition 3 (Dual SDP Formula for Doeblin Coefficient)**.** *Let $\mathcal{N}_{A\to B}$ be a quantum channel. Then $\alpha(\mathcal{N})$ defined in (3.1) is equal to the following expression:*

$$\alpha(\mathcal{N}) = \inf_{Y_{AB}\ge 0} \left\{\operatorname{Tr}[Y_{AB}\Gamma_{AB}^{\mathcal{N}}] \ :\ \operatorname{Tr}_A[Y_{AB}] = I_B\right\}, \tag{3.3}$$

where $\Gamma_{AB}^{\mathcal{N}}$ is defined in (2.2).

Proof. In the following proof, we omit system labels for simplicity. Consider that

$$\alpha(\mathcal{N}) = \sup_{X\in\text{Herm}} \left\{\operatorname{Tr}[X] : I\otimes X \le \Gamma^{\mathcal{N}}\right\} \tag{3.4}$$

$$= \sup_{X\in\text{Herm}} \left\{\operatorname{Tr}[X] + \inf_{Y\ge 0} \operatorname{Tr}[Y(\Gamma^{\mathcal{N}} - I\otimes X)]\right\} \tag{3.5}$$

$$= \sup_{X\in\text{Herm}} \inf_{Y\ge 0} \left\{\operatorname{Tr}[X] + \operatorname{Tr}[Y(\Gamma^{\mathcal{N}} - I\otimes X)]\right\} \tag{3.6}$$

$$= \sup_{X \in \text{Herm}} \inf_{Y \geq 0} \left\{ \text{Tr}[Y\Gamma^{\mathcal{N}}] + \text{Tr}[X\,(I - \text{Tr}_A[Y])] \right\} \tag{3.7}$$

$$\leq \inf_{Y \geq 0} \sup_{X \in \text{Herm}} \left\{ \text{Tr}[Y\Gamma^{\mathcal{N}}] + \text{Tr}[X\,(I - \text{Tr}_A[Y])] \right\} \tag{3.8}$$

$$= \inf_{Y \geq 0} \left\{ \text{Tr}[Y\Gamma^{\mathcal{N}}] : I = \text{Tr}_A[Y] \right\}. \tag{3.9}$$

Strong duality holds by picking strongly feasible solution $X = -I_B$ in the primal and the feasible solution $Y = \frac{1}{d_A} I_A \otimes I_B$ in the dual. Then it follows that

$$\alpha(\mathcal{N}) = \inf_{Y \geq 0} \left\{ \text{Tr}[Y\Gamma^{\mathcal{N}}] : I = \text{Tr}_A[Y] \right\}. \tag{3.10}$$

This concludes the proof. □

3.2 Alternative Notions of Quantum Doeblin Coefficients

In this section, we introduce some alternative notions of quantum Doeblin coefficient. All of these quantities reduce to the classical definition when the channel is a classical channel, as stated in Proposition 9 at the end of this section.

The first alternative definition we consider is from [11, Theorem 8.17], subsequently used in [45, Definition 1]:

Definition 5. *For a quantum channel $\mathcal{N}$, define the following quantity:*

$$\alpha_I(\mathcal{N}) \coloneqq \sup_{X \in \text{Herm}} \left\{ \text{Tr}[X] : \mathcal{N} - \mathcal{R}^X \in \text{Pos} \right\} \tag{3.11}$$

$$= \sup_{X \in \text{Herm}} \left\{ \text{Tr}[X] : \mathcal{R}^X(\rho) \leq \mathcal{N}(\rho) \quad \forall \rho \in \mathcal{D} \right\}, \tag{3.12}$$

where Pos *denotes the set of positive maps.*

This quantity leads to an improved upper bound on the trace-distance contraction coefficient than that given by $\alpha(\mathcal{N})$, as shown in [11, Theorem 8.17]. However, the main drawback of it is that it is not efficiently computable, because it is computationally difficult to optimize over the set of positive maps.

We now introduce another alternative definition of quantum Doeblin coefficient, denoted by $\alpha_{\mp}(\mathcal{N})$ and inspired by methods employed in [46, 47]. Although it does not give an improved upper bound on the trace distance contraction coefficient, it has several desirable properties, including 1) it is efficiently computable as a semidefinite program and 2) it is multiplicative under tensor products of quantum channels. As such, the quantity $\alpha_{\mp}(\mathcal{N})$ is useful in applications.

Definition 6. *For a quantum channel $\mathcal{N}$, let us define $\alpha_{\mp}(\mathcal{N})$ as*

$$\alpha_{\mp}(\mathcal{N}) \coloneqq \sup_{X_B \in \text{Herm}} \left\{ \text{Tr}[X_B] : -\Gamma^{\mathcal{N}}_{AB} \leq I_A \otimes X_B \leq \Gamma^{\mathcal{N}}_{AB} \right\}. \tag{3.13}$$

We note that alternative expressions for $\alpha_{\mp}(\mathcal{N})$ are as follows:

$$\alpha_{\mp}(\mathcal{N}) = \sup_{X_B \in \text{Herm}} \left\{ \text{Tr}[X_B] : -\mathcal{N} \leq \mathcal{R}^X \leq \mathcal{N} \right\} \tag{3.14}$$

$$= \left[\inf_{\lambda \geq 0, \tau \in \text{aff}(\mathcal{D})} \left\{ \lambda : -\lambda \Phi^{\mathcal{N}} \leq \pi_d \otimes \tau \leq \lambda \Phi^{\mathcal{N}} \right\} \right]^{-1} \tag{3.15}$$

$$= \left[\inf_{\tau \in \text{aff}(\mathcal{D})} \left\| (\Phi^{\mathcal{N}})^{-1/2} (\pi_d \otimes \tau) (\Phi^{\mathcal{N}})^{-1/2} \right\|_\infty \right]^{-1}. \tag{3.16}$$

The following inequality is an immediate consequence of Definition 4 and Definition 6, because the optimization in Definition 6 has more constraints than that in Definition 4:

$$\alpha(\mathcal{N}) \geq \alpha_{\mp}(\mathcal{N}). \tag{3.17}$$

Proposition 4 (Dual of $\alpha_{\mp}(\mathcal{N})$)**.** *For a quantum channel $\mathcal{N}_{A\to B}$, the following equality holds:*

$$\alpha_{\mp}(\mathcal{N}) = \inf_{Y^1_{AB}, Y^2_{AB} \geq 0} \left\{ \text{Tr}\left[\left(Y^1_{AB} + Y^2_{AB} \right) \Gamma^{\mathcal{N}}_{AB} \right] : I_B = \text{Tr}_A[Y^2_{AB} - Y^1_{AB}] \right\}, \tag{3.18}$$

where $\Gamma^{\mathcal{N}}_{AB}$ is defined in (2.2).

Proof. Consider that

$$\alpha_{\mp}(\mathcal{N}) = \sup_{X \in \text{Herm}} \left\{ \text{Tr}[X] : -\Gamma^{\mathcal{N}} \leq I \otimes X \leq \Gamma^{\mathcal{N}} \right\} \tag{3.19}$$

$$= \sup_{X \in \text{Herm}} \left\{ \text{Tr}[X] + \inf_{Y_1, Y_2 \geq 0} \text{Tr}\left[\left(I \otimes X + \Gamma^{\mathcal{N}} \right) Y_1 \right] + \text{Tr}\left[\left(\Gamma^{\mathcal{N}} - I \otimes X \right) Y_2 \right] \right\} \tag{3.20}$$

$$= \sup_{X \in \text{Herm}} \inf_{Y_1, Y_2 \geq 0} \left\{ \text{Tr}[X] + \text{Tr}\left[\left(I \otimes X + \Gamma^{\mathcal{N}} \right) Y_1 \right] + \text{Tr}\left[\left(\Gamma^{\mathcal{N}} - I \otimes X \right) Y_2 \right] \right\} \tag{3.21}$$

$$= \sup_{X \in \text{Herm}} \inf_{Y_1, Y_2 \geq 0} \left\{ \text{Tr}\left[\Gamma^{\mathcal{N}} (Y_1 + Y_2) \right] + \text{Tr}[X (I + \text{Tr}_A[Y_1 - Y_2])] \right\} \tag{3.22}$$

$$\leq \inf_{Y_1, Y_2 \geq 0} \sup_{X \in \text{Herm}} \left\{ \text{Tr}\left[\Gamma^{\mathcal{N}} (Y_1 + Y_2) \right] + \text{Tr}[X (I + \text{Tr}_A[Y_1 - Y_2])] \right\} \tag{3.23}$$

$$= \inf_{Y_1, Y_2 \geq 0} \left\{ \text{Tr}\left[\Gamma^{\mathcal{N}} (Y_1 + Y_2) \right] : I = \text{Tr}_A[Y_2 - Y_1] \right\}. \tag{3.24}$$

Strong duality holds by picking $X = 0$ (primal feasible) and $Y_2 = (1+\varepsilon) \frac{I}{|A|} \otimes I$ and $Y_1 = \varepsilon \frac{I}{|A|} \otimes I$, for $\varepsilon > 0$, so that these choices are strictly feasible for the dual. □

We define more alternative Doeblin coefficients by restricting the $X \in \text{Herm}$ condition to $X \geq 0$.

Definition 7. *For a quantum channel $\mathcal{N}$, define the following quantities:*

$$\alpha_+(\mathcal{N}) \coloneqq \sup_{X \geq 0} \left\{ \text{Tr}[X] : I \otimes X \leq \Gamma^{\mathcal{N}} \right\} \tag{3.25}$$

$$\alpha_{+,I}(\mathcal{N}) \coloneqq \sup_{X \geq 0} \left\{ \text{Tr}[X] : \mathcal{R}^X \leq_P \mathcal{N} \right\}, \tag{3.26}$$

where $\mathcal{R}^X(\cdot) = \text{Tr}[\cdot] X$.

Note that

$$\alpha_+(\mathcal{N}) \leq \alpha_{\mp}(\mathcal{N}) \tag{3.27}$$

because $X \geq 0$ and $I \otimes X \leq \Gamma^{\mathcal{N}}$ imply that $-\Gamma^{\mathcal{N}} \leq I \otimes X \leq \Gamma^{\mathcal{N}}$, so that optimizing over Hermitian X satisfying $-\Gamma^{\mathcal{N}} \leq I \otimes X \leq \Gamma^{\mathcal{N}}$ can only give the same or a larger value for $\text{Tr}[X]$.

3.2.1 Induced and Cone-Restricted Doeblin Coefficients

In this section we consider even more variants of quantum Doeblin coefficients. We begin by defining an induced version of the Doeblin coefficient, which arises from considering the classical Doeblin coefficient evaluated over all classical channels induced by a given quantum channel. We then establish a cone program for this quantity, and one of our findings is that the induced Doeblin coefficient is equal to the expression in (3.11). By generalizing the cone program, we obtain 'cone-restricted' Doeblin coefficients, which recover the quantum Doeblin coefficient $\alpha(\mathcal{N})$ as a special case.

Recall the expression for the classical Doeblin coefficient from (1.3). We can then consider the 'induced Doeblin coefficient' of a quantum channel by the use of prepared states and POVMs to map the quantum output system to a classical system. This is analogous to what was done in [48], where the communication value of a quantum channel is induced by the communication value of a classical channel.

Definition 8. *For a general channel $\mathcal{N}_{A\to B}$ and $n, n' \in \mathbb{N}$, define the $n \to n'$ best-case induced Doeblin coefficient as*

$$\alpha_I^{n\to n'}(\mathcal{N}) \coloneqq \min_{\substack{(\Lambda_y)_{y\in[n']}\in \mathrm{POVM} \\ (\rho_x)_{x\in[n]}}} \left\{ \sum_{y\in[n']} \min_{x\in[n]} \mathrm{Tr}[\Lambda_y \mathcal{N}(\rho_x)] \right\} . \tag{3.28}$$

Moreover, the best-case induced Doeblin coefficient is defined as

$$\alpha_I(\mathcal{N}) \coloneqq \inf_{n,n'\in\mathbb{N}} \alpha_I^{n\to n'}(\mathcal{N}) . \tag{3.29}$$

As we show later on (Proposition 20), the Doeblin coefficient $\alpha(\mathcal{N})$ satisfies a data-processing inequality. Using this now, we conclude that the induced Doeblin coefficient $\alpha_I(\mathcal{N})$ cannot be smaller than the quantum Doeblin coefficient $\alpha(\mathcal{N})$.

Proposition 5. *For a quantum channel $\mathcal{N}$, the following inequality holds:*

$$\alpha_I(\mathcal{N}) \geq \alpha(\mathcal{N}), \tag{3.30}$$

where $\alpha_I(\mathcal{N})$ and $\alpha(\mathcal{N})$ are defined in (3.29) and (3.1), respectively.

Proof. Consider preparation and measurement channels $\mathcal{F}_{X\to A}$, and $\mathcal{E}_{B\to Y}$ respectively. Then by Proposition 20, it follows that $\alpha(\mathcal{E}\circ\mathcal{N}\circ\mathcal{F}) \leq \alpha(\mathcal{N})$. We finally conclude that

$$\alpha_I(\mathcal{N}) = \inf_{\mathcal{F}_{X\to A},\mathcal{E}_{B\to Y}} \alpha(\mathcal{F}\circ\mathcal{N}\circ\mathcal{E}) \geq \alpha(\mathcal{N}) , \tag{3.31}$$

where we used the definition of the induced Doeblin coefficient. □

Next, we aim to show that $\alpha_I(\mathcal{N})$ can be expressed as a conic program. This follows in steps and closely follows proof methods from [49].

Proposition 6. *Let $\mathcal{N}_{A\to B}$ be a quantum channel, and fix $n \in \mathbb{N}$. Then*

$$\alpha_I^{n\to n}(\mathcal{N}) = \inf_{\substack{(\Lambda_x)_{x\in[n]} \\ (|\psi_x\rangle)_{x\in[n]}}} \sum_{x\in[n]} \mathrm{Tr}[\Lambda_x \mathcal{N}(|\psi_x\rangle\langle\psi_x|)] , \tag{3.32}$$

$$\alpha_I(\mathcal{N}) = \alpha_I^{d_B^2\to d_B^2}(\mathcal{N}) . \tag{3.33}$$

Proof. Identifying $\mathcal{Y} = [n']$, $\mathcal{X} = [n]$, for every choice of strategy, we may consider the function $f : \mathcal{Y} \to \mathcal{X}$ defined by $f(y) \coloneqq \operatorname{argmin}_x \operatorname{Tr}[\Lambda_y \mathcal{N}(\rho_x)]$, so that $\alpha^{n\to n'} = \sum_y \operatorname{Tr}[\Lambda_y \mathcal{N}(\rho_{f(y)})]$. First we will show that, without loss of generality, the induced Doeblin coefficient $\alpha_I(\mathcal{N})$ is achieved using a strategy where f is bijective, at which point a relabeling will allow us to reduce the induced quantum Doeblin coefficient to the expression $\alpha_I(\mathcal{N}) = \sup_{n\in\mathbb{N}} \alpha_I^{n\to n}(\mathcal{N})$.

Consider an induced classical channel $\mathcal{W}$. If there is $\hat{x} \in \mathcal{X}$ such that $\hat{x} \notin \operatorname{Im}(f)$, then one may remove $\rho_{\hat{x}}$ to obtain a new induced channel with the same induced Doeblin coefficient, as $\rho_{\hat{x}}$ is not the minimizer for any $y \in \mathcal{Y}$. Thus, f is surjective.

Now imagine there exist $y_1 \neq y_2$ such that $f(y_1) = f(y_2) = \hat{x}$; i.e., f is not injective. Define $\overline{W}_{\overline{Y}|X}(\overline{y}|x)$ where $\overline{\mathcal{Y}} \coloneqq \mathcal{Y} \setminus \{y_1, y_2\} \cup \{\hat{y}\}$, which is induced by altering the POVM by $\Lambda_{\widetilde{y}} \coloneqq \Lambda_y$ for $y \in \mathcal{Y} \setminus \{y_1, y_2\}$ and $\Lambda_{\hat{y}} \coloneqq \Lambda_{y_1} + \Lambda_{y_2}$. Then,

$$\alpha(\mathcal{W}) = \left(\sum_{y\in\mathcal{Y}\setminus\{y_1,y_2\}} \min_x \operatorname{Tr}[\Lambda_y \mathcal{N}(\rho_x)] \right) + \operatorname{Tr}[\Lambda_{y_1} \mathcal{N}(\rho_{\hat{x}})] + \operatorname{Tr}[\Lambda_{y_2} \mathcal{N}(\rho_{\hat{x}})] \tag{3.34}$$

$$\geq \sum_{y\in\mathcal{Y}\setminus\{y_1,y_2\}} \min_x \operatorname{Tr}[\Lambda_y \mathcal{N}(\rho_x)] + \min_x \operatorname{Tr}[\Lambda_{\hat{y}} \mathcal{N}(\rho_x)] \tag{3.35}$$

$$= \alpha(\overline{\mathcal{W}}) \ , \tag{3.36}$$

where the inequality follows from linearity and the definition of $\Lambda_{\hat{y}}$. Since the induced Doeblin coefficient involves an infimization, we can repeat this procedure iteratively to obtain an injective function. Thus, an optimal f is always bijective. Taking everything together, this implies that $\alpha_I(\mathcal{N}) = \inf_{n\in\mathbb{N}} \alpha_I^{n\to n}(\mathcal{N})$, where

$$\alpha_I^{n\to n}(\mathcal{N}) = \inf_{\substack{(\Lambda_x)_{x\in[n]} \\ (\rho_x)_{x\in[n]}}} \sum_{x\in[n]} \operatorname{Tr}[\Lambda_x \mathcal{N}(\rho_{\sigma(x)})] \ , \tag{3.37}$$

and σ is a permutation on n objects, as f is a bijection and thus a permutation. Lastly, we may relabel the states with the inverse permutation σ^{-1} to obtain the form

$$\alpha_I^{n\to n}(\mathcal{N}) = \inf_{\substack{(\Lambda_x)_{x\in[n]} \\ (\rho_x)_{x\in[n]}}} \sum_{x\in[n]} \operatorname{Tr}[\Lambda_x \mathcal{N}(\rho_x)] \ . \tag{3.38}$$

Now, we show that we can restrict the optimization to be over pure states. To do this, for each $x \in \mathcal{X}$ let $\rho_x = \sum_{y=1}^{r_x} p^x(y) |\phi_y^x\rangle\langle\phi_y^x|$ where r_x is the rank of ρ_X and $p^x(y)$ is a probability distribution. The existence of this decomposition follows from the spectral theorem. Then we have

$$\sum_x \operatorname{Tr}[\Lambda_x \mathcal{N}(\rho_x)] = \sum_x \operatorname{Tr}\left[\Lambda_x \sum_y p^x(y) \mathcal{N}(|\phi_y^x\rangle\langle\phi_y^x|)\right] \geq \sum_x \operatorname{Tr}[\Lambda_x \mathcal{N}(|\phi_{y^\star}^x\rangle\langle\phi_{y^\star}^x|] \ , \tag{3.39}$$

where $|\phi_{y^\star}^x\rangle \coloneqq \operatorname{argmin}_y \operatorname{Tr}[\Lambda_x |\phi_y^x\rangle\langle\phi_y^x|]$, which justifies the inequality. Thus, one can always decrease the value by replacing ρ_x with one of the pure states in its spectral decomposition, and so the restriction to pure states is without loss of generality. This establishes (3.32).

Finally, we will establish that $\alpha(\mathcal{N}) = \alpha_I^{d_B^2 \to d_B^2}(\mathcal{N})$. We do this by a common convexity argument for bounding the cardinality of the necessary POVM outcome size. For clarity,

we provide the argument in full. First, recall that a POVM on a space B, $(\Lambda_x)_{x\in[n]}$, is extremal if it cannot be expressed as a convex combination of other POVMs on that space, in the sense that it cannot be expressed as $\Lambda_x = \sum_y p(y)\Lambda_x^y$ where p is a non-degenerate probability distribution and $((\Lambda_x^y)_x)_y$ are POVMs. Second, recalling Minkowski's theorem [25, Theorem 1.10], and noting that the set of POVMs on (finite-dimensional) B is compact, all POVMs may be expressed as convex combinations of extremal POVMs; i.e., $(\Lambda_x)_x$ is such that $\Lambda_x = \sum_y p(y)\Lambda_x^y$ where p is a distribution, and for each y, the POVM $(\Lambda_x^y)_x$ is an extremal POVM on B. Finally, recall that an extremal POVM on B has at most d_B^2 non-zero outcomes [25, Corollary 2.48]. It follows that for all $n \geq d_B^2$, the inequality $\alpha_I^{n\to n} \geq \alpha_I^{d_B^2\to d_B^2}$ holds, following from linearity in the choice of POVM. That is, for all $n \geq d_B^2$, every set $(|\psi_x\rangle)_{x\in[n]}$ and (possibly non-extremal) POVM $(\Lambda_x)_{x\in[n]}$ such that $\Lambda_x \neq 0$ for all $x \in [n]$,

$$\sum_x \operatorname{Tr}[\Lambda_x \mathcal{N}(|\psi_x\rangle\langle\psi_x|)] = \sum_x \operatorname{Tr}\left[\sum_y p(y)\Lambda_x^y \mathcal{N}(|\psi_x\rangle\langle\psi_x|)\right] \tag{3.40}$$

$$= \sum_y p(y) \sum_x \operatorname{Tr}[\Lambda_x^y \mathcal{N}(|\psi_x\rangle\langle\psi_x|)] \tag{3.41}$$

$$\geq \min_y \sum_x \operatorname{Tr}[\Lambda_x^y \mathcal{N}(|\psi_x\rangle\langle\psi_x|)] \tag{3.42}$$

$$\geq \alpha_I^{d_B^2\to d_B^2}(\mathcal{N}) \ , \tag{3.43}$$

where we used the definition of $\alpha_I^{n\to n}$ in the final inequality and that an extremal POVM $\{\Lambda_x^y : \Lambda_x^y \neq 0\}$ with the $\{|\psi_x\rangle : \Lambda_x^y \neq 0\}$ is always feasible for $\alpha_I^{d_B^2\to d_B^2}$. As $\alpha_I(\mathcal{N}) = \inf_n \alpha_I^{n\to n}(\mathcal{N})$ and $\alpha_I^{n\to n}(\mathcal{N})$ is non-increasing in n, this completes the proof. □

Proposition 7. *Given a quantum channel $\mathcal{N}_{A\to B}$, the induced Doeblin coefficient $\alpha_I(\mathcal{N})$ from (3.29) is equal to both of the following conic optimization programs:*

Primal problem		Dual problem		
maximize:	$\operatorname{Tr}[X]$	*minimize:*	$\operatorname{Tr}[Y\Gamma^{\mathcal{N}}]$	
subject to:	$\Gamma^{\mathcal{N}} - I_A \otimes X \in \operatorname{Sep}^*(A:B)$	*subject to:*	$\operatorname{Tr}_A[Y] = I_B$	(3.45)
	$X \in \operatorname{Herm}(B)$.		$Y \in \operatorname{Sep}(A:B)$	

(3.44)

Proof. The proof is nearly identical to that of [48, Proposition 2], but we provide it for completeness. By Proposition 6, letting $n = d_B^2$, we conclude that

$$\alpha_I(\mathcal{N}) = \min_{\substack{(\Lambda_x)_{x\in[n]}\in\mathrm{POVM} \\ (\rho_x)_{x\in[n]}}} \sum_{x\in[n]} \operatorname{Tr}[\Lambda_x \mathcal{N}(\rho_x)] \tag{3.46}$$

$$= \min_{\substack{(\Lambda_x)_{x\in[n]}\in\mathrm{POVM} \\ (\rho_x)_{x\in[n]}}} \sum_{x\in[n]} \operatorname{Tr}[(\rho_x^T \otimes \Lambda_x)\Gamma^{\mathcal{N}}] \ , \tag{3.47}$$

where we used the action of the channel in terms of its Choi operator. Note that the operator $Y := \sum_x \rho_x^T \otimes \Lambda_x \geq 0$ is separable and satisfies $\operatorname{Tr}_A[Y] = \sum_x \Lambda_x = I_B$. Similarly, if Y is separable, then there exists a finite alphabet $\mathcal{Y}$ such that $Y = \sum_y |\psi_y\rangle\langle\psi_y| \otimes P_y$ where $P_y \geq 0$ for all $y \in \mathcal{Y}$. Moreover, $\operatorname{Tr}_A[Y] = I_B$ implies that $\sum_y P_y = I_B$. Thus, this defines a feasible

strategy. This derives what we labeled the dual problem. This is already in the form of (2.16) where we let $\Phi := \mathrm{Tr}_A$, so we obtain the primal problem immediately by recalling the adjoint map of Tr_A consists of tensoring the identity on the A space. □

With this cone program established, we are able to show that the induced Doeblin coefficient of this section is indeed the same as the quantity defined in Definition 5.

Proposition 8. *The quantities denoted by $\alpha_I(\mathcal{N})$ in Definitions 5 and 8 are equal.*

Proof. As the objective functions of Definition 5 and the primal problem of Proposition 7 are the same, it suffices to show that the constraints of these optimization programs are equivalent. We have the following equivalences:

$$\mathcal{N} - \mathcal{R}^X \in \mathrm{Pos} \quad \Leftrightarrow \quad \Gamma^{\mathcal{N}-\mathcal{R}^X} \in \mathrm{Sep}^*(A:B) \tag{3.48}$$

$$\Leftrightarrow \quad \Gamma^{\mathcal{N}} - I_A \otimes X \in \mathrm{Sep}^*(A:B)\ , \tag{3.49}$$

where the first equivalence is given in Proposition 2 and the second equivalence follows from the definition of the Choi operator and a replacer map. As in both cases, $X \in \mathrm{Herm}(A)$, this shows that the constraints are equivalent and completes the proof. □

Finally, following [50], by replacing the separable cone with the positive cone $\mathcal{K} \subset \mathrm{Pos}(A \otimes B)$, we define the cone-restricted Doeblin coefficient $\alpha_{I,\mathcal{K}}(\mathcal{N})$ as follows:

$$\begin{array}{rl} & \textit{Primal problem} \\ \text{maximize:} & \mathrm{Tr}[X] \\ \text{subject to:} & \Gamma^{\mathcal{N}} - I_A \otimes X \in \mathcal{K}^*(A:B) \\ & X \in \mathrm{Herm}(B)\ . \end{array} \qquad \begin{array}{rl} & \textit{Dual problem} \\ \text{minimize:} & \mathrm{Tr}[Y\Gamma^{\mathcal{N}}] \\ \text{subject to:} & \mathrm{Tr}_A[Y] = I_B \\ & Y \in \mathcal{K} \end{array} \tag{3.50, 3.51}$$

We note here that we need extra structure to guarantee strong duality in this case. A sufficient condition is for I_{AB} to be in the relative interior of the cone $\mathcal{K}$ in order to apply Slater's criterion for strong duality. This sufficient condition will be used later to work with $\alpha_{I,\mathcal{K}}$.

We end this section by stressing that all these quantities studied above recover the classical Doeblin coefficient for classical channels, which justifies identifying them as quantum Doeblin coefficients.

Proposition 9. *Let $\mathcal{W}_{Y|X}$ be a classical-to-classical channel. Then for* $\mathrm{Sep}(A:B) \subseteq \mathcal{K} \subset \mathrm{Pos}(A \otimes B)$*,*

$$\alpha_{\mathrm{cl}}(\mathcal{W}) = \alpha_{I,\mathcal{K}}(\mathcal{W}) = \alpha(\mathcal{W}) = \alpha_{\mp}(\mathcal{W})\ , \tag{3.52}$$

where α_{cl} denotes the classical Doeblin coefficient.

Proof. The key point is that $\mathcal{W}$ is invariant under dephasing, and all the conic constraints we consider are preserved under this action, so we can reduce back to the classical case. There are various forms of such a proof. For concreteness, we provide one such proof for $\alpha_I = \alpha_{I,\mathrm{Sep}}$. The proofs for α and $\alpha_{\mp}$ are similar, and then the cones listed in the proposition statement take the claimed value, as they are bounded from below and above by $\alpha_I(\mathcal{W})$ and $\alpha_{\mp}(\mathcal{W})$, respectively.

Let X be an optimizer for $\alpha_I(\mathcal{W})$. First we show, without loss of generality, that it is diagonal in the computational basis. By Proposition 2, $\mathcal{W} - \mathcal{R}^X$ is a positive map. Letting Δ_Y denote the completely dephasing map, we have that $\Delta_Y \circ (\mathcal{W} - \mathcal{R}^X) = \mathcal{W} - \mathcal{R}^{\Delta_Y(X)}$ is positive. Moreover $\mathrm{Tr}[X] = \mathrm{Tr}[\Delta_Y(X)]$. Thus, without loss of generality, an optimizer X of $\alpha_I(\mathcal{W})$ is diagonal. As X is Hermitian, its diagonal entries are all real. Thus, expressing the y^{th} diagonal entry as $r(y)$, we are interested in $\Gamma^{\mathcal{W}} - I \otimes X = \sum_{x,y} (W(y|x) - r(y)) |x\rangle\langle x| \otimes |y\rangle\langle y|$. By (2.22), $\mathrm{Sep}^*(X : Y) \ni \Gamma^{\mathcal{W}} - I \otimes X$ if and only if $r(y) \leq W(y|x)$ for all x and y. Noting that $\mathrm{Tr}[X] = \sum_y r(y)$ and $W(y|x) \geq 0$, without loss of generality $r(y) \geq 0$ for all y. Putting these points together,

$$\alpha_I(\mathcal{W}) = \max_{r \in \mathbb{R}^{|\mathcal{Y}|}_{\geq 0}} \left\{ \sum_{y \in \mathcal{Y}} r(y) : r(y) \leq W(y|x) \ \forall x \in \mathcal{X}, y \in \mathcal{Y} \right\} = \alpha_{\mathrm{cl}}(\mathcal{W}) \ , \tag{3.53}$$

where the final equality follows from (1.1). □

4 Interpretations of the Quantum Doeblin Coefficient

In the following subsections, we establish various interpretations of and alternate expressions for the quantum Doeblin coefficient $\alpha(\mathcal{N})$, as summarized in Table 1. After that, we do the same for the induced Doeblin coefficient $\alpha_I(\mathcal{N})$, and note that these latter interpretations are summarized in Table 2.

4.1 Minimum Achievable Singlet Fraction

In this section, we show that the quantum Doeblin coefficient in (3.1) is proportional to the minimum achievable singlet fraction, which is a counterpart of the maximum achievable singlet fraction considered in [51]. Alternatively, due to the connection with a task known as quantum state exclusion [8], this task could also be called fully quantum exclusion.

Theorem 1. *The quantum Doeblin coefficient of a channel $\mathcal{N}$ is proportional to the minimum achievable singlet fraction:*

$$\alpha(\mathcal{N}) = d^2 \inf_{\mathcal{D} \in \mathrm{CPTP}} \mathrm{Tr}[\Phi_d (\mathrm{id} \otimes (\mathcal{D} \circ \mathcal{N}))(\Phi_d)] \tag{4.1}$$

$$= d^2 \inf_{\mathcal{D} \in \mathrm{CPTP}} F(\Phi_d, (\mathrm{id} \otimes (\mathcal{D} \circ \mathcal{N}))(\Phi_d)), \tag{4.2}$$

where $F(\omega, \tau) \coloneqq \|\sqrt{\omega}\sqrt{\tau}\|_1^2$, Φ_d is defined in (2.1), and d is the dimension of the input system of $\mathcal{N}$.

Proof. Recall from (3.3) that the Doeblin coefficient can be expressed as

$$\alpha(\mathcal{N}) = \inf_{Y_{AB} \geq 0} \left\{ \mathrm{Tr}[Y_{AB} \Gamma^{\mathcal{N}}_{AB}] : I = \mathrm{Tr}_A[Y_{AB}] \right\}. \tag{4.3}$$

The conditions $Y_{AB} \geq 0$ and $I = \mathrm{Tr}_A[Y_{AB}]$ imply that Y_{AB} is the Choi operator of a completely positive unital (CPU) map from A to B. Thus,

$$\inf_{Y_{AB} \geq 0} \left\{ \mathrm{Tr}[Y_{AB} \Gamma^{\mathcal{N}}_{AB}] : I = \mathrm{Tr}_A[Y_{AB}] \right\}$$

$$= \inf_{\mathcal{M}\in\mathrm{CPU}} \mathrm{Tr}[\Gamma_{AB}^{\mathcal{M}}\Gamma_{AB}^{\mathcal{N}}] \tag{4.4}$$

$$= d^2 \inf_{\mathcal{M}\in\mathrm{CPU}} \mathrm{Tr}[(\mathrm{id}\otimes\mathcal{M})(\Phi_d)(\mathrm{id}\otimes\mathcal{N})(\Phi_d)] \tag{4.5}$$

$$= d^2 \inf_{\mathcal{M}\in\mathrm{CPU}} \mathrm{Tr}[\Phi_d(\mathrm{id}\otimes(\mathcal{M}^\dagger\circ\mathcal{N}))(\Phi_d)] \tag{4.6}$$

$$= d^2 \inf_{\mathcal{D}\in\mathrm{CPTP}} \mathrm{Tr}[\Phi_d(\mathrm{id}\otimes(\mathcal{D}\circ\mathcal{N}))(\Phi_d)] \tag{4.7}$$

$$= d^2 \inf_{\mathcal{D}\in\mathrm{CPTP}} F(\Phi_d, (\mathcal{D}\circ\mathcal{N})(\Phi_d)), \tag{4.8}$$

where in the penultimate equality we used the fact that the Hilbert–Schmidt adjoint of a unital map is a trace-preserving map. □

Remark 2. *By taking the decoding channel $\mathcal{D}$ in (4.2) to be completely depolarizing, a direct calculation verifies that the value is equal to one. Thus one may use the above to see that $\alpha(\mathcal{N}) \leq 1$ via this interpretation.*

Note that the analysis accommodates channels with a different input and output dimension (i.e., it is not required that $d_A = d_B$).

As indicated previously, the quantity on the right side of (4.2) is a counterpart of "quantum correlation" or "maximum achievable singlet fraction" considered in [51, Eq. (20)]. As such, it is thus reasonable to call it the "minimum achievable singlet fraction,"as we have done.

Remark 3 (Fully Quantum Counterpart of Exclusion)**.** *If the channel $\mathcal{N}$ is a classical–quantum channel of the form*

$$\mathcal{N}(\omega) = \sum_x \langle x|\omega|x\rangle \rho_x, \tag{4.9}$$

then after some direct calculations we obtain that the right-hand side of (4.2) becomes

$$\inf_{\mathcal{D}\in\mathrm{CPTP}} F(\Phi_d, (\mathrm{id}\otimes(\mathcal{D}\circ\mathcal{N}))(\Phi_d)) = \frac{1}{d^2} \inf_{(\Lambda_x)_x} \sum_x \mathrm{Tr}[\Lambda_x \rho_x], \tag{4.10}$$

where $(\Lambda_x)_x$ is a POVM, so that indeed we have recovered a fully quantum counterpart of the exclusion task, as considered in [8].

4.2 Exclusion Value of a Quantum Channel

The induced Doeblin coefficient of a quantum channel and its cone-restricted variants have similar mathematical structure to the communication value of a quantum channel [48] and its cone-restricted variants [50], respectively. Thus, one may expect that there is an operational interpretation for the quantum Doeblin coefficient analogous to the communication value. Indeed, the communication value of a quantum channel $\mathcal{N}$ may be expressed in terms of the minimal error probability of state discrimination over the quantum channel $\mathcal{N}$, which is one reason that it is a measure of the ability to communicate over the channel $\mathcal{N}$. It also generalizes to entanglement-assisted variations by using the cone-restricted variant [50].

In this section, we relate the quantum Doeblin coefficient to performing quantum state exclusion (QSE) over the channel. We do this by first observing that the induced Doeblin coefficient can be expressed in terms of the error probability of QSE, then extending generalizing this to the entanglement-assisted setting for cone-restricted Doeblin coefficients (Theorem 2).

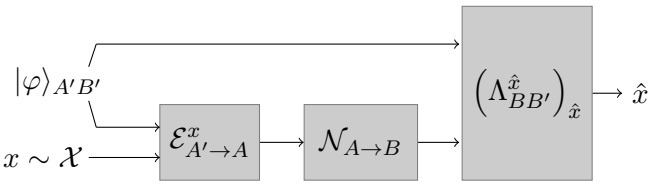


Figure 1: Depiction of the game that characterizes the value of $\alpha(\mathcal{N})$. One infimizes over choices of encoders and decoders, so it is minimizing the quantum state exclusion over all codes.

In particular, this proves that $\alpha(\mathcal{N})$ is the optimal classical Doeblin coefficient taken over all classical channels achievable using $\mathcal{N}$ and arbitrary entanglement assistance.

Recalling the definition of error probability for QSE (Definition 3), we observe that the induced Doeblin coefficient may be expressed as an infimization involving quantum state exclusion.

Proposition 10. *The induced Doeblin coefficient optimizes a rescaled version of state exclusion (with uniform prior) over all possible ensembles:*

$$\alpha_I(\mathcal{N}) = \inf_{n\in\mathbb{N}} n \cdot \inf_{(\rho^x)_{x\in[n]}} P_{\text{err}}\Big((n^{-1}, \mathcal{N}(\rho_x))_x\Big) \ . \tag{4.11}$$

Proof. By Proposition 6,

$$\alpha_I(\mathcal{N}) = \inf_{n\in\mathbb{N}} \inf_{\substack{(\Lambda_x)_{x\in[n]} \\ (\rho_x)_{x\in[n]}}} \sum_{x\in[n]} \text{Tr}[\Lambda_x \mathcal{N}(\rho_x)] \tag{4.12}$$

$$= \inf_{n\in\mathbb{N}} n \inf_{(\rho_x)_{x\in[n]}} \inf_{(\Lambda_x)_{x\in[n]}} \frac{1}{n} \sum_{x\in[n]} \text{Tr}[\Lambda_x \mathcal{N}(\rho_x)] \tag{4.13}$$

$$= \inf_{n\in\mathbb{N}} n \inf_{(\rho_x)_{x\in[n]}} P_{\text{err}}\left(\left(\frac{1}{n}, \mathcal{N}(\rho_x)\right)_x\right) \ , \tag{4.14}$$

concluding the proof. □

Our goal is now to generalize the above proposition by adding entanglement assistance and calling it a 'channel state exclusion game.' The game is depicted in Fig. 1.

Definition 9. *Given a cone $\mathcal{K}(B\otimes B') \subseteq \text{Pos}(B\otimes B')$, the channel state exclusion game (see Fig. 1) is characterized by*

$$\mathcal{G}^{\mathcal{K}}_{\text{QSE}}(\mathcal{N}) \coloneqq \inf_{\mathcal{X},(\Lambda^x)_x,(\mathcal{E}^x)_x,|\varphi\rangle} \sum_x P(x|x) \ , \tag{4.15}$$

where the infimum is over every decoder (POVM) $(\Lambda^x)_x$ such that each element is in the cone $\mathcal{K}$, every encoding channel $(\mathcal{E}^x)_x$, and

$$P(x'|x) \coloneqq \text{Tr}[\Lambda^{x'}_{BB'}(\mathcal{N}_{A\to B} \circ \mathcal{E}^x_{A'\to A})(\varphi_{A'B'})], \tag{4.16}$$

where $\varphi_{A'B'}$ is an arbitrary pure bipartite state.

This is a state exclusion game as, up to a rescaling by the input dimension, it optimizes the state exclusion error probability, with uniform prior, when using the channel $\mathcal{N}$. In other words, it is a generalization of Proposition 10 to the setting where there is entanglement-assistance.

Theorem 2. *Let $\mathcal{K}(A \otimes B) \subseteq \mathrm{Pos}(A \otimes B)$ be a closed, convex cone such that it is invariant under local CP maps on the A space and $I \in \mathrm{relint}(\mathcal{K})$. Then $\alpha_{I,\mathcal{K}}(\mathcal{N}) = \mathcal{G}^{\mathcal{K}}_{\mathrm{QSE}}(\mathcal{N})$.*

Proof. See Section A. □

As the proof is nearly identical to one in [50] which itself is a rather direct generalization of a proof in [48], we relegate it to Section A. We now state the most important case of Theorem 2, which makes use of the definition of the game (Definition 9) and (3.32).

Corollary 1. *The quantum Doeblin coefficient $\alpha(\mathcal{N})$ is equal to the best-case Doeblin coefficient optimized over all classical channels induced by $\mathcal{N}$ and arbitrary entanglement assistance. That is to say,*

$$\alpha(\mathcal{N}) = \inf_{\substack{\mathcal{X},\varphi\in\mathcal{S}(A'B') \\ (\mathcal{E}^x_{A'\to A})_x,\mathcal{D}_{BB'\to X}}} \alpha(\mathcal{W}^{(\mathcal{E}^x),\mathcal{D},\varphi}) \ , \tag{4.17}$$

where the infimization is over every finite alphabet $\mathcal{X}$, pure state $\varphi_{A'B'}$, encoder $(\mathcal{E}^x_{A'\to A})_{x\in\mathcal{X}}$, and decoder $\mathcal{D}_{BB'\to X}$, such that they induce a classical channel with conditional probabilities given by

$$\mathcal{W}^{(\mathcal{E}^x),\mathcal{D},\phi}(x'|x) \coloneqq \mathrm{Tr}[\Lambda^{x'}_{BB'}(\mathrm{id}_{B'} \otimes (\mathcal{N}_{A\to B} \circ \mathcal{E}^x_{A'\to A}))(\varphi_{A'B'})] \ . \tag{4.18}$$

We remark that, in the same way as identified in [50, Corollary 19] for the communication value, what Theorem 2 and Corollary 1 show is that $\alpha(\mathcal{N})$ and $\alpha_I(\mathcal{N})$ are interpolated between by providing arbitrary entanglement assistance, but limiting the entangling power of the decoder. We state this formally using the relevant cones, which generalizes the previous corollary.

Corollary 2. *For every $k \in \mathbb{N}$, define the entanglement-rank-k cone [25]*

$$\mathrm{Ent}_k(A:B) \coloneqq \{P_{AB} \geq 0 : \ \exists \mathcal{I} \ : \exists\{X_i\} \subset \mathrm{Lin}(A,B) : P = \sum_{i\in\mathcal{I}} \mathrm{vec}(X_i)\,\mathrm{vec}(X_i)^\dagger\} \ ,$$

where vec *is the 'vec mapping' (see, e.g., [25, Eq. (1.128)]). Then for channel $\mathcal{N}_{A\to B}$ and every $k \in \{1,\ldots,d\}$ where $d = \min\{|A|,|B|\}$*

$$\alpha_I(\mathcal{N}) = \alpha_{I,\mathrm{Ent}_1}(\mathcal{N}) \geq \alpha_{I,\mathrm{Ent}_k}(\mathcal{N}) \geq \alpha_{I,\mathrm{Ent}_d}(\mathcal{N}) = \alpha_{I,\mathrm{Pos}}(\mathcal{N}) = \alpha(\mathcal{N}) \ . \tag{4.19}$$

In other words, one interpolates between the quantum Doeblin coefficients $\alpha_I(\mathcal{N})$ and $\alpha(\mathcal{N})$ operationally by considering the channel state exclusion game (Definition 9) and interpolating between the decoder POVM being a separable measurement, $(\Lambda^x)_x \subset \mathcal{K}_1 = \mathrm{Sep}(A:B)$, and the decoder elements being arbitrary, $(\Lambda^x)_x \subset \mathcal{K}_d(A:B) = \mathrm{Pos}(A\otimes B)$ by restricting the POVM elements to a given entanglement rank cone k.

Proof. First, it is known [25, Section 6.1] that for Hilbert spaces A and B

$$\mathrm{Sep}(A:B) = \mathrm{Ent}_1(A:B) \subseteq \mathrm{Ent}_k(A:B) \subseteq \mathrm{Pos}(A \otimes B) = \mathrm{Ent}_{\min\{|A|,|B|\}}(A:B) \ ,$$

which implies the ordering between coefficients by (3.45). Second, it is known that the entanglement rank cones are closed under completely positive maps acting on one system (special case of [25, Theorem 6.23]). It is also known $I \in \mathrm{relint}(\mathrm{Ent}_k(A:B))$ for all k [50, Corollary 1]. Thus, the cones satisfy the conditions needed to apply Theorem 2. The proof is then completed by considering Proposition 7 and Definition 4. □

While the above claim uses a finite-sized hierarchy, it is not known to be efficiently computable. A similar claim may be made using the cones generated by the intersection of the PPT and symmetrically extendible cones, which are SDP representable. Such a claim was not made explicit for the communication value.

Corollary 3. *For every $k \in \mathbb{N}$, define the cone $\mathcal{K}_k(A:B) \coloneqq \mathrm{PPT}(A:B) \cap \mathrm{Sym}_k(A:B)$ where these sets are defined in (2.23) and Definition 2. Then for every $k \in \mathbb{N}$,*

$$\alpha_I(\mathcal{N}) = \alpha_{I,\mathrm{Sep}}(\mathcal{N}) \geq \alpha_{I,\mathcal{K}_k}(\mathcal{N}) \geq \alpha_{I,\mathcal{K}_1}(\mathcal{N}) = \alpha_{I,\mathrm{Pos}}(\mathcal{N}) = \alpha(\mathcal{N}) \ . \tag{4.20}$$

In other words, one operationally interpolates from the quantum Doeblin coefficient $\alpha(\mathcal{N})$ to $\alpha_I(\mathcal{N})$ by considering the channel state exclusion game (Definition 9) with the decoder being made up of POVM elements that are PPT and k-symmetrically extendible and increasing the value of k.

Proof. One can verify using the definition of the PPT cone and $\mathrm{Sym}_k(A:B)$ that if an element is contained in the cone, then it still is after applying a completely positive map to the A space. It follows that $\mathcal{K}_k(A:B)$ also satisfies this property. Moreover, for every $k \in \mathbb{N}$, $\mathrm{Sep}(A:B) \subseteq \mathcal{K}_k(A:B)$ and $I \in \mathrm{relint}(\mathrm{Sep}(A:B))$ due to [52] (see [50, Corollary 1] to see this proven explicitly). Thus, for all $k \in \mathbb{N}$, $\mathcal{K}_k(A:B)$ satisfies the conditions on the cone to apply Theorem 2. The claimed inequalities and equalities then follow from Proposition 7, Definition 4, and that $\mathrm{Pos}(A \otimes B) = \mathrm{Sym}_1(A:B) \supseteq \mathrm{Sym}_2(A:B) \supseteq \cdots \supseteq \mathrm{Sym}_k(A:B) \supseteq \mathrm{Sep}(A:B)$ for every $k \in \mathbb{N}$. □

4.3 Reverse Max-Mutual Information

In this section, we observe that the Doeblin coefficient $\alpha(\mathcal{N})$ can be written in terms of the reverse max-mutual information $\inf_{\tau \in \mathrm{aff}(\mathcal{D})} D_{\max}(\mathcal{R}^\tau \| \mathcal{N})$ of the channel $\mathcal{N}$ (alternatively, this quantity can be called the Doeblin information of the channel). This is a counterpart of the fact that the entanglement-assisted communication value can be written in terms of the max-mutual information of the channel (see Remark 4 below). The formulation given in Proposition 11 can also be compared to the concept of resource weight [53–55], studied in the context of quantum resource theories, especially if one considers the replacer map $\mathcal{R}^\tau$ to play the role of a free map.

Proposition 11. *The following equality holds for a quantum channel $\mathcal{N}$:*

$$\alpha(\mathcal{N}) = \exp\left(- \inf_{\tau \in \mathrm{aff}(\mathcal{D})} D_{\max}(\mathcal{R}^\tau \| \mathcal{N}) \right), \tag{4.21}$$

where $D_{\max}(\mathcal{R}^\tau \| \mathcal{N})$ is defined from (2.12) and $\mathcal{R}^\tau(\cdot) \coloneqq \mathrm{Tr}[\cdot]\tau$ is a replacer superoperator.

Proof. Consider that

$$\alpha(\mathcal{N}) = \sup_{X \in \text{Herm}} \left\{ \text{Tr}[X] : \mathcal{R}^X \leq \mathcal{N} \right\} \tag{4.22}$$

$$= \sup_{\substack{\lambda \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \left\{ \text{Tr}[\lambda \tau] : \mathcal{R}^{\lambda \tau} \leq \mathcal{N} \right\} \tag{4.23}$$

$$= \sup_{\substack{\lambda \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \{ \lambda : \lambda \mathcal{R}^\tau \leq \mathcal{N} \} \tag{4.24}$$

$$= \sup_{\substack{\mu \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \left\{ \frac{1}{\mu} : \frac{1}{\mu} \mathcal{R}^\tau \leq \mathcal{N} \right\} \tag{4.25}$$

$$= \left[\inf_{\substack{\mu \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \{ \mu : \mathcal{R}^\tau \leq \mu \mathcal{N} \} \right]^{-1} \tag{4.26}$$

$$= \left[\inf_{\tau \in \text{aff}(\mathcal{D})} \exp(D_{\max}(\mathcal{R}^\tau \| \mathcal{N})) \right]^{-1} \tag{4.27}$$

$$= \exp\left(- \inf_{\tau \in \text{aff}(\mathcal{D})} D_{\max}(\mathcal{R}^\tau \| \mathcal{N}) \right). \tag{4.28}$$

The second equality follows because $X = 0$ is a feasible solution, so that it suffices to restrict the optimization to be over $X \in \text{Herm}$ such that $\text{Tr}[X] \geq 0$. Furthermore, all such X can be written as $X = \lambda \tau$, where $\lambda \geq 0$ and $\tau \in \text{aff}(\mathcal{D})$. The penultimate equality follows from the definition of $D_{\max}$ in (2.8). □

Remark 4. *By following a proof similar to the above, we note that the entanglement-assisted communication value* $\text{cv}^*(\mathcal{N})$ *of a quantum channel* $\mathcal{N}$ *(see [48, Theorem 6] for a mathematical expression) can be written in terms of the max-mutual information:*

$$\ln \text{cv}^*(\mathcal{N}) = \inf_{\sigma \in \mathcal{D}} D_{\max}(\mathcal{N} \| \mathcal{R}^\sigma). \tag{4.29}$$

4.4 Reverse Robustness

In this section, we interpret the quantum Doeblin coefficient $\alpha(\mathcal{N})$ as a reverse robustness, related to robustness quantities previously considered in quantum information (see, e.g., [54, 55]). Here, the goal is to express the original channel as a convex combination of a trace-preserving replacer map and an another arbitrary channel, with as much weight as possible on the replacer map. As before, we allow the replacer map to be non-physical and replace the input state with an arbitrary quasi-state. This characterization is helpful in establishing $1 - \alpha(\mathcal{N})$ as an upper bound on the trace distance contraction coefficient of $\mathcal{N}$ (Proposition 30), as it neatly separates $\mathcal{N}$ into a component that has a trivial action on an arbitrary input and a component that does not. This characterization is also helpful in establishing Proposition 19 below, where we use the following slightly different representation of the quantum Doeblin coefficient:

$$1 - \alpha(\mathcal{N}) = \inf_{\substack{\lambda \in [0,1], \\ \tau \in \text{aff}(\mathcal{D}), \\ \mathcal{M} \in \text{CPTP}}} \{ \lambda : (1 - \lambda) \mathcal{R}^\tau + \lambda \mathcal{M} = \mathcal{N} \}, \tag{4.30}$$

which is equivalent to (4.31) below.

Proposition 12. *The following equality holds for a quantum channel $\mathcal{N}$:*

$$\alpha(\mathcal{N}) = \sup_{\substack{\lambda\in[0,1],\\ \tau\in\text{aff}(\mathcal{D}),\\ \mathcal{M}\in\text{CPTP}}} \left\{\lambda : \mathcal{N} = \lambda\mathcal{R}^\tau + (1-\lambda)\,\mathcal{M}\right\}. \tag{4.31}$$

Proof. Consider that

$$\begin{aligned}
\alpha(\mathcal{N}) &= \sup_{\substack{\lambda\geq 0,\\ \tau\in\text{aff}(\mathcal{D})}} \left\{\lambda : \lambda\mathcal{R}^\tau \leq \mathcal{N}\right\} && (4.32)\\
&= \sup_{\substack{\lambda\geq 0,\\ \tau\in\text{aff}(\mathcal{D}),\\ \mathcal{M}'\in\text{CP}}} \left\{\lambda : \lambda\mathcal{R}^\tau + \mathcal{M}' = \mathcal{N}\right\} && (4.33)\\
&= \sup_{\substack{\lambda\in[0,1],\\ \tau\in\text{aff}(\mathcal{D}),\\ \mathcal{M}\in\text{CPTP}}} \left\{\lambda : \mathcal{N} = \lambda\mathcal{R}^\tau + (1-\lambda)\,\mathcal{M}\right\}. && (4.34)
\end{aligned}$$

The first equality follows from (4.24). The second equality follows because the inequality $\lambda\mathcal{R}^\tau \leq \mathcal{N}$ is equivalent to the existence of a completely positive map $\mathcal{M}'$ such that $\mathcal{M}' = \mathcal{N} - \lambda\mathcal{R}^\tau$. The final equality follows because the trace-preserving constraint on $\mathcal{N}$ implies that the following equality holds for every state ρ:

$$\begin{aligned}
1 = \text{Tr}[\rho] &= \text{Tr}[\mathcal{N}(\rho)] && (4.35)\\
&= \text{Tr}[\left(\lambda\mathcal{R}^\tau + \mathcal{M}'\right)(\rho)] && (4.36)\\
&= \lambda\,\text{Tr}[\mathcal{R}^\tau(\rho)] + \text{Tr}[\mathcal{M}'(\rho)] && (4.37)\\
&= \lambda\,\text{Tr}[\tau] + \text{Tr}[\mathcal{M}'(\rho)] && (4.38)\\
&= \lambda + \text{Tr}[\mathcal{M}'(\rho)]. && (4.39)
\end{aligned}$$

Since $\mathcal{M}'$ is completely positive, it follows that $\text{Tr}[\mathcal{M}'(\rho)] \geq 0$, so that $\lambda \in [0,1]$. Furthermore, it then follows that $\text{Tr}[\mathcal{M}'(\rho)] = 1-\lambda \in [0,1]$ for every input state ρ. Thus, in the case that $\lambda \in [0,1)$, it follows that $\mathcal{M} \coloneqq \frac{1}{1-\lambda}\mathcal{M}'$ is a completely positive trace-preserving map, and so the last equality holds, concluding the proof for $\lambda \in [0,1)$. In the case that $\lambda = 1$, it follows that $\text{Tr}[\mathcal{M}'(\rho)] = 0$ for every input state ρ, and thus $\mathcal{M}'$ is the zero map in this case (i.e., such that $\mathcal{M}'(\rho) = 0$ for every input state). Thus, $\mathcal{N} = \mathcal{R}^\tau$ in this case (i.e., $\mathcal{N}$ is a replacer channel), concluding the proof. □

4.5 Reverse Hypothesis Testing Mutual Information

In this section, we establish a relationship between the Doeblin coefficient and the extended hypothesis testing relative entropy defined in (3.1) and (2.6), respectively. Indeed, expressing the Doeblin coefficient in this way is helpful for relating it to other distinguishability measures, such as the extended sandwiched Rényi relative entropy, as done in Section E.

Proposition 13. *Let $\mathcal{N}$ be a quantum channel. We have that*

$$-\ln\alpha(\mathcal{N}) = -2\ln d + \inf_{\tau\in\text{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}\left(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}\right), \tag{4.40}$$

where $\mathrm{aff}(\mathcal{D})$ *is the affine hull of the set of density operators (i.e., Hermitian operators with unit trace),* $D_H^{1-\frac{1}{d^2}}$ *is given in* (2.6)*,* $\pi_d \coloneqq I/d$*, and* $\Phi^{\mathcal{N}} \coloneqq \frac{1}{d}\Gamma^{\mathcal{N}}_{AB}$ *with* $\Gamma^{\mathcal{N}}_{AB}$ *in* (2.2)*.*

Proof. See Section B. □

4.6 Interpretations of the Induced Quantum Doeblin Coefficient

In this subsection, we show how to lift the interpretations for $\alpha(\mathcal{N})$ to $\alpha_I(\mathcal{N})$.

4.6.1 Minimum Achievable Singlet Fraction

We begin by showing that $\alpha_{I,\mathcal{K}}(\mathcal{N})$, as defined in (3.51), inherits a singlet fraction interpretation almost immediately in the same manner as [50, Proposition 12]. This recovers Theorem 1 as a specific case. It also establishes $\alpha_I(\mathcal{N})$ as having the same singlet-fraction interpretation but with the minimization over *entanglement-breaking channels* [56].

Proposition 14. *Let* $\mathcal{K}(A:B) \subset \mathrm{Pos}(A \otimes B)$ *be such that* I_{AB} *is in the relative interior of* $\mathcal{K}(A:B)$*. Then,*

$$\alpha_{I,\mathcal{K}}(\mathcal{N}) \coloneqq d^2 \min_{\mathcal{E}\in\mathrm{CPTP}(B,A):\Gamma^{\mathcal{E}^\dagger}\in\mathcal{K}} \mathrm{Tr}[\Phi_d(\mathrm{id}_A \otimes (\mathcal{E}\circ\mathcal{N}))(\Phi_d)] \ , \tag{4.41}$$

where d *denotes the dimension of the input system of* $\mathcal{N}$*. In particular,*

$$\alpha_I(\mathcal{N}) = d^2 \min_{\mathcal{E}\in\mathrm{EB}} \mathrm{Tr}[\Phi_d(\mathrm{id}_A \otimes (\mathcal{E}\circ\mathcal{N}))(\Phi_d)] \ , \tag{4.42}$$

where EB *denotes the set of entanglement-breaking channels.*

Proof. The condition that I_{AB} is in the relative interior of $\mathcal{K}(A:B)$ guarantees that strong duality holds, so that we may work with the dual problem. Note that $\mathrm{Tr}_A[Y] = I_B$ implies that $Y = \Gamma^{\mathcal{M}}$ for some completely positive, unital (CPU) map $\mathcal{M}: A \to B$. Thus, similar to (4.4)–(4.8), we have that

$$\begin{aligned}
\alpha_{I,\mathcal{K}}(\mathcal{N}) &= \min_{\mathcal{M}\in\mathrm{CPU}:\Gamma^{\mathcal{M}}\in\mathcal{K}} \mathrm{Tr}[\Gamma^{\mathcal{N}}\Gamma^{\mathcal{M}}] && (4.43)\\
&= d \min_{\mathcal{M}\in\mathrm{CPU}:\Gamma^{\mathcal{M}}\in\mathcal{K}} \mathrm{Tr}[\Gamma^{\mathcal{N}}(\mathrm{id}_A \otimes \mathcal{M})(\Phi_d)] && (4.44)\\
&= d \min_{\mathcal{M}\in\mathrm{CPU}:\Gamma^{\mathcal{M}}\in\mathcal{K}} \mathrm{Tr}[\mathcal{M}^\dagger(\Gamma^{\mathcal{N}})\Phi_d] && (4.45)\\
&= d^2 \min_{\mathcal{M}\in\mathrm{CPU}:\Gamma^{\mathcal{M}}\in\mathcal{K}} \mathrm{Tr}\left[\left(\mathcal{M}^\dagger_{B\to A}\circ\mathcal{N}_{A\to B}(\Phi_d)\right)(\Phi_d)\right] \ . && (4.46)
\end{aligned}$$

To conclude the proof, note that $\mathcal{M}^\dagger$ is a CPTP map as it is the adjoint of a CPU map and we replace optimizing over $\mathcal{M}^\dagger$ with $\mathcal{E}$ for notational clarity.

To obtain the statement in (4.42), recall that $\Gamma^{\mathcal{E}^\dagger}_{AB} = T_{AB}(\Gamma^{\mathcal{E}}_{BA})$ where T_{AB} the transpose map acting on the joint space, and $\Gamma^{\mathcal{E}}_{BA}$ is $\Gamma^{\mathcal{E}}_{AB}$, with the spaces swapped [50]. It follows that if $\Gamma^{\mathcal{E}^\dagger} \in \mathrm{Sep}(A:B)$, then $\Gamma^{\mathcal{E}} \in \mathrm{Sep}(A:B)$. As a CPTP map $\mathcal{E}$ is entanglement breaking if and only if $\Gamma^{\mathcal{E}} \in \mathrm{Sep}(A:B)$, this completes the proof. □

4.6.2 Max-oveloH Information

In this section, we represent the induced Doeblin coefficient $\alpha_I(\mathcal{N})$ in terms of the max-oveloH information of the channel $\mathcal{N}$. This is a counterpart of the fact that the communication value of a channel can be written in terms of the max-Holevo information [48, Theorem 1].

Proposition 15. *The following equality holds for a quantum channel $\mathcal{N}$:*

$$\alpha_I(\mathcal{N}) = \exp\left(- \inf_{\tau \in \text{aff}(\mathcal{D})} \sup_{\rho \in \mathcal{D}} D_{\max}(\tau \| \mathcal{N}(\rho)) \right). \tag{4.47}$$

Proof. Consider that

$$\begin{aligned}
\alpha_I(\mathcal{N}) &= \sup_{X \in \text{Herm}} \left\{ \text{Tr}[X] : \mathcal{R}^X \leq_P \mathcal{N} \right\} && (4.48)\\
&= \sup_{\substack{\lambda \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \left\{ \text{Tr}[\lambda \tau] : \mathcal{R}^{\lambda \tau} \leq_P \mathcal{N} \right\} && (4.49)\\
&= \sup_{\substack{\lambda \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \{ \lambda : \lambda \mathcal{R}^\tau \leq_P \mathcal{N} \} && (4.50)\\
&= \sup_{\substack{\mu \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \left\{ \frac{1}{\mu} : \frac{1}{\mu} \mathcal{R}^\tau \leq_P \mathcal{N} \right\} && (4.51)\\
&= \left[\inf_{\substack{\mu \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \{ \mu : \mathcal{R}^\tau \leq_P \mu \mathcal{N} \} \right]^{-1} && (4.52)\\
&= \left[\inf_{\substack{\mu \geq 0, \\ \tau \in \text{aff}(\mathcal{D})}} \{ \mu : \mathcal{R}^\tau(\rho) \leq \mu \mathcal{N}(\rho) \ \forall \rho \in \mathcal{D} \} \right]^{-1} && (4.53)\\
&= \left[\inf_{\tau \in \text{aff}(\mathcal{D})} \inf_{\mu \geq 0} \{ \mu : \tau \leq \mu \mathcal{N}(\rho) \ \forall \rho \in \mathcal{D} \} \right]^{-1} && (4.54)\\
&= \left[\inf_{\tau \in \text{aff}(\mathcal{D})} \inf_{\mu \geq 0} \sup_{\rho \in \mathcal{D}} \{ \mu : \tau \leq \mu \mathcal{N}(\rho) \} \right]^{-1} && (4.55)\\
&= \left[\inf_{\tau \in \text{aff}(\mathcal{D})} \sup_{\rho \in \mathcal{D}} \inf_{\mu \geq 0} \{ \mu : \tau \leq \mu \mathcal{N}(\rho) \} \right]^{-1} && (4.56)\\
&= \left[\inf_{\tau \in \text{aff}(\mathcal{D})} \sup_{\rho \in \mathcal{D}} \exp(D_{\max}(\tau \| \mathcal{N}(\rho))) \right]^{-1} && (4.57)\\
&= \exp\left(- \inf_{\tau \in \text{aff}(\mathcal{D})} \sup_{\rho \in \mathcal{D}} D_{\max}(\tau \| \mathcal{N}(\rho)) \right). && (4.58)
\end{aligned}$$

The equality in (4.55) holds because the constraint should hold for all $\rho \in \mathcal{D}$. The equality in (4.56) holds by applying the extended Sion minimax theorem from [57, Theorem 2.11]; indeed, the objective function is linear in μ and ρ, the set of μ satisfying $\mu \geq 0, \tau \leq \mu \mathcal{N}(\rho)$ is convex, and the set of ρ satisfying $\rho \in \mathcal{D}, \tau \leq \mu \mathcal{N}(\rho)$ is convex and compact. The penultimate equality follows from applying the definition of $D_{\max}$ in (2.8). □

4.6.3 Reverse Robustness

Here we write the induced Doeblin coefficient $\alpha_I(\mathcal{N})$ as a reverse robustness, analogous to what we found in Proposition 12. The main difference between Proposition 16 and Proposition 12 is that Proposition 16 involves an optimization over the set of positive, trace-preserving maps.

Proposition 16. *The following equality holds for a quantum channel $\mathcal{N}$:*

$$\alpha_I(\mathcal{N}) = \sup_{\substack{\lambda\in[0,1],\\ \tau\in\mathrm{aff}(\mathcal{D}),\\ \mathcal{M}\in\mathrm{PTP}}} \{\lambda : \mathcal{N} = \lambda\mathcal{R}^\tau + (1-\lambda)\,\mathcal{M}\}\,, \tag{4.59}$$

where PTP *denotes the set of positive, trace-preserving maps.*

Proof. The proof is identical to the proof of Proposition 12, with the main exception being that we use the ordering induced by positive maps rather than completely positive maps. □

4.6.4 Hypothesis-testing oveloH information

Here we write the induced Doeblin coefficient $\alpha_I(\mathcal{N})$ in terms of the hypothesis testing oveloH information of the channel $\mathcal{N}$, analogously to what we did in Proposition 13 for the Doeblin coefficient $\alpha(\mathcal{N})$.

Proposition 17. *The following equality holds for a quantum channel $\mathcal{N}$:*

$$-\ln\alpha_I(\mathcal{N}) = -2\ln d + \sup_{(\psi^x)_{x=1}^{d^2}} \inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}(\rho_X\otimes\tau_B\|\rho_{XB}), \tag{4.60}$$

where d is the dimension of the channel input system and each ψ_A^x is a pure state for all $x\in[d^2]$ and

$$\rho_{XB} \coloneqq \frac{1}{d^2}\sum_{x=1}^{d^2} |x\rangle\langle x|\otimes\mathcal{N}_{A\to B}\left(\psi_A^x\right). \tag{4.61}$$

Proof. Consider that

$$-\ln\alpha_I(\mathcal{N}) = -\ln \inf_{(\psi^x,\Lambda^x)_{x=1}^{d^2}} \sum_{x=1}^{d^2} \mathrm{Tr}[\Lambda_B^x\mathcal{N}_{A\to B}(\psi_A^x)] \tag{4.62}$$

$$= -\ln\left(d^2 \inf_{(\psi^x)_{x=1}^{d^2}} \inf_{(\Lambda^x)_{x=1}^{d^2}} \sum_{x=1}^{d^2} \frac{1}{d^2}\,\mathrm{Tr}[\Lambda_B^x\mathcal{N}_{A\to B}(\psi_A^x)]\right) \tag{4.63}$$

$$= -2\ln d + \sup_{(\psi^x)_{x=1}^{d^2}} \left(-\ln \inf_{(\Lambda^x)_{x=1}^{d^2}} \sum_{x=1}^{d^2} \frac{1}{d^2}\,\mathrm{Tr}[\Lambda_B^x\mathcal{N}_{A\to B}(\psi_A^x)]\right) \tag{4.64}$$

$$= -2\ln d + \sup_{(\psi^x)_{x=1}^{d^2}} \inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}(\rho_X\otimes\tau_B\|\rho_{XB}) \tag{4.65}$$

The first equality follows from Proposition 6, and the last equality follows from [32, Proposition 7]. □

5 Properties of Quantum Doeblin Coefficients

5.1 Normalization

Proposition 18 (Normalization)**.** *For a quantum channel $\mathcal{N}$, the following inequalities hold:*

$$0 \leq \alpha(\mathcal{N}) \leq 1. \tag{5.1}$$

Moreover, $\alpha(\mathcal{N}) = 1$ if and only if $\mathcal{N}$ is a replacer channel, and $\alpha(\mathcal{N}) = 0$ if and only if $\mathcal{N}$ is orthogonal to a completely positive unital map under the Hilbert–Schmidt inner product through the Choi isomorphism.

Proof. To establish the inequality $0 \leq \alpha(\mathcal{N})$, note that $X_B = 0$ is feasible in the primal SDP in (3.1), and the objective function evaluates to zero for this choice. To establish the inequality $\alpha(\mathcal{N}) \leq 1$, we note that $Y_{AB} = \frac{1}{d_A} I_A \otimes I_B$ is feasible for the dual SDP in (3.3), and the objective function evaluates to one for this choice.

We now establish the saturation conditions. We begin by establishing that $\alpha(\mathcal{N}) = 1$ if and only if $\mathcal{N}$ is a replacer channel. Let $\mathcal{N}$ be a replacer channel. Then $\Gamma^{\mathcal{N}} = I_A \otimes \sigma_B$ for some $\sigma_B \in \mathcal{D}(B)$. Then $X_B = \sigma_B$ is feasible for (3.1), and thus $\alpha(\mathcal{N}) = 1$. For the other direction, let $\alpha(\mathcal{N}) = 1$. Recall that $1 - \alpha(\mathcal{N}) \geq \eta_{\mathrm{Tr}}(\mathcal{N})$ [11, 14] (see also Proposition 30 for a strengthening), and so $\eta_{\mathrm{Tr}}(\mathcal{N}) = 0$. It follows that $\sup_{\rho,\sigma\in\mathcal{D}} \|\mathcal{N}(\rho) - \mathcal{N}(\sigma)\|_1 = 0$. Thus, $\mathcal{N}(\rho) = \mathcal{N}(\sigma)$ for all $\rho, \sigma \in \mathcal{D}$. It follows that $\mathcal{N}$ is a replacer channel.

The case that $\alpha(\mathcal{N}) = 0$ is in effect an immediate consequence of the dual problem (3.3). By the Choi isomorphism, the set $\{Y_{AB} \geq 0 : \mathrm{Tr}_A[Y_{AB}] = I_B\}$ consists of Choi operators of the space of completely positive, unital maps from A to B. Note that, for fixed spaces A and B, this is a compact set, and so we may minimize over it. Thus, by (3.3), $\alpha(\mathcal{N}) = 0$ if and only if $\min_{\mathcal{E}\in\mathrm{CPU}} \mathrm{Tr}[\Gamma^{\mathcal{N}}\Gamma^{\mathcal{E}}] = 0$, which means it is orthogonal to some CPU map under the Hilbert–Schmidt inner product. □

We note that while the saturation for the lower bound is more mathematical than physical, it is relevant. Namely, an example of such a channel that is noisy but satisfies this condition will be considered in Example 1. Another way of stating the saturation condition is based on the minimum achievable singlet fraction interpretation from Theorem 1: the Doeblin coefficient $\alpha(\mathcal{N}) = 0$ if and only if there exists a decoding channel $\mathcal{D}$ acting on the channel output of the Choi state $\Phi^{\mathcal{N}}$ to make it orthogonal to the maximally entangled state Φ.

5.2 Concatenation

In this section, we consider how the quantum Doeblin coefficients $\alpha(\mathcal{N})$ and $\alpha_I(\mathcal{N})$ behave under sequential composition of channels. In particular, we find that the quantities $1 - \alpha(\mathcal{N}_2 \circ \mathcal{N}_1)$ and $1 - \alpha_I(\mathcal{N}_2 \circ \mathcal{N}_1)$ are submultiplicative under such a sequential composition, and we note that this property is strongly related to the behavior of the contraction coefficient of a quantum channel, studied later on in Section 6.1.

Proposition 19 (Concatenation)**.** *For channels $\mathcal{N}_1$ and $\mathcal{N}_2$, the following inequalities hold:*

$$1 - \alpha(\mathcal{N}_2 \circ \mathcal{N}_1) \leq (1 - \alpha(\mathcal{N}_1))\,(1 - \alpha(\mathcal{N}_2))\,, \tag{5.2}$$

$$1 - \alpha_I(\mathcal{N}_2 \circ \mathcal{N}_1) \leq (1 - \alpha_I(\mathcal{N}_1))\,(1 - \alpha_I(\mathcal{N}_2))\,. \tag{5.3}$$

Proof. The proof is analogous to that given for [13, Corollary 2.2] in the classical case, and it makes use of the representation in (4.30). For $i \in \{1,2\}$, let $\lambda_i \in [0,1]$, $\tau_i \in \operatorname{aff}(\mathcal{D})$, and $\mathcal{M}_i \in \mathrm{CPTP}$ be such that $(1-\lambda_i)\,\mathcal{R}_{\tau_i} + \lambda_i \mathcal{M}_i = \mathcal{N}_i$. Then it follows that

$$\begin{aligned}
\mathcal{N}_2 \circ \mathcal{N}_1 &= [(1-\lambda_2)\,\mathcal{R}_{\tau_2} + \lambda_2 \mathcal{M}_2] \circ [(1-\lambda_1)\,\mathcal{R}_{\tau_1} + \lambda_1 \mathcal{M}_1] && (5.4)\\
&= (1-\lambda_2)\,(1-\lambda_1)\,\mathcal{R}_{\tau_2} \circ \mathcal{R}_{\tau_1} + \lambda_2\,(1-\lambda_1)\,\mathcal{M}_2 \circ \mathcal{R}_{\tau_1} \\
&\quad + (1-\lambda_2)\,\lambda_1 \mathcal{R}_{\tau_2} \circ \mathcal{M}_1 + \lambda_2 \lambda_1 \mathcal{M}_2 \circ \mathcal{M}_1 && (5.5)\\
&= (1-\lambda_2)\,(1-\lambda_1)\,\mathcal{R}_{\tau_2} + \lambda_2\,(1-\lambda_1)\,\mathcal{R}_{\mathcal{M}_2(\tau_1)} \\
&\quad + (1-\lambda_2)\,\lambda_1 \mathcal{R}_{\tau_2} + \lambda_2 \lambda_1 \mathcal{M}_2 \circ \mathcal{M}_1 && (5.6)\\
&= (1-\lambda_2)\,\mathcal{R}_{\tau_2} + \lambda_2\,(1-\lambda_1)\,\mathcal{R}_{\mathcal{M}_2(\tau_1)} + \lambda_2 \lambda_1 \mathcal{M}_2 \circ \mathcal{M}_1 && (5.7)\\
&= (1-\lambda_1\lambda_2)\,\mathcal{R}_{\tau'} + \lambda_2 \lambda_1 \mathcal{M}_2 \circ \mathcal{M}_1, && (5.8)
\end{aligned}$$

where

$$\tau' := \frac{1-\lambda_2}{1-\lambda_1\lambda_2}\tau_2 + \frac{\lambda_2\,(1-\lambda_1)}{1-\lambda_1\lambda_2}\mathcal{M}_2(\tau_1). \tag{5.9}$$

Thus, the choices $\lambda_2\lambda_1$, τ', and $\mathcal{M}_2 \circ \mathcal{M}_1$ are feasible for $1-\alpha(\mathcal{N}_2 \circ \mathcal{N}_1)$, and it follows that

$$1 - \alpha(\mathcal{N}_2 \circ \mathcal{N}_1) \leq \lambda_2 \lambda_1. \tag{5.10}$$

Since $\lambda_i \in [0,1]$, $\tau_i \in \operatorname{aff}(\mathcal{D})$, and $\mathcal{M}_i \in \mathrm{CPTP}$ are arbitrary feasible choices for $1-\alpha(\mathcal{N}_i)$, the desired inequality follows.

The proof for (5.3) proceeds in exactly the same way, but instead using the representation in Proposition 16. □

Corollary 4. *For a channel $\mathcal{N}$ and $n \in \mathbb{N}$, the following inequalities hold:*

$$1 - \alpha(\mathcal{N}^{\circ n}) \leq (1-\alpha(\mathcal{N}))^n, \tag{5.11}$$

$$1 - \alpha_I(\mathcal{N}^{\circ n}) \leq (1-\alpha_I(\mathcal{N}))^n. \tag{5.12}$$

Proof. Direct consequence of Proposition 19. □

5.3 Data Processing under Pre- and Post-Processing

Next, we show that both Doeblin coefficients, $\alpha(\mathcal{N})$ and $\alpha_I(\mathcal{N})$, of a channel $\mathcal{N}$ can only stay the same or increase under pre- and post-processing, and they are invariant if both these pre- and post-processing channels are isometric-reversal and isometric channels, respectively.

Given an isometric channel $\mathcal{U}$, define the corresponding isometric-reversal channel as [23, Eq. (4.264)]

$$\mathcal{P}^{\mathcal{U}}(\cdot) := \mathcal{U}^\dagger(\cdot) + \operatorname{Tr}[(\mathrm{id} - \mathcal{U}^\dagger)(\cdot)]\omega, \tag{5.13}$$

where ω is an arbitrary state. It then follows that $\mathcal{P}^{\mathcal{U}}$ indeed reverses the action of $\mathcal{U}$; i.e.,

$$\mathcal{P}^{\mathcal{U}} \circ \mathcal{U} = \mathrm{id}\,. \tag{5.14}$$

Proposition 20 (Data Processing under Pre- and Post-Processing and Isometric Invariance)**.** *Let $\mathcal{E}_{B\to B'}$, $\mathcal{G}_{A'\to A}$, and $\mathcal{N}_{A\to B}$ be quantum channels. Then*

$$\alpha(\mathcal{E} \circ \mathcal{N} \circ \mathcal{G}) \geq \alpha(\mathcal{N}), \tag{5.15}$$

$$\alpha_I(\mathcal{E} \circ \mathcal{N} \circ \mathcal{G}) \geq \alpha_I(\mathcal{N}). \tag{5.16}$$

That is, the quantum Doeblin coefficient does not decrease under pre- and post-processing. Moreover, the quantum Doeblin coefficient is invariant under isometric-reversal and isometric pre- and post-processing, respectively; i.e.,

$$\alpha(\mathcal{N}) = \alpha(\mathcal{U} \circ \mathcal{N} \circ \mathcal{P}^{\mathcal{V}}), \tag{5.17}$$
$$\alpha_I(\mathcal{N}) = \alpha_I(\mathcal{U} \circ \mathcal{N} \circ \mathcal{P}^{\mathcal{V}}), \tag{5.18}$$

where $\mathcal{U}$ and $\mathcal{V}$ are isometric channels.

Proof. To establish the inequality in (5.16), consider that

$$\alpha_I(\mathcal{E} \circ \mathcal{N} \circ \mathcal{F}) = \inf_{(\rho_x, \Lambda_x)_x} \sum_x \operatorname{Tr}[\Lambda_x (\mathcal{E} \circ \mathcal{N} \circ \mathcal{F})(\rho_x)] \tag{5.19}$$
$$= \inf_{(\rho_x, \Lambda_x)_x} \sum_x \operatorname{Tr}[\mathcal{E}^\dagger(\Lambda_x) \mathcal{N}(\mathcal{F}(\rho_x))] \tag{5.20}$$
$$\geq \inf_{(\Lambda_y, \rho_y)_y} \sum_y \operatorname{Tr}[\Lambda_y \mathcal{N}(\rho_y)] \tag{5.21}$$
$$= \alpha_I(\mathcal{N}), \tag{5.22}$$

where we made use of (3.29). The equality in (5.18) follows because

$$\alpha_I(\mathcal{U} \circ \mathcal{N} \circ \mathcal{P}^{\mathcal{V}}) \geq \alpha_I(\mathcal{N}) \tag{5.23}$$
$$\alpha_I(\mathcal{N}) = \alpha_I(\mathcal{P}^{\mathcal{U}} \circ \mathcal{U} \circ \mathcal{N} \circ \mathcal{P}^{\mathcal{V}} \circ \mathcal{V}) \geq \alpha_I(\mathcal{U} \circ \mathcal{N} \circ \mathcal{P}^{\mathcal{V}}), \tag{5.24}$$

as a consequence of (5.16) and (5.14). By appealing to Corollary 1, we can employ the same proofs to establish the claims for $\alpha(\mathcal{N})$. □

5.4 Submultiplicativity and Multiplicativity

In Proposition 19, we considered channels being applied sequentially. We now consider when they are applied in parallel. This means that we are interested in the multiplicativity of the quantum Doeblin coefficient $\alpha(\mathcal{N})$. First, we show that it is submultiplicative.

Proposition 21 (Submultiplicativity)**.** *Let $\mathcal{N}$ and $\mathcal{M}$ be quantum channels. The following submultiplicativity inequalities hold:*

$$\alpha(\mathcal{N} \otimes \mathcal{M}) \leq \alpha(\mathcal{N})\alpha(\mathcal{M}), \tag{5.25}$$
$$\alpha_I(\mathcal{N} \otimes \mathcal{M}) \leq \alpha_I(\mathcal{N})\alpha_I(\mathcal{M}). \tag{5.26}$$

Proof. Let Y^1, Y^2 be optimizers for $\mathcal{N}, \mathcal{M}$, respectively, to the dual program given in (3.3). Then $Y^1 \otimes Y^2$ is feasible for $\mathcal{N} \otimes \mathcal{M}$, satisfying $\operatorname{Tr}_{A_1A_2}[Y^1 \otimes Y^2] = I_{B_1} \otimes I_{B_2} = I_{B_1B_2}$, with value

$$\operatorname{Tr}[\Gamma^{\mathcal{N}\otimes\mathcal{M}}(Y^1 \otimes Y^2)] = \operatorname{Tr}[(\Gamma^{\mathcal{N}} \otimes \Gamma^{\mathcal{M}})(Y^1 \otimes Y^2)] = \alpha(\mathcal{N})\alpha(\mathcal{M}). \tag{5.27}$$

As (3.3) involves a minimization, we conclude that this is an upper bound. The same proof works for α_I using (3.45) and that if $Y^1 \in \operatorname{Sep}(A_1 : B_1)$ and $Y^2 \in \operatorname{Sep}(A_2 : B_2)$, then $Y^1 \otimes Y^2 \in \operatorname{Sep}(A_1A_2 : B_1B_2)$. □

Corollary 5. *Let $\mathcal{N}$ be a quantum channel satisfying $\alpha(\mathcal{N}) = 0$. Then, for every channel $\mathcal{M}$, we have*

$$\alpha(\mathcal{N} \otimes \mathcal{M}) = 0. \tag{5.28}$$

Consequently for $n \in \mathbb{N}$,

$$\alpha(\mathcal{N}^{\otimes n}) = 0. \tag{5.29}$$

Proof. The proof follows by applying Proposition 21 together with the assumption $\alpha(\mathcal{N}) = 0$ and the non-negativity of the quantum Doeblin coefficient. □

Before continuing forward, we show that even qubit channels with classical inputs can be strictly submultiplicative. This differs from the communication value, which is multiplicative for qubit channels [49]. This also differs from the minimal entropy and Holevo capacity (resp. maximal p-norms) which are additive (resp. multiplicative) for all entanglement-breaking channels [58, 59]. Interestingly, the example uses the same sort of construction as given for the PBR theorem [15] and is directly related to the operational task of state exclusion considered there, given the connection of $\alpha(\mathcal{N})$ to state exclusion for classical-quantum channels (see Remark 9).

Example 1. *The qubit-to-qubit classical-quantum channel*

$$\mathcal{N}(X) = \langle 0|X|0\rangle |0\rangle\langle 0| + \langle 1|X|1\rangle |+\rangle\langle +| \tag{5.30}$$

is such that

$$\alpha_I(\mathcal{N} \otimes \mathcal{N}) = \alpha(\mathcal{N} \otimes \mathcal{N}) < \alpha(\mathcal{N})^2 = \alpha_I(\mathcal{N})^2 \ . \tag{5.31}$$

Proof. By Corollary 10, the equality $\alpha_I(\mathcal{N}) = \alpha(\mathcal{N})$ holds for a classical-quantum channel. Note that applying classical-quantum channels in parallel realizes another classical-quantum channel. This justifies the equalities in (5.31). One can verify that the Hermitian operator $X \coloneqq \begin{bmatrix} 0.3964 & 0.25 \\ 0.25 & -0.1036 \end{bmatrix}$ satisfies $X \leq |0\rangle\langle 0|$, $X \leq |+\rangle\langle +|$. Thus, $\alpha(\mathcal{N}) \geq 0.2929$. Using the dual problem (3.3) directly, one can verify that

$$\begin{aligned} Y_{AB} = |00\rangle\langle 00| \otimes \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + |01\rangle\langle 01| \otimes \frac{1}{4}\begin{bmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix} \\ + |10\rangle\langle 10| \otimes \frac{1}{4}\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix} + |11\rangle\langle 11| \otimes \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{5.32}$$

is such that $\mathrm{Tr}_A[Y] = I_B$ and $\mathrm{Tr}[\Gamma^{\mathcal{N}\otimes\mathcal{N}} Y] = 0$, which implies $\alpha(\mathcal{N} \otimes \mathcal{N}) = 0$.[3] □

Remark 5. *We note the above shows that $\alpha(\mathcal{N})$ can be a (trivially) loose bound on $\eta_{\mathrm{Tr}}(\mathcal{N})$ as, for $\mathcal{N}$ as defined in (5.30), $\eta_{\mathrm{Tr}}(\mathcal{N}\otimes\mathcal{N}) < 1$ as all outputs are overlapping, but $\alpha(\mathcal{N}\otimes\mathcal{N}) = 0$.*

[3] Alternatively, one may verify this using a standard SDP solver.

Remark 6 (Multiplicativity of Doeblin)**.** *In Proposition 21, we showed that* $\alpha(\mathcal{N}\otimes\mathcal{M}) \leq \alpha(\mathcal{N})\alpha(\mathcal{M})$*. However, it is not possible to obtain the reverse inequality in general, i.e.,* $\alpha(\mathcal{N}\otimes\mathcal{M}) \not\geq \alpha(\mathcal{N})\alpha(\mathcal{M})$*. We obtain counterexamples that violate the reverse inequality by choosing generalized amplitude damping channels (defined using the Kraus operators in* (7.95) *and* (7.96)*) as* $\mathcal{N}$ *and* $\mathcal{M}$ *with the choice of parameters* $p = 1$ *and* $\eta = 0.1$*.*

We remark that the issue for establishing multiplicativity is that it is not clear that the optimization program in (3.1) is multiplicative because if for $X_1, X_2 \in \text{Herm}$, $\Gamma^{\mathcal{N}} \geq I_{A_1}\otimes X_1$ and $\Gamma^{\mathcal{M}} \geq I_{A_2}\otimes X_2$, it does not follow that $\Gamma^{\mathcal{N}}\otimes\Gamma^{\mathcal{M}} \geq I_{A_1A_2}\otimes X_1\otimes X_2$ as may be verified with real numbers. In other words, the basic concern is that $\alpha(\mathcal{N})$ and $\alpha(\mathcal{M})$ can be truly optimized by Hermitian operators X_1 and X_2, respectively, such that the negative eigenspaces cause an issue when tensored together. For instance, for the generalized amplitude damping channel with the parameters $p = 1$ and $\eta = 0.1$, through numerical evaluation, we find that the optimal $X = \begin{bmatrix} 0.6838 & 0 \\ 0 & -0.2162 \end{bmatrix}$. Then, with that, we have that $I_{A_1A_2}\otimes X\otimes X \not\leq \Gamma^{\mathcal{N}_{p,\eta}}\otimes\Gamma^{\mathcal{N}_{p,\eta}}$.

In contrast, if there exist optimizers of $\alpha(\mathcal{N})$ and $\alpha(\mathcal{M})$, given by X_1 and X_2, respectively, that are positive semidefinite, then multiplicativity does hold, as we will now prove. To that end, let us begin with the following simple lemma.

Lemma 1. *Let* $A, B, C \geq 0$*. If* $A \leq B$*, then* $A\otimes C \leq B\otimes C$*.*

Proof. Consider the map $\Phi(\cdot) = (I\otimes\sqrt{C})\cdot(I\otimes\sqrt{C})^\dagger$. This is CP by the characterization of CP maps, thus $0 \leq \Phi(B-A) = \Phi(B) - \Phi(A) = B\otimes C - A\otimes C$. Re-ordering completes the proof. □

Using this, we obtain the following general result on multiplicativity, which relies on $\alpha_+(\mathcal{N})$ defined in Eq. (3.25).

Lemma 2. *Given a set of channels* $\mathcal{N}_{A_i\to B_i}$ *where* $i \in [n]$*. If* $\alpha_+(\mathcal{N}_i) = \alpha(\mathcal{N}_i)$ *for all* $i \in [n]$*, then*

$$\alpha\left(\bigotimes_{i\in[n]}\mathcal{N}_i\right) = \prod_{i\in[n]}\alpha(\mathcal{N}_i)\ . \tag{5.33}$$

Proof. We first prove the case when $n = 2$ and then explain how the proof generalizes. Recall that Eq. (3.25) is the same as Definition 4 except the optimization is restricted to a positive semidefinite X. Thus, if $\alpha_+(\mathcal{N}_1) = \alpha(\mathcal{N}_1)$ and $\alpha_+(\mathcal{N}_2) = \alpha(\mathcal{N}_2)$, then $\alpha(\mathcal{N}_1), \alpha(\mathcal{N}_2)$ can be optimized by optimizers of $\alpha_+(\mathcal{N}_1), \alpha_+(\mathcal{N}_2)$, respectively, given by $X^1_{B_1}, X^2_{B_2} \geq 0$. Then,

$$(I_{A_1}\otimes X^1_{B_1})\otimes(I_{A_2}\otimes X^2_{B_2}) \leq \Gamma^{\mathcal{N}_1}\otimes(I_{A_2}\otimes X^2_{B_2}) \tag{5.34}$$

$$\leq \Gamma^{\mathcal{N}_1}\otimes\Gamma^{\mathcal{N}_2}\ , \tag{5.35}$$

where the first inequality follows from tensoring $(I_{A_2}\otimes X^2_{B_2})$ to $(I_{A_1}\otimes X^1_{B_1}) \leq \Gamma^{\mathcal{N}_1}$ and the second follows by tensoring $\Gamma^{\mathcal{N}_1}$ to $(I_{A_2}\otimes X^2_{B_2}) \leq \Gamma^{\mathcal{N}_2}$. Both of these inequalities hold via Lemma 1. Combining these relations, we obtain

$$I_{A_1}\otimes I_{A_2}\otimes X^1_{B_1}\otimes X^2_{B_2} \leq \Gamma^{\mathcal{N}_1}\otimes\Gamma^{\mathcal{N}_2}\ . \tag{5.36}$$

Thus, $X_{B_1B_2} \coloneqq X^1 \otimes X^2$ is feasible for $\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$ via (3.1) and achieves $\mathrm{Tr}[X^1 \otimes X^2] = \mathrm{Tr}[X^1]\,\mathrm{Tr}[X^2] = \alpha(\mathcal{N}_1)\alpha(\mathcal{N}_2)$. As (3.1) is a supremization, combining this with Proposition 21 allows us to conclude $\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) = \alpha(\mathcal{N}_1)\alpha(\mathcal{N}_2)$. This completes the proof for the $n = 2$ case. For $n > 2$, note that we can use the argument above iteratively to construct feasible $X_{B_1^n} \coloneqq \otimes_{i\in[n]} X^i_{B_i}$ that achieves $\prod_{i\in[n]} \alpha(\mathcal{N}_i)$. By Proposition 21, this completes the proof. □

Note that as both $\alpha(\mathcal{N})$ and $\alpha_+(\mathcal{N})$ are semidefinite programs, the conditions of Lemma 2 can be checked numerically. However, it is reasonable to look for sufficient structure of the channel to know without numerically checking. We provide two cases. First, as will be shown in Corollary 11, for every measurement channel $\mathcal{M}$, it is the case that $\alpha_+(\mathcal{M}) = \alpha(\mathcal{M})$. Thus we obtain the following:

Theorem 3. *For measurement channels $\mathcal{N}_{A\to X}$ and $\mathcal{M}_{B\to Y}$,*

$$\alpha(\mathcal{N} \otimes \mathcal{M}) = \alpha(\mathcal{N})\alpha(\mathcal{M}). \tag{5.37}$$

We remark this case is important as it recovers that the Doeblin coefficient is multiplicative for arbitrary classical channels [2, Theorem 1]. Also note that Example 1 shows that one cannot make a similar claim for preparation channels.

The second case where we obtain multiplicativity involves channels that are covariant with respect to some irreducible representation on their outputs. We refer to such channels as mictodiactic covariant channels, due to the fact that, for such channels, the output is a maximally mixed state if the input is maximally mixed.[4]

Definition 10. *Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let G be a group. Let $g \to U_g$ and $g \to V_g$ be unitary representations on A and B respectively, where $g \in G$. We say $\mathcal{N}$ is (G, U_g, V_g)-covariant if $V_g\mathcal{N}(X)V_g^\dagger = \mathcal{N}(U_gXU_g^\dagger)$ for all $g \in G$ and $X \in \mathrm{Lin}(A)$.*

Lemma 3. *Let $\mathcal{N}_{A\to B}$ be a (G, U_g, V_g)-covariant channel, where $g \to V_g$ is an irreducible representation of a compact group G. Then*

$$\alpha(\mathcal{N}) = \max\{c : I_A \otimes c\pi_B \le \Gamma^{\mathcal{N}}\} = \alpha_+(\mathcal{N}), \tag{5.38}$$

where π_B is the maximally mixed state of system B.

Proof. Let Y be an optimizer for (3.1). Then, using that the twirling map

$$\mathcal{T}(\cdot) \coloneqq \int_G (\overline{U}_g \otimes V_g) \cdot (\overline{U}_g \otimes V_g)^\dagger \, d\mu(g) \tag{5.39}$$

is CP, we find that

$$\begin{aligned} 0 &\le \mathcal{T}(\Gamma^{\mathcal{N}} - I_A \otimes Y) && (5.40)\\ &= \Gamma^{\mathcal{N}} - \int_G (\overline{U}_g \otimes V_g) I_A \otimes Y (\overline{U}_g \otimes V_g)^\dagger \, d\mu(g) && (5.41)\\ &= \Gamma^{\mathcal{N}} - I_A \otimes \int_G V_g Y V_g^\dagger \, d\mu(g) && (5.42)\\ &= \Gamma^{\mathcal{N}} - I_A \otimes \mathrm{Tr}[Y]\pi_B \ , && (5.43) \end{aligned}$$

[4] The term "mictodiactic" comes from Greek for "mixtures passing through."

where the first equality follows from linearity of the twirling map and covariance of the channel, the second from linearity and a standard property of unitaries, and the third follows from the fact that an irreducible representation of a compact group G forms a one-design. It follows from the last equation that $\mathrm{Tr}[Y]\pi_B$ is feasible. Since Y is optimal, this is also optimal as $\mathrm{Tr}[\mathrm{Tr}[Y]\pi_B] = \mathrm{Tr}[Y]$. Finally, note that $\mathrm{Tr}[Y] \geq 0$ without loss of generality given (3.1). Thus, $\mathrm{Tr}[Y]\pi_B$ is feasible for $\alpha_+(\mathcal{N})$. As $\alpha_+(\mathcal{N}) \leq \alpha(\mathcal{N})$, this shows that $\alpha_+(\mathcal{N}) = \alpha(\mathcal{N})$. Combining these points completes the proof. □

Theorem 4. *For all $i \in [n]$, let $\mathcal{N}_i \in \mathcal{C}(A_i, B_i)$ be (G_i, U_{g_i}, V_{g_i})-covariant where $g_i \to V_{g_i}$ is an irreducible representation of a compact group G_i. Then*

$$\alpha\left(\bigotimes_{i\in[n]} \mathcal{N}_i\right) = \prod_{i\in[n]} \alpha(\mathcal{N}_i). \tag{5.44}$$

That is, multiplicativity holds in this setting.

Proof. This follows from Lemma 3 and Lemma 2. □

5.5 Concavity

Next, we show that the quantum Doeblin coefficients are concave under a convex combination of channels. This may find use in applications where a set of noise channels are applied randomly.

Proposition 22 (Concavity)**.** *Let $\lambda_i \geq 0$ for all i and $\sum_i \lambda_i = 1$, and let $(\mathcal{N}_i)_i$ be a tuple of quantum channels. The quantum Doeblin coefficients $\alpha(\mathcal{N})$ and $\alpha_I(\mathcal{N})$ are concave; i.e., for $\mathcal{N} \coloneqq \sum_i \lambda_i \mathcal{N}_i$, we have*

$$\alpha(\mathcal{N}) \geq \sum_i \lambda_i \alpha(\mathcal{N}_i), \tag{5.45}$$

$$\alpha_I(\mathcal{N}) \geq \sum_i \lambda_i \alpha_I(\mathcal{N}_i). \tag{5.46}$$

Proof. First, we see that $\Gamma^{\mathcal{N}} = \sum_i \lambda_i \Gamma^{\mathcal{N}_i}$. Let $X_{\mathcal{N}_i}$ be chosen such that $I \otimes X_{\mathcal{N}_i} \leq \Gamma^{\mathcal{N}_i}$ for all i. With that, we have

$$\sum_i \lambda_i I \otimes X_{\mathcal{N}_i} \leq \sum_i \lambda_i \Gamma^{\mathcal{N}_i} = \Gamma^{\mathcal{N}}. \tag{5.47}$$

This leads to

$$\alpha(\mathcal{N}) \geq \mathrm{Tr}\left[\sum_i \lambda_i X_{\mathcal{N}_i}\right] = \sum_i \lambda_i \alpha(\mathcal{N}_i). \tag{5.48}$$

Finally, we conclude the proof by optimizing $X_{\mathcal{N}_i}$ on the right-hand side, together with (3.1).

The claim for $\alpha_I(\mathcal{N})$ follows from similar reasoning. □

5.6 Properties of Alternative Notions of Quantum Doeblin Coefficient

Proposition 23. *For a quantum channel $\mathcal{N}$, the following bounds hold:*

$$0 \leq \alpha_I(\mathcal{N}) \leq 1, \tag{5.49}$$

where the upper bound holds with equality if and only if $\mathcal{N}$ is a replacer channel.

Proof. That $\alpha_I(\mathcal{N}) \geq 0$ follows from Proposition 5 and Proposition 18. That $\alpha_I(\mathcal{N}) \leq 1$ follows from Definition 8 and that the classical Doeblin coefficient is bounded from above by 1.

We now focus on establishing the saturation condition. If $\mathcal{N}$ is a replacer channel, it is straightforward to conclude that $\alpha_I(\mathcal{N}) = 1$. The other direction follows from the inequality $\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha_I(\mathcal{N})$ (see (6.34)), and an argument similar to that given in the proof of Proposition 18. Indeed, this inequality and $\alpha_I(\mathcal{N}) = 1$ imply that $\eta_{\mathrm{Tr}}(\mathcal{N}) = 0$. It follows that $\sup_{\rho,\sigma\in\mathcal{D}} \|\mathcal{N}(\rho) - \mathcal{N}(\sigma)\|_1 = 0$. Thus, $\mathcal{N}(\rho) = \mathcal{N}(\sigma)$ for all $\rho, \sigma \in \mathcal{D}$. It follows that $\mathcal{N}$ is a replacer channel. □

Proposition 24 (Normalization)**.** *For a quantum channel $\mathcal{N}$, the following inequalities hold:*

$$0 \leq \alpha_{\mp}(\mathcal{N}) \leq 1. \tag{5.50}$$

Moreover, it is equal to one if and only if $\mathcal{N}$ is a replacer channel.

Proof. To establish the inequality $0 \leq \alpha_{\mp}(\mathcal{N})$, we note that $X_B = 0$ is feasible in the primal SDP in (3.13), and the objective function evaluates to zero for this choice. To establish the inequality $\alpha_{\mp}(\mathcal{N}) \leq 1$, we note that $Y^2_{AB} = \frac{1}{d_A} I_A \otimes I_B$ and $Y^1_{AB} = 0$ are feasible for the dual SDP in (3.18), and the objective function evaluates to one for these choices.

For the equality $\alpha_{\mp}(\mathcal{N}) = 1$, note that it is clear that if $\mathcal{N}$ is a replacer channel, then $\alpha_{\mp}(\mathcal{N}) = 1$ is achievable. Furthermore, if $\alpha_{\mp}(\mathcal{N}) = 1$, then $1 = \alpha_{\mp}(\mathcal{N}) \leq \alpha(\mathcal{N}) \leq 1$ and, by Proposition 18, the channel $\mathcal{N}$ must be a replacer channel. □

Proposition 25 (Multiplicativity of $\alpha_{\mp}(\mathcal{N})$)**.** *For channels $\mathcal{N}$ and $\mathcal{M}$, the following multiplicativity equality holds:*

$$\alpha_{\mp}(\mathcal{N} \otimes \mathcal{M}) = \alpha_{\mp}(\mathcal{N}) \cdot \alpha_{\mp}(\mathcal{M}). \tag{5.51}$$

Proof. To show this equality, we employ proof techniques similar to those presented in [46, Proposition 6].

The inequality $\alpha_{\mp}(\mathcal{N}\otimes\mathcal{M}) \geq \alpha_{\mp}(\mathcal{N})\cdot\alpha_{\mp}(\mathcal{M})$ follows by picking X_1 to be primal feasible for $\alpha_{\mp}(\mathcal{N})$ and X_2 to be primal feasible for $\alpha_{\mp}(\mathcal{M})$ (as given in (3.13)), and noting that $X_1 \otimes X_2$ is feasible for $\alpha_{\mp}(\mathcal{N}\otimes\mathcal{M})$, by [27, Lemma 12.35] and the fact that the Choi operator of $\mathcal{N}\otimes\mathcal{M}$ is $\Gamma^{\mathcal{N}} \otimes \Gamma^{\mathcal{M}}$.

For the opposite inequality, suppose that $Z^1_{A_1B_1}$ and $Z^2_{A_1B_1}$ are dual feasible for $\alpha_{\mp}(\mathcal{N})$ and that $W^1_{A_2B_2}$ and $W^2_{A_2B_2}$ are dual feasible for $\alpha_{\mp}(\mathcal{M})$, as defined in (3.18). Thus, the following constraints hold:

$$Z^1_{A_1B_1}, Z^2_{A_1B_1}, W^1_{A_2B_2}, W^2_{A_2B_2} \geq 0, \tag{5.52}$$

$$\mathrm{Tr}_{A_1}\left[Z^2_{A_1B_1} - Z^1_{A_1B_1}\right] = I_{B_1}, \tag{5.53}$$

$$\mathrm{Tr}_{A_2}\left[W^2_{A_2B_2} - W^1_{A_2B_2}\right] = I_{B_2}. \tag{5.54}$$

Then pick

$$Y^1_{A_1B_1A_2B_2} \coloneqq Z^1_{A_1B_1} \otimes W^1_{A_2B_2} + Z^2_{A_1B_1} \otimes W^2_{A_2B_2}, \tag{5.55}$$
$$Y^2_{A_1B_1A_2B_2} \coloneqq Z^1_{A_1B_1} \otimes W^2_{A_2B_2} + Z^2_{A_1B_1} \otimes W^1_{A_2B_2}. \tag{5.56}$$

It follows that $Y^1_{A_1B_1A_2B_2}$ and $Y^2_{A_1B_1A_2B_2}$ are dual feasible for $\alpha_{\mp}(\mathcal{N} \otimes \mathcal{M})$ because

$$Y^1_{A_1B_1A_2B_2}, Y^2_{A_1B_1A_2B_2} \geq 0, \tag{5.57}$$
$$\mathrm{Tr}_{A_1A_2}\left[Y^2_{A_1B_1A_2B_2} - Y^1_{A_1B_1A_2B_2}\right] = \mathrm{Tr}_{A_1A_2}\left[\left(Z^2_{A_1B_1} - Z^1_{A_1B_1}\right) \otimes \left(W^2_{A_2B_2} - W^1_{A_2B_2}\right)\right] \tag{5.58}$$
$$= I_{B_1} \otimes I_{B_2} \tag{5.59}$$
$$= I_{B_1B_2}. \tag{5.60}$$

Furthermore, the objective function value $\mathrm{Tr}[\Gamma^{\mathcal{N}\otimes\mathcal{M}}(Y^1_{A_1B_1A_2B_2} + Y^2_{A_1B_1A_2B_2})]$ is given by

$$\mathrm{Tr}\left[\Gamma^{\mathcal{N}\otimes\mathcal{M}}\left(Y^1_{A_1B_1A_2B_2} + Y^2_{A_1B_1A_2B_2}\right)\right]$$
$$= \mathrm{Tr}\left[\left(\Gamma^{\mathcal{N}} \otimes \Gamma^{\mathcal{M}}\right)\left(\left(Z^1_{A_1B_1} + Z^2_{A_1B_1}\right) \otimes \left(W^1_{A_2B_2} + W^2_{A_2B_2}\right)\right)\right] \tag{5.61}$$
$$= \mathrm{Tr}\left[\Gamma^{\mathcal{N}}\left(Z^1_{A_1B_1} + Z^2_{A_1B_1}\right) \otimes \Gamma^{\mathcal{M}}\left(W^1_{A_2B_2} + W^2_{A_2B_2}\right)\right] \tag{5.62}$$
$$= \mathrm{Tr}\left[\Gamma^{\mathcal{N}}\left(Z^1_{A_1B_1} + Z^2_{A_1B_1}\right)\right] \mathrm{Tr}\left[\Gamma^{\mathcal{M}}\left(W^1_{A_2B_2} + W^2_{A_2B_2}\right)\right]. \tag{5.63}$$

As such, we conclude that

$$\mathrm{Tr}\left[\Gamma^{\mathcal{N}}\left(Z^1_{A_1B_1} + Z^2_{A_1B_1}\right)\right] \mathrm{Tr}\left[\Gamma^{\mathcal{M}}\left(W^1_{A_2B_2} + W^2_{A_2B_2}\right)\right]$$
$$= \mathrm{Tr}\left[\Gamma^{\mathcal{N}\otimes\mathcal{M}}\left(Y^1_{A_1B_1A_2B_2} + Y^2_{A_1B_1A_2B_2}\right)\right] \tag{5.64}$$
$$\geq \alpha_{\mp}(\mathcal{N} \otimes \mathcal{M}). \tag{5.65}$$

Since the inequality holds for every $Z^1_{A_1B_1}$ and $Z^2_{A_1B_1}$ dual feasible for $\alpha_{\mp}(\mathcal{N})$ and $W^1_{A_2B_2}$ and $W^2_{A_2B_2}$ dual feasible for $\alpha_{\mp}(\mathcal{M})$, we conclude the desired inequality:

$$\alpha_{\mp}(\mathcal{N}) \cdot \alpha_{\mp}(\mathcal{M}) \geq \alpha_{\mp}(\mathcal{N} \otimes \mathcal{M}). \tag{5.66}$$

This concludes the proof. □

Corollary 6. *For channels $\mathcal{N}$ and $\mathcal{M}$, if $\alpha_{\mp}(\mathcal{N}) = \alpha(\mathcal{N})$ and $\alpha_{\mp}(\mathcal{M}) = \alpha(\mathcal{M})$, then the following multiplicativity equality holds*

$$\alpha(\mathcal{N} \otimes \mathcal{M}) = \alpha(\mathcal{N}) \cdot \alpha(\mathcal{M}). \tag{5.67}$$

Proof. Note the following chain of arguments,

$$\alpha(\mathcal{N}) \cdot \alpha(\mathcal{M}) \geq \alpha(\mathcal{N} \otimes \mathcal{M}) \geq \alpha_{\mp}(\mathcal{N} \otimes \mathcal{M}) = \alpha_{\mp}(\mathcal{N}) \cdot \alpha_{\mp}(\mathcal{M}) = \alpha(\mathcal{N}) \cdot \alpha(\mathcal{M}). \tag{5.68}$$

The first inequality follows from Proposition 21, and the first equality follows from Proposition 25. □

Proposition 26 (Data Processing under Pre- and Post-Processing and Isometric Invariance)**.** *Let $\mathcal{E}_{B\to B'}$, $\mathcal{G}_{A'\to A}$, and $\mathcal{N}_{A\to B}$ be quantum channels. Then*

$$\alpha_{\mp}(\mathcal{E}\circ\mathcal{N}\circ\mathcal{G}) \geq \alpha_{\mp}(\mathcal{N}). \tag{5.69}$$

Moreover, the quantity $\alpha_{\mp}(\mathcal{N})$ is invariant under isometric-reversal and isometric pre- and post-processing, respectively; i.e.,

$$\alpha(\mathcal{N}) = \alpha(\mathcal{U}\circ\mathcal{N}\circ\mathcal{P}^{\mathcal{V}}),, \tag{5.70}$$

where $\mathcal{U}$ and $\mathcal{V}$ are isometric channels and $\mathcal{P}^{\mathcal{V}}$ is defined in (5.13).

Proof. Let $X \in \mathrm{Herm}$ be feasible for $\alpha_{\mp}(\mathcal{N})$. Then $-\mathcal{N} \leq \mathcal{R}^X \leq \mathcal{N}$ holds, and this implies that $-\mathcal{E}\circ\mathcal{N}\circ\mathcal{G} \leq \mathcal{E}\circ\mathcal{R}^X\circ\mathcal{G} \leq \mathcal{E}\circ\mathcal{N}\circ\mathcal{G}$ holds. Observe that

$$\mathcal{E}\circ\mathcal{R}^X\circ\mathcal{G} = \mathcal{E}\circ\mathcal{R}^X \tag{5.71}$$
$$= \mathcal{R}^{\mathcal{E}(X)} \tag{5.72}$$
$$\mathrm{Tr}[X] = \mathrm{Tr}[\mathcal{E}(X)], \tag{5.73}$$

where the last equality follows because $\mathcal{E}$ is trace preserving. Then $\mathcal{E}(X)$ is feasible for $\alpha_{\mp}(\mathcal{E}\circ\mathcal{N}\circ\mathcal{G})$, and we conclude that

$$\mathrm{Tr}[X] = \mathrm{Tr}[\mathcal{E}(X)] \leq \alpha_{\mp}(\mathcal{E}\circ\mathcal{N}\circ\mathcal{G}). \tag{5.74}$$

Since the inequality holds for every X feasible for $\alpha(\mathcal{N})$, we conclude that $\alpha_{\mp}(\mathcal{N}) \leq \alpha_{\mp}(\mathcal{E}\circ\mathcal{N}\circ\mathcal{G})$. The statement about isometric invariance follows the same proof given for (5.18). □

Proposition 27 (Concavity)**.** *Let $\lambda_i \geq 0$ for all i and $\sum_i \lambda_i = 1$, and let $(\mathcal{N}_i)_i$ be a tuple of quantum channels. Then $\alpha_{\mp}$ is concave; i.e., for $\mathcal{N} \coloneqq \sum_i \lambda_i\mathcal{N}_i$, we have*

$$\alpha_{\mp}(\mathcal{N}) \geq \sum_i \lambda_i\alpha_{\mp}(\mathcal{N}_i). \tag{5.75}$$

Proof. The proof is similar to the proof of Proposition 22 but incorporates the extra constraint $-\Gamma^{\mathcal{N}} \leq X$. □

We lastly establish that $\alpha_{\mp}$ satisfies the same concatenation property that α and α_I satisfy. This follows the same proof method as for the other quantities, as given in Proposition 19. For this reason, the complete proof is provided in Section C, and the approach is summarized here. First, we establish the following alternative representation of $\alpha_{\mp}$:

Lemma 4. *For a quantum channel $\mathcal{N}$,*

$$\alpha_{\mp}(\mathcal{N}) = \max_{\substack{\lambda\in[0,1],\\ \tau\in\mathrm{aff}(\mathcal{D}),\\ \mathcal{M},\mathcal{G}\in\mathrm{CPTP}}} \left\{\lambda : \lambda\mathcal{R}_\tau + (1-\lambda)\mathcal{M} = \mathcal{N}\,,\, \mathcal{N} = (1+\lambda)\mathcal{G} - \lambda\mathcal{R}_\tau\right\}\,. \tag{5.76}$$

Proof. The proof is provided in Section C. □

This result and its proof method are analogous to (4.24), which was used to establish the robustness interpretation of $\alpha(\mathcal{N})$.

Second, we use optimizers of the above representation for $\alpha_{\mp}(\mathcal{N}_1)$ and $\alpha_{\mp}(\mathcal{N}_2)$ to construct a feasible solution to the above representation such that the concatenation property follows as a consequence.

Proposition 28. *For quantum channels $\mathcal{N}_1$ and $\mathcal{N}_2$ such that $\mathcal{N}_2 \circ \mathcal{N}_1$ is well defined,*

$$1 - \alpha_{\mp}(\mathcal{N}_2 \circ \mathcal{N}_1) \leq (1 - \alpha_{\mp}(\mathcal{N}_2))(1 - \alpha_{\mp}(\mathcal{N}_1)) \ . \tag{5.77}$$

Proof. The proof is provided in Section C. □

6 Contraction and Expansion Coefficients

6.1 Contraction Coefficients and Quantum Doeblin Coefficients

Let $\mathbb{D}(\rho\|\sigma)$ be a generalized divergence, i.e., a function of a pair of quantum states onto the reals that obeys the data-processing inequality for all states ρ and σ and every channel $\mathcal{N}$ [28, 60]:

$$\mathbb{D}(\rho\|\sigma) \geq \mathbb{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{6.1}$$

We call a generalized divergence faithful if $\mathbb{D}(\rho\|\rho) \geq 0$ holds for every quantum state ρ. Contraction coefficients give strong data-processing inequalities for these divergences and hence have numerous applications. There is a large body of literature exploring contraction coefficients for different divergences (see, e.g., [61, 62]). Here, we define a generalized contraction coefficient that allows for a reference system.

Definition 11 (Generalized Contraction Coefficient)**.** *For a quantum channel $\mathcal{N}$, a generalized divergence $\mathbb{D}$, and $r \in \mathbb{N}$, we define*

$$\eta_{\mathbb{D}}(\mathcal{N}, r) \coloneqq \sup_{\substack{\rho_{RA} \neq \sigma_{RA}, \\ \rho_R = \sigma_R, \\ |R| = r}} \frac{\mathbb{D}((\mathrm{id} \otimes \mathcal{N})(\rho_{RA})\|(\mathrm{id} \otimes \mathcal{N})(\sigma_{RA}))}{\mathbb{D}(\rho_{RA}\|\sigma_{RA})}, \tag{6.2}$$

leading to the contraction coefficient and the complete contraction coefficient,

$$\eta_{\mathbb{D}}(\mathcal{N}) \coloneqq \eta_{\mathbb{D}}(\mathcal{N}, 1) = \sup_{\rho_A \neq \sigma_A} \frac{\mathbb{D}(\mathcal{N}(\rho_A)\|\mathcal{N}(\sigma_A))}{\mathbb{D}(\rho_A\|\sigma_A)}, \tag{6.3}$$

$$\eta^c_{\mathbb{D}}(\mathcal{N}) \coloneqq \sup_{r \geq 1} \eta_{\mathbb{D}}(\mathcal{N}, r), \tag{6.4}$$

respectively.

We also define a generalized input-dependent contraction coefficient that allows for a reference system.

Definition 12 (Input-Dependent Generalized Contraction Coefficient)**.** *For a quantum channel $\mathcal{N}_{A\to B}$, a reference state ω_A, a generalized divergence $\mathbb{D}$, and $r \in \mathbb{N}$, we define*

$$\eta_{\mathbb{D}}(\mathcal{N}, \omega, r) \coloneqq \sup_{\substack{\rho_{RA} \neq \sigma_{RA}, \\ \sigma_A = \omega_A, \\ \rho_R = \sigma_R, \\ |R| = r}} \frac{\mathbb{D}((\mathrm{id} \otimes \mathcal{N})(\rho_{RA})\|(\mathrm{id} \otimes \mathcal{N})(\sigma_{RA}))}{\mathbb{D}(\rho_{RA}\|\sigma_{RA})}, \tag{6.5}$$

leading to the input-dependent contraction coefficient and the input-dependent complete contraction coefficient,

$$\eta_{\mathbb{D}}(\mathcal{N},\omega) \coloneqq \eta_{\mathbb{D}}(\mathcal{N},1) = \sup_{\rho_A \neq \omega_A} \frac{\mathbb{D}(\mathcal{N}(\rho_A)\|\mathcal{N}(\omega_A))}{\mathbb{D}(\rho_A\|\omega_A)}, \tag{6.6}$$

$$\eta^c_{\mathbb{D}}(\mathcal{N},\omega) \coloneqq \sup_{r\geq 1} \eta_{\mathbb{D}}(\mathcal{N},\omega,r), \tag{6.7}$$

respectively.

The idea of allowing for additional reference systems was previously investigated in [63, Section 4]. In the following, we provide a number of properties of these coefficients.

Proposition 29 (Properties of Contraction Coefficients)**.** *For quantum channels $\mathcal{N}$ and $\mathcal{M}$, the following hold:*

1. *For faithful $\mathbb{D}$, the contraction coefficients are bounded as follows:*

$$0 \leq \eta_{\mathbb{D}}(\mathcal{N}) \leq \eta^c_{\mathbb{D}}(\mathcal{N}) \leq 1, \tag{6.8}$$

$$0 \leq \eta_{\mathbb{D}}(\mathcal{N},\omega) \leq \eta^c_{\mathbb{D}}(\mathcal{N},\omega) \leq 1. \tag{6.9}$$

2. *Under concatenation, the following inequality holds for all $r \in \mathbb{N}$:*

$$\eta_{\mathbb{D}}(\mathcal{N}\circ\mathcal{M},r) \leq \eta_{\mathbb{D}}(\mathcal{N},r)\,\eta_{\mathbb{D}}(\mathcal{M},r). \tag{6.10}$$

3. *For isometric channels $\mathcal{U}$ and $\mathcal{V}$, the following equality holds for all $r \in \mathbb{N}$:*

$$\eta_{\mathbb{D}}(\mathcal{U}\circ\mathcal{N}\circ\mathcal{P}^{\mathcal{V}},r) = \eta_{\mathbb{D}}(\mathcal{N},r), \tag{6.11}$$

where $\mathcal{P}^{\mathcal{V}}$ is defined in (5.13).

Proof. 1. In (6.8), the first inequality follows from faithfulness of the generalized divergence $\mathbb{D}$, the second by definition, and the third by the data-processing inequality. The same reasoning holds for the chain of inequalities in the input-dependent case.

2. Here, observe that for all ρ_{RA} and σ_{RA} with $\rho_R = \sigma_R$ and $|R| = r$, we have

$$\begin{aligned}
&\frac{\mathbb{D}((\mathrm{id}\otimes(\mathcal{N}\circ\mathcal{M}))(\rho_{RA})\|(\mathrm{id}\otimes(\mathcal{N}\circ\mathcal{M}))(\sigma_{RA}))}{\mathbb{D}(\rho_{RA}\|\sigma_{RA})} \\
&= \frac{\mathbb{D}((\mathrm{id}\otimes(\mathcal{N}\circ\mathcal{M}))(\rho_{RA})\|(\mathrm{id}\otimes(\mathcal{N}\circ\mathcal{M}))(\sigma_{RA}))}{\mathbb{D}((\mathrm{id}\otimes\mathcal{M})(\rho_{RA})\|(\mathrm{id}\otimes\mathcal{M})(\sigma_{RA}))} \frac{\mathbb{D}((\mathrm{id}\otimes\mathcal{M})(\rho_{RA})\|(\mathrm{id}\otimes\mathcal{M})(\sigma_{RA}))}{\mathbb{D}(\rho_{RA}\|\sigma_{RA})}
\end{aligned} \tag{6.12}$$

$$\leq \eta_{\mathbb{D}}(\mathcal{N},r)\,\eta_{\mathbb{D}}(\mathcal{M},r). \tag{6.13}$$

3. As a consequence of 1. and 2., the following inequality holds for every pre- and post-processing channel $\mathcal{E}$ and $\mathcal{G}$:

$$\eta_{\mathbb{D}}(\mathcal{G}\circ\mathcal{N}\circ\mathcal{E},r) \leq \eta_{\mathbb{D}}(\mathcal{N},r). \tag{6.14}$$

The claim then follows from reasoning similar to that given for (5.17)–(5.18). □

Based on the above, the complete trace-distance contraction coefficient of a quantum channel is defined as

$$\eta_{\mathrm{Tr}}^{c}(\mathcal{N}) \coloneqq \sup_{r\geq 1} \sup_{\substack{\rho_{RA},\sigma_{RA} \\ \rho_R=\sigma_R \\ |R|=r}} \frac{T((\mathrm{id}\otimes\mathcal{N})(\rho_{RA}),(\mathrm{id}\otimes\mathcal{N})(\sigma_{RA}))}{T(\rho_{RA},\sigma_{RA})}, \tag{6.15}$$

where $T(\rho,\sigma)$ is defined in (2.5). The non-complete version was previously discussed in [61,64]. One main observation from [11, Eq. (8.86)] and [14, Eq. (III.57)] is that $1-\alpha(\mathcal{N})$ is an upper bound on $\eta_{\mathrm{Tr}}(\mathcal{N})$. Here we prove the stronger statement that it is also an upper bound on the complete contraction coefficient.

Proposition 30. *For $\mathcal{N}$ a quantum channel, the following inequality holds:*

$$\eta_{\mathrm{Tr}}^{c}(\mathcal{N}) \leq 1-\alpha(\mathcal{N}). \tag{6.16}$$

Proof. Via the reverse robustness representation of the Doeblin coefficient (Proposition 12), let $\lambda\in[0,1], \tau\in\mathrm{aff}(\mathcal{D}), \mathcal{M}\in\mathrm{CPTP}$ be optimizers for $\alpha(\mathcal{N})$, such that $\lambda\mathcal{R}^{\tau}+(1-\lambda)\mathcal{M}=\mathcal{N}$. Then,

$$\begin{aligned}
&\|(\mathrm{id}\otimes\mathcal{N})(\rho_{RA})-(\mathrm{id}\otimes\mathcal{N})(\sigma_{RA})\|_1 \\
&= \|(\mathrm{id}\otimes(\lambda\mathcal{R}^{\tau}+(1-\lambda)\mathcal{M}))(\rho_{RA})-(\mathrm{id}\otimes(\lambda\mathcal{R}^{\tau}+(1-\lambda)\mathcal{M}))(\sigma_{RA})\|_1 && (6.17)\\
&= \|\lambda\rho_R\otimes\tau-\lambda\sigma_R\otimes\tau+(1-\lambda)[(\mathrm{id}\otimes\mathcal{M})(\rho_{RA})-(\mathrm{id}\otimes\mathcal{M})(\sigma_{RA})]\|_1 && (6.18)\\
&= (1-\lambda)\,\|(\mathrm{id}\otimes\mathcal{M})(\rho_{RA}-\sigma_{RA})\|_1 && (6.19)\\
&\leq (1-\lambda)\,\|\rho_{RA}-\sigma_{RA}\|_1\,. && (6.20)
\end{aligned}$$

Here, the third equality makes use of the identical marginals on R (i.e., $\rho_R=\sigma_R$), and the inequality follows from the data-processing inequality for the trace distance. Noting that $\lambda=\alpha(\mathcal{N})$ concludes the proof. □

Let us employ the shorthand $\eta_{\gamma}^{c}\equiv\eta_{E_\gamma}^{c}$, when choosing the generalized divergence to be the hockey-stick divergence (see (2.3)). Next, we show that the following upper bound holds.

Proposition 31 (Complete Contraction of Hockey-Stick Divergence)**.** *For $\mathcal{N}$ a quantum channel and $\gamma\geq 1$, the following inequality holds:*

$$\eta_{\gamma}^{c}(\mathcal{N}) \leq 1-\alpha_{+}(\mathcal{N}), \tag{6.21}$$

where $\alpha_+(\mathcal{N})$ is defined in (3.25).

Proof. Recall from [14, Proposition III.3] that

$$\alpha_{+}(\mathcal{N}) = \sup_{\substack{\lambda\in[0,1], \\ \sigma\in\mathcal{D}, \\ \mathcal{M}\in\mathrm{CPTP}}} \{\lambda : \mathcal{N}=\lambda\mathcal{R}^{\sigma}+(1-\lambda)\,\mathcal{M}\}\,. \tag{6.22}$$

Let $\lambda\in[0,1]$, $\sigma\in\mathcal{D}$, and $\mathcal{M}\in\mathrm{CPTP}$ be optimizers for $\alpha_+(\mathcal{N})$, such that $\lambda\mathcal{R}^{\sigma}+(1-\lambda)\mathcal{M}=\mathcal{N}$.

Furthermore, let ρ_{RA} and ω_{RA} be quantum states such that $\rho_R = \omega_R$. Then consider that

$$
\begin{aligned}
&E_\gamma((\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA}) \| (\mathrm{id}_R \otimes \mathcal{N})(\omega_{RA})) \\
&= E_\gamma((\mathrm{id} \otimes (\lambda \mathcal{R}^\sigma + (1-\lambda)\mathcal{M}))(\rho_{RA}) \| (\mathrm{id} \otimes (\lambda \mathcal{R}^\sigma + (1-\lambda)\mathcal{M}))(\omega_{RA})) && (6.23) \\
&= \sup_{0 \leq M_{BR} \leq I_{BR}} \mathrm{Tr}[M_{BR} \left[(\mathrm{id} \otimes (\lambda \mathcal{R}^\sigma + (1-\lambda)\mathcal{M}))(\rho_{RA}) - \gamma (\mathrm{id} \otimes (\lambda \mathcal{R}^\sigma + (1-\lambda)\mathcal{M}))(\omega_{RA})\right]] && (6.24) \\
&= \sup_{0 \leq M_{BR} \leq I_{BR}} \mathrm{Tr}[M_{BR} \left(\lambda \rho_R \otimes \sigma - \gamma \lambda \omega_R \otimes \sigma + (1-\lambda)[(\mathrm{id} \otimes \mathcal{M})(\rho_{RA}) - \gamma (\mathrm{id} \otimes \mathcal{M})(\omega_{RA})]\right)] && (6.25) \\
&\leq \sup_{0 \leq M_{BR} \leq I_{BR}} \mathrm{Tr}[M_{BR} \left((1-\lambda)[(\mathrm{id} \otimes \mathcal{M})(\rho_{RA}) - \gamma (\mathrm{id} \otimes \mathcal{M})(\omega_{RA})]\right)] && (6.26) \\
&= (1-\lambda) E_\gamma((\mathrm{id}_R \otimes \mathcal{M})(\rho_{RA}) \| (\mathrm{id}_R \otimes \mathcal{M})(\omega_{RA})) && (6.27) \\
&\leq (1-\lambda) E_\gamma(\rho_{RA} \| \omega_{RA}). && (6.28)
\end{aligned}
$$

The second equality follows by using (2.4) for the hockey-stick divergence; the first inequality follows by upper bounding the negative term by zero, because $1 - \gamma \leq 0$, $\rho_R = \omega_R$, and $\sigma \otimes \rho_R \geq 0$; and the last inequality by the data-processing inequality for the hockey-stick divergence.

Finally, by supremizing over quantum states ρ_{RA} and ω_{RA} satisfying $\rho_R = \omega_R$, we conclude the proof. □

Let $f : (0, \infty) \to \mathbb{R}$ be a convex and twice differentiable function satisfying $f(1) = 0$. Then, for all quantum states ρ and σ, recall the quantum f-divergence defined in [65, Definition 2.3]:

$$
D_f(\rho \| \sigma) \coloneqq \int_1^\infty f''(\gamma) E_\gamma(\rho \| \sigma) + \gamma^{-3} f''(\gamma^{-1}) E_\gamma(\sigma \| \rho) \, \mathrm{d}\gamma. \tag{6.29}
$$

Corollary 7 (Complete Contraction of f-Divergences)**.** *Let D_f be an f-divergence of the form in (6.29). Then, we have that*

$$
\eta^c_{D_f}(\mathcal{N}) \leq 1 - \alpha_+(\mathcal{N}), \tag{6.30}
$$

where $\alpha_+(\mathcal{N})$ is defined in (3.25).

Proof. Let ρ_{RA} and σ_{RA} be quantum states such that $\rho_R = \sigma_R$. Then, consider that

$$
\begin{aligned}
&D_f((\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA}) \| (\mathrm{id}_R \otimes \mathcal{N})(\sigma_{RA})) \\
&= \int_1^\infty f''(\gamma) E_\gamma((\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA}) \| (\mathrm{id}_R \otimes \mathcal{N})(\sigma_{RA})) \\
&\qquad + \gamma^{-3} f''(\gamma^{-1}) E_\gamma((\mathrm{id}_R \otimes \mathcal{N})(\sigma_{RA}) \| (\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA})) \, \mathrm{d}\gamma && (6.31) \\
&\leq (1 - \alpha_+(\mathcal{N})) \int_1^\infty f''(\gamma) E_\gamma(\rho_{RA} \| \sigma_{RA}) + \gamma^{-3} f''(\gamma^{-1}) E_\gamma(\sigma_{RA} \| \rho_{RA}) \, \mathrm{d}\gamma && (6.32) \\
&= (1 - \alpha_+(\mathcal{N})) \, D_f(\rho_{RA} \| \sigma_{RA}), && (6.33)
\end{aligned}
$$

where the last inequality follows from Proposition 31. With this and supremizing over quantum states ρ_{RA} and σ_{RA} such that $\rho_R = \sigma_R$, we conclude the proof. □

If we do not allow for reference systems, one can in fact deduce from [11, Theorem 8.17] that also

$$\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha_I(\mathcal{N}) \tag{6.34}$$

holds. Bounds on the trace-distance contraction coefficient are particularly interesting because they give upper bounds on the contraction coefficient for a wide class of quantum f-divergences [65, Lemma 4.1]. As $\alpha_I(\mathcal{N})$ is not known to be computationally efficient to evaluate in general, as it requires optimizing over the separable cone, we note the following hierarchy of SDP computable upper bounds that can only improve upon relaxing to considering $\alpha(\mathcal{N})$.

Proposition 32. *Consider quantum channel $\mathcal{N}_{A\to B}$. For every $k \in \mathbb{N}$, define the cone $\mathcal{K}_k(A:B) \coloneqq \mathrm{PPT}(A:B) \cap \mathrm{Sym}_k(A:B)$ where the cones on the right are defined in (2.23) and Definition 2. Then for every $k \in \mathbb{N}$, $\alpha_{I,\mathcal{K}_k}(\mathcal{N})$, as defined via (3.51), is SDP representable and satisfies the inequality*

$$\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha_{I,\mathcal{K}_k}(\mathcal{N}) \ , \tag{6.35}$$

which can only ever be a tighter bound than $\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha(\mathcal{N})$.

Proof. By Corollary 3, $\alpha_I(\mathcal{N}) \geq \alpha_{I,\mathcal{K}_k}(\mathcal{N}) \geq \alpha(\mathcal{N})$. Combining this with (6.34) establishes the inequality and the claim that this bound can only be tighter. That $\alpha_{I,\mathcal{K}_k}(\mathcal{N})$ is SDP representable follows from the constraints that define $\mathrm{PPT}(A:B)$ and $\mathrm{Sym}_k(A:B)$ being SDP representable as is well known. So in this case, (3.45) is SDP representable. □

We remark that $\alpha_{I,\mathcal{K}_k}(\mathcal{N})$ has the appealing property of having an operational interpretation given in Corollary 3.

We define the contraction coefficient for the hockey-stick divergence $E_\gamma(\rho\|\sigma)$ for $\gamma \geq 1$ from (6.3), resulting in the following [66, Eq. (II.4)]:

$$\eta_\gamma(\mathcal{N}) \coloneqq \sup_{\substack{\rho,\sigma\in\mathcal{D}:\\ E_\gamma(\rho,\sigma)\neq 0}} \frac{E_\gamma(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))}{E_\gamma(\rho\|\sigma)} \tag{6.36}$$

$$= \sup_{\psi\perp\phi} E_\gamma(\mathcal{N}(\psi)\|\mathcal{N}(\phi)), \tag{6.37}$$

where the second equality is a consequence of [66, Theorem II.2] and the notation $\psi \perp \phi$ indicates that ψ and ϕ are orthogonal pure states. (See also [67, Proposition 11] for an alternative proof of (6.37).) Note that

$$\eta_1(\mathcal{N}) = \eta_{\mathrm{Tr}}(\mathcal{N}) \tag{6.38}$$

because $E_1(\rho\|\sigma) = T(\rho,\sigma)$.

Here, we provide an alternative proof for a refined connection between $\eta_\gamma(\mathcal{N})$ and $\alpha(\mathcal{N})$, originally considered in [11, Theorem 8.17] and [14, Lemma III.6]. First we give an alternative expression for the hockey-stick divergence.

Lemma 5. *Let ρ be a Hermitian operator with unit trace, and let σ be a positive semidefinite operator. For $\gamma \geq 0$, we have*

$$E_\gamma(\rho\|\sigma) + (1-\gamma)_+ = 1 - \sup_{Y \in \mathrm{Herm}} \{\mathrm{Tr}[Y] : Y \leq \rho, Y \leq \gamma\sigma\}. \tag{6.39}$$

Proof. A similar expression was previously given for the minimum probability of error in a state discrimination problem (see [68, Eq. (III.17)] and [8, Section II]),

$$p_e(p,\rho,\sigma) = \sup_{Y \in \mathrm{Herm}} \{\mathrm{Tr}[Y] : Y \leq p\rho, Y \leq (1-p)\sigma\}. \tag{6.40}$$

Observing that the minimum error probability can be written as a function of the hockey-stick divergence as follows [65, Eq. (2.51)]

$$\begin{aligned} p_e(p,\rho,\sigma) &\coloneqq \inf_{M:0\leq M\leq I} \{p\,\mathrm{Tr}(I-M)\rho + (1-p)\,\mathrm{Tr}\,M\sigma\} && (6.41)\\ &= p - pE_{\frac{1-p}{p}}(\rho\|\sigma), && (6.42) \end{aligned}$$

and fixing $p = \frac{1}{1+\gamma}$, we have

$$\begin{aligned} E_\gamma(\rho\|\sigma) + (1-\gamma)_+ &= 1 - (\gamma+1)p_e(p,\rho,\sigma) && (6.43)\\ &= 1 - (\gamma+1)\sup_{Y \in \mathrm{Herm}} \{\mathrm{Tr}[Y] : Y \leq \frac{1}{1+\gamma}\rho, Y \leq \frac{\gamma}{1+\gamma}\sigma\} && (6.44)\\ &= 1 - \sup_{Y \in \mathrm{Herm}} \{\mathrm{Tr}[Y] : Y \leq \rho, Y \leq \gamma\sigma\}, && (6.45) \end{aligned}$$

concluding the proof. □

Corollary 8 below provides an alternative formulation of the hockey-stick contraction coefficient $\eta_\gamma(\mathcal{N})$.

Corollary 8 (Contraction Coefficient of Hockey-Stick Divergence)**.** *Let $\mathcal{N}$ be a quantum channel. The following equality holds for $\gamma \geq 1$:*

$$\eta_\gamma(\mathcal{N}) = 1 - \inf_{\psi \perp \phi} \sup_{Y \in \mathrm{Herm}} \{\mathrm{Tr}[Y] : Y \leq \mathcal{N}(\psi), Y \leq \gamma\mathcal{N}(\phi)\}, \tag{6.46}$$

where $\eta_\gamma(\mathcal{N})$ is defined in (6.36).

Proof. The proof of (6.46) follows by adapting Lemma 5 together with (6.37). □

Remark 7 (Efficient Computability of $\alpha(\mathcal{N}^{\otimes n})$)**.** *Due to*

$$\eta_{\mathrm{Tr}}(\mathcal{N}^{\otimes n}) \leq 1 - \alpha(\mathcal{N}^{\otimes n}), \tag{6.47}$$

the computation of $\alpha(\mathcal{N}^{\otimes n})$ is useful in providing bounds on $\eta_{\mathrm{Tr}}(\mathcal{N}^{\otimes n})$. Recall from the dual SDP in (3.3) that

$$\alpha(\mathcal{N}^{\otimes n}) = \inf_{Y_{A^nB^n} \geq 0} \left\{ \mathrm{Tr}\left[(\Gamma^{\mathcal{N}}_{AB})^{\otimes n} Y_{A^nB^n} \right] \,:\, \mathrm{Tr}_{A^n}[Y_{A^nB^n}] = I_{B^n} \right\}. \tag{6.48}$$

By observing the SDP as it is, it seems that the time complexity to evaluate this SDP would be $O(\exp(n))$. However, one can utilize permutation symmetry to reduce its complexity to $O(\mathrm{poly}(n))$, by following the approach demonstrated in [69]. In particular, we can establish the claimed efficiency with the use of properties in [70, Proposition 3].

Corollary 9. *Let $\mathcal{N}$ be a quantum channel. The following inequality holds for all $n \in \mathbb{N}$:*

$$\eta_{\mathrm{Tr}}(\mathcal{N}^{\otimes n}) \leq 1 - \alpha_{\mp}(\mathcal{N})^n \leq n(1 - \alpha_{\mp}(\mathcal{N})). \tag{6.49}$$

Proof. The first inequality follows from (6.34), Proposition 5, (3.17), and multiplicativity of $\alpha_{\mp}(\mathcal{N})$ (Proposition 25). The second inequality follows because $1 - x^n \leq n(1-x)$ for all $n \in \mathbb{N}$ and $x \in [0,1]$. □

6.2 Expansion Coefficients and Reverse Doeblin Coefficients

So far, we considered contraction coefficients and Doeblin coefficients. Next, we focus on expansion coefficients and reverse Doeblin coefficients. Expansion coefficients were mentioned in [71] and then further explored in [14, 72, 73]. We give a general definition here that allows for arbitrary divergences and additional reference systems.

Definition 13. *For a quantum channel $\mathcal{N}$, a generalized divergence $\mathbb{D}$, and $r \in \mathbb{N}$, we define*

$$\check{\eta}_{\mathbb{D}}(\mathcal{N}, r) \coloneqq \inf_{\substack{\rho_{RA} \neq \sigma_{RA}, \\ \rho_R = \sigma_R, \\ |R| = r}} \frac{\mathbb{D}((\mathrm{id} \otimes \mathcal{N})(\rho_{RA}) \| (\mathrm{id} \otimes \mathcal{N})(\sigma_{RA}))}{\mathbb{D}(\rho_{RA} \| \sigma_{RA})}, \tag{6.50}$$

leading to the expansion coefficient and the complete expansion coefficient,

$$\check{\eta}_{\mathbb{D}}(\mathcal{N}) \coloneqq \check{\eta}_{\mathbb{D}}(\mathcal{N}, 1) = \inf_{\rho_A \neq \sigma_A} \frac{\mathbb{D}(\mathcal{N}(\rho_A) \| \mathcal{N}(\sigma_A))}{\mathbb{D}(\rho_A \| \sigma_A)} \tag{6.51}$$

$$\check{\eta}^{c}_{\mathbb{D}}(\mathcal{N}) \coloneqq \inf_{r \geq 1} \check{\eta}_{\mathbb{D}}(\mathcal{N}, r), \tag{6.52}$$

respectively.

It should be remarked that, in this generality, expansion coefficients are often trivial, i.e., equal to zero [14, 72, 73]. A notable exception is the case of the trace distance. It was shown in [14, Lemma IV.1] that the expansion coefficient of the trace distance is lower bounded as follows by the reverse Doeblin coefficient $\check{\alpha}(\mathcal{N})$:

$$\check{\eta}_{\mathrm{Tr}}(\mathcal{N}) \geq 1 - \check{\alpha}(\mathcal{N}), \tag{6.53}$$

where

$$\check{\alpha}(\mathcal{N}) \coloneqq \inf_{X \in \mathrm{Herm}} \{p : \mathcal{N} \geq_{\mathrm{deg}} \mathcal{Q}_{p,X}\} \tag{6.54}$$

with

$$\mathcal{Q}_{p,X}(\rho) \coloneqq (1-p)\rho + p \,\mathrm{Tr}[\rho] X \tag{6.55}$$

for $X \in \mathrm{Herm}$ and for all ρ. The notation $\mathcal{N} \geq_{\mathrm{deg}} \mathcal{M}$ is equivalent to the existence of a channel $\mathcal{A}$ such that $\mathcal{A} \circ \mathcal{N} = \mathcal{M}$. Also, note that (6.54) is SDP computable by [14, Corollary IV.2]. Here, we state a stronger result that includes the complete expansion coefficient for the trace distance.

Proposition 33. *Let $\mathcal{N}$ be a quantum channel. Then, we have*

$$\check{\eta}_{\mathrm{Tr}}(\mathcal{N}) \geq \check{\eta}^{c}_{\mathrm{Tr}}(\mathcal{N}) \geq 1 - \check{\alpha}(\mathcal{N}). \tag{6.56}$$

Proof. The first inequality holds by definition. Now, let p, $\mathcal{A}$, and X be optimizers for $\check{\alpha}(\mathcal{N})$. Consider the following:

$$\begin{aligned}
&\|(\mathrm{id} \otimes \mathcal{N})(\rho_{RA}) - (\mathrm{id} \otimes \mathcal{N})(\sigma_{RA})\|_1 \\
&\geq \|(\mathrm{id} \otimes \mathcal{A} \circ \mathcal{N})(\rho_{RA}) - (\mathrm{id} \otimes \mathcal{A} \circ \mathcal{N})(\sigma_{RA})\|_1 && (6.57)\\
&= \|(\mathrm{id} \otimes \mathcal{Q}_{p,X})(\rho_{RA}) - (\mathrm{id} \otimes \mathcal{Q}_{p,X})(\sigma_{RA})\|_1 && (6.58)\\
&= (1-p)\,\|\rho_{RA} - \sigma_{RA}\|_1\,. && (6.59)
\end{aligned}$$

The second inequality follows directly from the fact that we have shown that

$$\frac{\|(\mathrm{id} \otimes \mathcal{N})(\rho_{RA}) - (\mathrm{id} \otimes \mathcal{N})(\sigma_{RA})\|_1}{\|\rho_{RA} - \sigma_{RA}\|_1} \geq 1 - p \tag{6.60}$$

holds for all states ρ_{RA} and σ_{RA} satisfying $\rho_R = \sigma_R$. □

Although not the main focus of this work, expansion coefficients and reverse Doeblin coefficient will play a role again later when we discuss applications.

7 Quantum Doeblin Coefficients for Example Channels

In this section, we consider quantum Doeblin coefficients for a variety of quantum channels, in some cases finding exact expressions.

7.1 Classical–Quantum Channels

First, we consider classical–quantum channels of the form

$$\mathcal{N}(\cdot) = \sum_i \langle i|(\cdot)|i\rangle \rho_i, \tag{7.1}$$

where $(\rho_i)_i$ is a tuple of quantum states and $\{|i\rangle\}_i$ is an orthonormal basis. For channels of this form, we obtain the following characterization of the Doeblin coefficient $\alpha(\mathcal{N})$, which demonstrates that, for such channels, it is equal to a quantity previously considered in [9, Eq. (185)], in the context of error exponents for quantum state exclusion.

Proposition 34. *For a classical–quantum channel $\mathcal{N}$ of the form in (7.1), the following equality holds:*

$$\alpha(\mathcal{N}) = \sup_{X \in \mathrm{Herm}} \{\mathrm{Tr}\, X : X \leq \rho_i \ \forall i\}\,. \tag{7.2}$$

Proof. For such channels, the Choi operator is given by

$$\Gamma^{\mathcal{N}} = \sum_{i,j} |i\rangle\langle j| \otimes \mathcal{N}(|i\rangle\langle j|) = \sum_i |i\rangle\langle i| \otimes \rho_i. \tag{7.3}$$

Hence,

$$\alpha(\mathcal{N}) = \sup_{X \in \mathrm{Herm}} \left\{\mathrm{Tr}\, X : I \otimes X \leq \Gamma^{\mathcal{N}}\right\} \tag{7.4}$$

$$= \sup_{X \in \mathrm{Herm}} \left\{ \mathrm{Tr}\, X : I \otimes X \leq \sum_i |i\rangle\langle i| \otimes \rho_i \right\} \tag{7.5}$$

$$= \sup_{X \in \mathrm{Herm}} \left\{ \mathrm{Tr}\, X : X \leq \rho_i \ \forall i \right\}, \tag{7.6}$$

concluding the proof. □

Remark 8 (Connection to Trace Distance and its Multivariate Generalizations)**.** *Comparing Proposition 34 with (6.39) for $\gamma = 1$ and applying Lemma 5, we notice that for a channel $\mathcal{N}$ as in (7.1) and associated with a pair (ρ, σ) of states, we have*

$$E_1(\rho\|\sigma) = \frac{1}{2} \|\rho - \sigma\|_1 = 1 - \alpha(\mathcal{N}). \tag{7.7}$$

For classical–quantum channels with more than two states (i.e., such that the dimension of the input system is larger than two), we can view the quantum Doeblin coefficient as a multivariate trace distance generalization. This generalizes classical results from [2].

Remark 9 (State Discrimination and Exclusion)**.** *We can relate Proposition 34 and (7.7) to the probability of error in state discrimination and state exclusion. Consider an equal prior $p = \frac{1}{2}$ setting. Then the probability of error evaluates to*

$$p_e(p, \rho, \sigma) = \frac{1}{2}\alpha(\mathcal{N}) = \frac{1}{2}(1 - E_1(\rho\|\sigma)). \tag{7.8}$$

Analogous considerations hold for non-equal priors and state exclusion between multiple states; see [9, Eq. (181)].

Corollary 10. *For a classical-quantum channel $\mathcal{N}$, the following equality holds:*

$$\alpha_I(\mathcal{N}) = \alpha(\mathcal{N}). \tag{7.9}$$

Proof. As Proposition 5 shows that $\alpha_I(\mathcal{N}) \geq \alpha(\mathcal{N})$, it suffices to prove the other inequality. Let X be an optimizer for $\alpha_I(\mathcal{N})$. We will show that it is feasible for $\alpha(\mathcal{N})$. By Proposition 7 and (2.22), we have that

$$\begin{aligned} & \langle\alpha|\langle\beta|(\Gamma^{\mathcal{N}} - I \otimes X)|\alpha\rangle|\beta\rangle \geq 0 \qquad \forall |\alpha\rangle|\beta\rangle \\ \Leftrightarrow \quad & \sum_i |\langle\alpha|i\rangle|^2 \langle\beta|\rho_i|\beta\rangle - \langle\beta|X|\beta\rangle \geq 0 \qquad \forall |\alpha\rangle|\beta\rangle \ , \end{aligned} \tag{7.10}$$

where the equivalence uses the Choi representation of the classical-quantum channel. By considering the case $|\alpha\rangle = |i'\rangle$, we find that $\langle\beta|\rho_{i'}|\beta\rangle \geq \langle\beta|X|\beta\rangle$ for all $|\beta\rangle$. This is the condition for positive semidefiniteness, i.e., $\rho_{i'} \geq X$. As this holds for each i', we have that $\rho_i \geq X$ for all i. By Proposition 34, this proves that X is feasible for $\alpha(\mathcal{N})$. As $\alpha(\mathcal{N})$ involves a maximization, this completes the proof. □

7.2 Measurement Channels

A measurement channel is defined as follows for a measurement comprised of POVM elements $(\Lambda_y)_{y \in \mathcal{Y}}$:

$$\mathcal{M}(\rho) := \sum_{y \in \mathcal{Y}} \mathrm{Tr}[\Lambda_y \rho] |y\rangle\langle y|, \tag{7.11}$$

where $\{|y\rangle\}_{y \in \mathcal{Y}}$ is an orthonormal basis.

Proposition 35. *The following formula holds for the Doeblin coefficient of a measurement channel $\mathcal{M}$, as defined in (7.11):*

$$\alpha(\mathcal{M}) = \sum_{y \in \mathcal{Y}} \lambda_{\min}(\Lambda_y). \tag{7.12}$$

Proof. One can check that the Choi operator of a measurement channel is as follows:

$$\Gamma^{\mathcal{M}} = \sum_{y \in \mathcal{Y}} \Lambda_y^T \otimes |y\rangle\langle y|. \tag{7.13}$$

Thus, the Doeblin coefficient for such channels is as follows:

$$\alpha(\mathcal{M}) = \sup_{X \in \mathrm{Herm}} \left\{ \mathrm{Tr}[X] : I \otimes X \leq \sum_{y \in \mathcal{Y}} \Lambda_y^T \otimes |y\rangle\langle y| \right\}. \tag{7.14}$$

Defining $\Delta(\cdot) \coloneqq \sum_{y \in \mathcal{Y}} |y\rangle\langle y|(\cdot)|y\rangle\langle y|$ and given that $\mathrm{Tr}[X] = \mathrm{Tr}[\Delta(X)]$ and

$$I \otimes X \leq \sum_{y \in \mathcal{Y}} \Lambda_y^T \otimes |y\rangle\langle y| \quad \implies \quad I \otimes \Delta(X) \leq \sum_{y \in \mathcal{Y}} \Lambda_y^T \otimes \Delta(|y\rangle\langle y|) = \sum_{y \in \mathcal{Y}} \Lambda_y^T \otimes |y\rangle\langle y|, \tag{7.15}$$

it suffices to optimize over diagonal X in (7.14). Then the SDP reduces to

$$\begin{aligned}
\alpha(\mathcal{M}) &= \sup_{(\mu_y \in \mathbb{R})_{y \in \mathcal{Y}}} \left\{ \sum_y \mu_y : I \otimes \sum_{y \in \mathcal{Y}} \mu_y |y\rangle\langle y| \leq \sum_{y \in \mathcal{Y}} \Lambda_y^T \otimes |y\rangle\langle y| \right\} && (7.16) \\
&= \sup_{(\mu_y \in \mathbb{R})_{y \in \mathcal{Y}}} \left\{ \sum_y \mu_y : \mu_y I \leq \Lambda_y^T \quad \forall y \in \mathcal{Y} \right\} && (7.17) \\
&= \sum_{y \in \mathcal{Y}} \lambda_{\min}(\Lambda_y^T) && (7.18) \\
&= \sum_{y \in \mathcal{Y}} \lambda_{\min}(\Lambda_y), && (7.19)
\end{aligned}$$

in analogy with the classical Doeblin coefficient, as expressed in (1.3). □

The following shows that all of the considered quantum Doeblin coefficients are equal for measurement channels.

Corollary 11. *For a measurement channel $\mathcal{M}$,*

$$\alpha_+(\mathcal{M}) = \alpha_{\mp}(\mathcal{M}) = \alpha(\mathcal{M}) = \alpha_I(\mathcal{M}). \tag{7.20}$$

Proof. In general, $\alpha_+(\mathcal{M}) \leq \alpha_{\mp}(\mathcal{M}) \leq \alpha(\mathcal{M}) \leq \alpha_I(\mathcal{M})$. The equality $\alpha_+(\mathcal{M}) = \alpha(\mathcal{M})$ follows from Proposition 35, showing that an optimizer X of $\alpha(\mathcal{M})$ is positive semidefinite without loss of generality. Therefore all that remains is to show that $\alpha(\mathcal{M}) \geq \alpha_I(\mathcal{M})$. To that end, let X be an optimizer of $\alpha_I(\mathcal{M})$. By (3.44) and using $\Gamma^{\mathcal{M}} = \sum_y \Lambda_y^T \otimes |y\rangle\langle y|$, consider that

$$\langle\alpha|\langle\beta|(\Gamma^{\mathcal{M}} - I \otimes X)|\alpha\rangle|\beta\rangle \geq 0 \,\forall |\alpha\rangle|\beta\rangle \Leftrightarrow \sum_y \langle\alpha|\Lambda_y^T|\alpha\rangle|\langle\beta|y\rangle|^2 - \langle\beta|X|\beta\rangle \geq 0 \,\forall |\alpha\rangle|\beta\rangle \ . \tag{7.21}$$

As the above holds for an arbitrary tensor-product pure state $|\alpha\rangle|\beta\rangle$, by making the choice $|\beta\rangle = |y'\rangle$ and $|\alpha\rangle$ the minimum eigenvector of $\Lambda_{y'}^T$, we obtain $\lambda_{\min}(\Lambda_{y'}^T) \geq \langle y'|X|y'\rangle$ for all $y' \in \mathcal{Y}$. Thus, using that the trace may be written as the sum over the diagonal in any basis, we have

$$\alpha_I(\mathcal{M}) = \mathrm{Tr}[X] = \sum_{y'} \langle y'|X|y'\rangle \leq \sum_{y'} \lambda_{\min}(\Lambda_{y'}^T) = \alpha(\mathcal{M}) \ , \tag{7.22}$$

where the last equality follows from Proposition 35. □

7.3 Entanglement-Breaking Channels

An entanglement-breaking channel has the following form [56]:

$$\mathcal{E}(\rho) \coloneqq \sum_x \mathrm{Tr}[\Lambda_x \rho]\sigma_x, \tag{7.23}$$

where $(\Lambda_x)_x$ is a POVM and $(\sigma_x)_x$ is a tuple of states.

Proposition 36. *Let $\mathcal{N}$ be an entanglement-breaking channel. Then for every channel $\mathcal{M}$, we have that*

$$\alpha_{I,+}(\mathcal{N} \otimes \mathcal{M}) \leq \alpha_{I,+}(\mathcal{N})\alpha_{I,+}(\mathcal{M}). \tag{7.24}$$

Proof. Let $\mathcal{N} : \mathcal{L}(A) \to \mathcal{L}(B)$ be an entanglement-breaking channel. Then,

$$\mathcal{N}(\rho_A) = \sum_x \mathrm{Tr}[\Lambda_x \rho_A]\sigma_x. \tag{7.25}$$

Also, $\mathcal{M} : \mathcal{L}(A') \to \mathcal{L}(B')$. Considering $\alpha_{I,+}(\mathcal{N} \otimes \mathcal{M})$ in Eq. (3.26), a feasible solution to the primal $X_{\mathcal{N}\otimes\mathcal{M}}$ needs to satisfy the following for every state $\rho_{AA'}$

$$X_{\mathcal{N}\otimes\mathcal{M}} \leq \sum_x \mathrm{Tr}[\Lambda_x \rho_A]\sigma_x \otimes \mathcal{M}(\omega_x), \tag{7.26}$$

where

$$\omega_x \coloneqq \frac{\mathrm{Tr}_A[(\Lambda_x \otimes I_{A'})\rho_{AA'}]}{\mathrm{Tr}[\Lambda_x \rho_A]} \tag{7.27}$$

Choose an optimizer $X_{\mathcal{N}} \geq 0$ of the primal of $\alpha_{I,+}(\mathcal{N})$ and $X_{\mathcal{M}} \geq 0$ for $\alpha_{I,+}(\mathcal{M})$. Thus, they satisfy, for all states ρ_A and $\omega_{A'}$,

$$X_{\mathcal{N}} \leq \mathcal{N}(\rho_A), \qquad X_{\mathcal{M}} \leq \mathcal{M}(\omega_{A'}). \tag{7.28}$$

Then, consider that

$$\sum_x \mathrm{Tr}[\Lambda_x \rho_A]\sigma_x \otimes \mathcal{M}(\omega_x) \geq \sum_x \mathrm{Tr}[\Lambda_x \rho_A]\sigma_x \otimes X_{\mathcal{M}} \tag{7.29}$$

$$\geq X_{\mathcal{N}} \otimes X_{\mathcal{M}}. \tag{7.30}$$

With this, we have that $X_{\mathcal{N}} \otimes X_{\mathcal{M}} \leq (\mathcal{N} \otimes \mathcal{M})(\rho_{AA'})$ for all $\rho_{AA'}$, leading to $\mathrm{Tr}[X_{\mathcal{N}} \otimes X_{\mathcal{M}}] \leq \alpha_{I,+}(\mathcal{N} \otimes \mathcal{M})$, so as to arrive at

$$\alpha_{I,+}(\mathcal{N})\alpha_{I,+}(\mathcal{M}) \leq \alpha_{I,+}(\mathcal{N} \otimes \mathcal{M}), \tag{7.31}$$

by recalling that $X_{\mathcal{N}}$ and $X_{\mathcal{M}}$ are optimizers for the primal of $\alpha_{I,+}(\mathcal{N})$ and $\alpha_{I,+}(\mathcal{M})$. □

7.4 Generalized Dephasing Channels

Define the generalized qudit dephasing channel $\mathcal{D}$ as follows:

$$\mathcal{D}(\rho) \coloneqq \sum_{i=0}^{d-1} \langle i|\rho|j\rangle |i\rangle\langle j| \langle \psi_j|\psi_i\rangle, \tag{7.32}$$

where $(|\psi_i\rangle)_i$ is a tuple of pure quantum states. An isometric extension of this channel is specified by $|i\rangle \to |i\rangle|\psi_i\rangle$. Then the Choi operator of the generalized dephasing channel is as follows:

$$\Gamma^{\mathcal{D}} = \sum_{i,j=1}^{d} \langle \psi_j|\psi_i\rangle |i\rangle\langle j| \otimes |i\rangle\langle j|. \tag{7.33}$$

Proposition 37 (Doeblin Coefficients for the Generalized Dephasing Channel)**.** *For a generalized dephasing channel $\mathcal{D}$, the following equalities hold:*

$$\alpha(\mathcal{D}) = 0, \tag{7.34}$$

$$\alpha_I(\mathcal{D}) = 0. \tag{7.35}$$

Proof. The equality $\eta_{\mathrm{Tr}}(\mathcal{D}) = 1$ holds because

$$\|\mathcal{D}(|0\rangle\langle 0|) - \mathcal{D}(|1\rangle\langle 1|)\|_1 = \||0\rangle\langle 0| - |1\rangle\langle 1|\|_1. \tag{7.36}$$

As $\eta_{\mathrm{Tr}}(\mathcal{D}) \leq 1-\alpha(\mathcal{D})$ (see (6.8) and Proposition 30), we conclude that $\alpha(\mathcal{D}) = 0$. Furthermore, by (6.34), we conclude that $\alpha_I(\mathcal{D}) = 0$. □

7.5 Qubit Channels

In this section, our main purpose is to analyze Doeblin coefficients and contraction and expansion coefficients for the trace distance for qubit channels. Before doing so, let us first analyse conditions such that the Doeblin coefficient is non-trivial (i.e., $\alpha(\mathcal{N}) > 0$) for general channels. Then we study the conditions such that $\alpha(\mathcal{N}) > 0$ for qubit channels.

We begin by noting whenever the Choi matrix is full rank, the quantum Doeblin coefficient $\alpha(\mathcal{N})$ is strictly greater than zero; i.e., this guarantees that the channel is contractive.

Proposition 38. *Let $\mathcal{N}\colon \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a quantum channel. If $\Gamma^{\mathcal{N}} > 0$, then $\alpha(\mathcal{N}) \geq d_B \lambda_{\min}(\Gamma^{\mathcal{N}}) > 0$.*

Proof. If $\Gamma^{\mathcal{N}} > 0$, then $t \coloneqq \lambda_{\min}(\Gamma^{\mathcal{N}}) > 0$. Then letting $X = tI_B$, we have for all $|\omega\rangle \in A \otimes B$, $0 < \langle \omega|I_A \otimes X|\omega\rangle = t = \lambda_{\min}(\Gamma^{\mathcal{N}}) \leq \langle \omega|\Gamma^{\mathcal{N}}|\omega\rangle$. Thus, X is feasible for the optimization in (3.1) and has $\mathrm{Tr}[X] = d_B \lambda_{\min}(\Gamma^{\mathcal{N}})$. As (3.1) involves a supremization, this completes the proof. □

One may take being full rank as a qualitative sign that the channel disturbs a quantum state strongly. A natural question is if, with more structure, more can be said about when the quantum Doeblin coefficient is non-trivial. Indeed, we are able to establish that for qubit channels, the quantum Doeblin coefficient is 'faithful' in the sense that it is non-zero if and only if the trace-distance contraction coefficient is not equal to one.

Theorem 5. *For a qubit channel $\mathcal{N}$, the following are equivalent:*

- $\alpha(\mathcal{N}) > 0$,
- $\eta_{\mathrm{Tr}}(\mathcal{N}) < 1$.

To establish Theorem 5, we will rely on a variety of parameterizations of qubit channels, which are reviewed in Section D for clarity. Of particular import will be the Stokes parameterization, which tells us that any qubit channel can be expressed as an affine transformation on the Bloch sphere according to $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$. The proof of Theorem 5 breaks into two steps. First, we find necessary and sufficient conditions for the quantum Doeblin coefficient to be strictly non-zero in terms of the Stokes parameterization of the channel (Lemma 6). Second, we establish necessary and sufficient conditions for the trace distance contraction coefficient to be strictly less than one in terms of the Stokes parameterization, by slightly strengthening a result of [61] (Lemma 7). Lastly, we establish these two claims in fact select for the same conditions on the Stokes parameters, which will complete the proof of Theorem 5. A key idea in both lemmas is to realize that the 'non-unital part' of a qubit channel is irrelevant for determining the value.

Lemma 6 below establishes necessary and sufficient conditions for $\alpha(\mathcal{N}) > 0$. While in effect purely mathematical, we can give a physical explanation for what the proof shows. The proof is made up of two reductions. The first reduction shows that only the 'unital part' of the qubit channel $\mathcal{N}$ is relevant for computing $\alpha(\mathcal{N})$. What we mean by this is that, under some pre- and post-processing, the linear shift of the Paulis that a qubit channel can realize does not affect the value of the contraction coefficient. The second reduction in effect shows that the 'unital part' of the channel is contractive if and only if it is a non-trivial mixture of all four Pauli channels, i.e., $\Phi(\cdot) = p_0 \,\mathrm{id}(\cdot) + p_1 X \cdot X^\dagger + p_2 Y \cdot Y^\dagger + p_3 Z \cdot Z^\dagger$ where $p_0, p_1, p_2, p_3 > 0$. We stress that this is not the same as saying that a unital channel is contractive if and only if it is a Pauli channel with all four Paulis, because the first step in the reduction involves pre- and post-processing to convert the channel to have a specific form in the Stokes representation.

Lemma 6. *Let $\mathcal{N}$ be a qubit channel given by the transformation $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$. Let $\mathcal{N}'$ be the unitary pre- and post-processed form of $\mathcal{N}$ such that*

$$\Gamma^{\mathcal{N}'} = \frac{1}{2}\left[I_A \otimes I_B + I_A \otimes \vec{r}' \cdot \vec{\sigma} + \sum_k t'_k \sigma_k \otimes \sigma_k \right] . \tag{7.37}$$

Then, $\alpha(\mathcal{N}) > 0$ if and only if $\vec{t}'$ is a full support convex combination of the set

$$\{(1,1,1), (1,-1,-1), (-1,1,-1), (-1,-1,1)\} . \tag{7.38}$$

Proof. By Proposition 20, without loss of generality, we can focus on any version of $\mathcal{N}$ that has undergone unitary pre- and post-processing. We thus choose the unitary pre- and post-processing from Proposition 55 so that, without loss of generality, we just need to consider $\alpha(\mathcal{N}')$ where

$$\Gamma^{\mathcal{N}'} = \frac{1}{2}\left[I_A \otimes I_B + I_A \otimes \vec{r}' \cdot \vec{\sigma} + \sum_k t'_k \sigma_k \otimes \sigma_k \right] . \tag{7.39}$$

Now, note that by (3.3) any feasible Y is the Choi operator of a unital CP map by the Choi isomorphism. Thus, by (D.5), we may express Y in the form

$$Y = \frac{1}{2}\left[I_A \otimes I_B + \vec{s}\cdot\vec{\sigma}\otimes I_B + \sum_{i,j} b_{i,j}\sigma_i \otimes \sigma_j\right] \tag{7.40}$$

and the positivity constraints are $|B_{11} \pm B_{22}| \leq |1 + B_{33}|$ by Proposition 54. Now,

$$\begin{aligned}\mathrm{Tr}[\Gamma^{\mathcal{N}'}Y] = \frac{1}{4}\,\mathrm{Tr}\Bigg[&\left(I_A \otimes I_B + I_A \otimes \vec{r}\,'\cdot\vec{\sigma} + \sum_k t'_k \sigma_k \otimes \sigma_k\right) \\ &\cdot\left(I_A \otimes I_B + \vec{s}\cdot\vec{\sigma}\otimes I_B + \sum_{i,j} b_{i,j}\sigma_i\otimes\sigma_j\right)\Bigg] \qquad (7.41)\\ = \frac{1}{4}&\left(4 + 4\sum_k t'_k b_{k,k}\right)\ , \qquad (7.42)\end{aligned}$$

where the second equality uses the fact that Pauli operators are traceless and the commutation relations between Pauli operators to simplify. Thus, by (3.3) and the positivity constraints being $|B_{11} \pm B_{22}| \leq |1 + B_{33}|$, we find that

$$\alpha(\mathcal{N}') = 1 + \inf\left\{\sum_k t'_k b_{k,k} : |b_{1,1} \pm b_{2,2}| \leq |1 \pm b_{3,3}|\right\}\ . \tag{7.43}$$

Note this shows that only the unital part of $\mathcal{N}'$ and Y (if we consider it also as a Choi operator) matter. In other words, we have reduced the problem to showing α is faithful on qubit channels to showing it is faithful on the equivalence class of unital qubit channels.[5] For clarity, we rewrite the optimization for (presumably CP) unital $\mathcal{M}$:

$$\alpha(\mathcal{M}) = 1 + \inf\left\{\sum_k t_k b_k : |b_1 \pm b_2| \leq |1 \pm b_3|\right\}\ . \tag{7.44}$$

We are now just interested in what values of $\vec{t}$ satisfying the CP constraints in (D.1) are such that the infimum equals unity. Using $\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha(\mathcal{N}) = 1 - (1 + \inf\{\sum_k t_k b_k : |b_1 \pm b_2| \leq |1 \pm b_3|\})$, we are interested in when the infimum is equal to negative one.

As established in [74], the conditions on the vector $\vec{b}$, which are positivity constraints, means that it lies in the tetrahedron defined by extreme points $\{(1,1,1), (1,-1,-1), (-1,1,-1), (-1,-1,1)\}$. By convexity, it will always be optimized on one of these extreme points. So we are in principle interested in

$$\min\left\{\sum_k t_k, t_1 - (t_2 + t_3), t_2 - (t_1 + t_3), t_3 - (t_1 + t_2)\right\} = -1 \tag{7.45}$$

over the $\vec{t}$ satisfying $|t_1 \pm t_2| \leq |1 \pm t_3|$. Again, this means

$$\vec{t} \in \mathrm{conv}(\{(1,1,1), (1,-1,-1), (-1,1,-1), (-1,-1,1)\})\ . \tag{7.46}$$

[5] The equivalence class is with respect to unitary pre- and post-processing.

We may then consider $\vec{t}$ as a convex combination of these points, which using $0 \leq p_0, p_1, p_2$ such that $p_0 + p_1 + p_2 \leq 1$, we can write

$$\vec{t} \coloneqq \begin{bmatrix} 2(p_0 + p_1) - 1 \\ 2(p_0 + p_2) - 1 \\ 1 - 2(p_1 + p_2) \end{bmatrix} , \tag{7.47}$$

where the parameterization is implicit. We are then interested when each function in (7.45) takes the value negative one as a function of p. Direct calculations lead to

$$\sum_k \vec{t}_k = -1 \Leftrightarrow p_0 = 0, \quad t_1 - (t_2 + t_3) \Leftrightarrow p_1 = 0, \tag{7.48}$$

$$t_2 - (t_1 + t_3) \Leftrightarrow p_2 = 0, \quad t_3 - (t_1 + t_2) \Leftrightarrow p_0 + p_1 + p_2 = 0 \ . \tag{7.49}$$

In other words, the set of $\vec{t}$ *not* satisfying (7.45) are those that are a mixture of *all* extreme points. □

The first part of the previous lemma reduced the consideration of $\alpha(\mathcal{N})$ for qubit channels to the 'unital part.' The following lemma shows that something similar holds for the trace distance contraction coefficient.

Lemma 7. *Let $\mathcal{N}$ be a qubit channel given by the transformation $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$. Then we have*

$$\eta_{\mathrm{Tr}}(\mathcal{N}) = \|T\|_\infty. \tag{7.50}$$

Proof. The same statement was proven in [61] for unital maps. It solely remains to show that for every channel $\mathcal{N}$, the non-unital part t does not contribute to the contraction coefficient, which is clear by direct calculation:

$$\eta_{\mathrm{Tr}}(\mathcal{N}) = \frac{1}{2} \sup_{w,w'} \left\| w_0 I + (t + Tw) \cdot \vec{\sigma} - (w_0 I + (t + Tw') \cdot \vec{\sigma}) \right\|_1 \tag{7.51}$$

$$= \frac{1}{2} \sup_{w,w'} \left\| Tw \cdot \vec{\sigma} - Tw' \cdot \vec{\sigma} \right\|_1 \tag{7.52}$$

$$= \frac{1}{2} \sup_{w,w'} \left\| T(w - w') \right\| \tag{7.53}$$

$$= \sup_w \|Tw\| \tag{7.54}$$

$$= \|T\|_\infty \ , \tag{7.55}$$

where $\|w\|, \|w'\| \leq 1$, as we optimize over quantum states and the final equality uses the optimal choice $w' = -w$ with $\|w'\| = 1$ and $\|X\|_1 = \mathrm{Tr}\,|X|$ to simplify. □

Finally, we use the above lemma to show that the conditions for $\eta_{\mathrm{Tr}}(\mathcal{N}) < 1$ are the same as $\alpha(\mathcal{N}) > 0$.

Proof of Theorem 5. By the unitary invariance of the trace-distance contraction coefficient, we may convert $\mathcal{N}$ to $\mathcal{N}'$, where the latter has a Choi operator of the form in (7.37). That is to say, $\mathcal{N}'$ is given by the transformation $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (\vec{r}' + T') \cdot \vec{\sigma}$ where $T' = \mathrm{diag}(t'_1, t'_2, t'_3)$.

It follows that $\eta_{\mathrm{Tr}}(\mathcal{N}) = \eta_{\mathrm{Tr}}(\mathcal{N}') = \|T'\|_\infty = \max\{t'_1, t'_2, t'_3\} =: \vec{t}'$, where the second equality used Lemma 7 and the third used the definition of the operator norm.

Re-parameterizing $\vec{t}'$ in terms of the extreme points of the CP qubit maps as was done in (7.47), we have that the contraction coefficient is the maximal entry of

$$\vec{t}' := \begin{bmatrix} 2(p_0+p_1)-1 \\ 2(p_0+p_2)-1 \\ 1-2(p_1+p_2) \end{bmatrix}, \tag{7.56}$$

where $0 \leq p_0, p_1, p_2$ and $p_0 + p_1 + p_2 \leq 1$. It is now easy to see that any of the entries being one, is equivalent to at least one coefficient being zero, e.g.

$$2(p_0+p_1)-1 = 1 \Leftrightarrow p_0 + p_1 = 1 \tag{7.57}$$

$$\Rightarrow p_2 = 0. \tag{7.58}$$

Hence the support of the corresponding convex combination can not be full. These are the same conditions that Lemma 6 specifies for $\alpha(\mathcal{N}) > 0$. Thus the conditions are equivalent. □

Proposition 39 (Doeblin Coefficients for Qubit Channels)**.** *Let $\mathcal{N}$ be a qubit channel given by the transformation $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$. Then, we have that*

$$\alpha(\mathcal{N}) = 1 - \lambda_1(T) - \lambda_2(T) + \lambda_3(T), \tag{7.59}$$

where $\lambda_1(T)$, $\lambda_2(T)$, $\lambda_3(T)$ are the ordered eigenvalues of matrix T (i.e., such that $\lambda_1(T) \geq \lambda_2(T) \geq \lambda_3(T)$).

Furthermore,

$$\alpha_I(\mathcal{N}) = 1 - \lambda_1(T) = 1 - \|T\|_\infty = 1 - \eta_{\mathrm{Tr}}(\mathcal{N}). \tag{7.60}$$

Proof. By (7.43) and the assumption on the channel $\mathcal{N}$ such that the following transformation happens $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$, we have that

$$\alpha(\mathcal{N}) = 1 + \inf\left\{ \sum_k t_k b_{k,k} : |b_{1,1} \pm b_{2,2}| \leq |1 \pm b_{3,3}| \right\}, \tag{7.61}$$

where $T = \mathrm{diag}(b_{1,1}, b_{2,2}, b_{3,3})$ and $t = (t_1, t_2, t_3)^T$. As established in [74], the conditions on the vector $\vec{b}$, which are positivity constraints, means that it lies in the tetrahedron defined by the extreme points $\{(1,1,1), (1,-1,-1), (-1,1,-1), (-1,-1,1)\}$. Then, the infimum is achieved when negative weights (-1) are assigned for the two largest eigenvalues and positive weight $(+1)$ is assigned for the smallest eigenvalue. This leads to the desired expression of

$$\alpha(\mathcal{N}) = 1 - \lambda_1(T) - \lambda_2(T) + \lambda_3(T). \tag{7.62}$$

To prove the next statement, we employ ideas presented around [48, Eq. (41)]. Recall the dual program for $\alpha_I(\mathcal{N})$ in Proposition 7. Note that a separable operator on two qubits can be generally written as $Y = \sum_k \alpha_k^T \otimes \beta_k$. Then, to demand the constraints of the dual, we need $\mathrm{Tr}[\alpha_k] = 1$ and $\sum_k \beta_k = I$. With that we also choose $\alpha_k := \frac{1}{2}(I + a_k \cdot \vec{\sigma})$ and $\beta_k := \gamma_k(I + b_k \cdot \vec{\sigma})$ with $\sum_k \gamma_k b_k = 0$, $\sum_k \gamma_k = 1$, and $\|b_k\|_2, \|a_k\|_2 \leq 1$. With these consider,

$$\mathrm{Tr}[\Gamma^{\mathcal{N}} Y] = \mathrm{Tr}\left[\Gamma^{\mathcal{N}} \left(\sum_k \alpha_k^T \otimes \beta_k \right) \right] \tag{7.63}$$

$$= \sum_k \operatorname{Tr}[\beta_k \mathcal{N}(\alpha_k)] \tag{7.64}$$

$$= \sum_k \frac{\gamma_k}{2} \operatorname{Tr}[(I + b_k \cdot \vec{\sigma})\,(I + (Ta_k + t) \cdot \vec{\sigma})] \tag{7.65}$$

$$= 1 + \sum_k \gamma_k b_k \cdot (Ta_k + t) \tag{7.66}$$

$$= 1 + \sum_k \gamma_k b_k^T T a_k, \tag{7.67}$$

where the last equality follows due to the condition $\sum_k \gamma_k b_k = 0$. To find $\alpha_I(\mathcal{N})$ we need to infimize $1 + \sum_k \gamma_k b_k^T T a_k$, which leads to infimizing $\sum_k \gamma_k b_k^T T a_k$. Then, the function of our interest is lower bounded by $\sum_k \gamma_k b_k^T T a_k \geq 1 - \lambda_1(T)$ since $\sum_k \gamma_k = 1$ and $b_k^T T a_k \geq -\sigma_{\max}(T) = -\lambda_1(T)$, where $\sigma_{\max}(\cdot)$ corresponds to the largest singular value. The inequality there is achieved by the following choice: choose $\gamma_k = 1/2$ for $k = \{1, 2\}$ and zero otherwise; $b_1 = -b_2 = -a_1 = a_2 = -u_1$, where u_1 is the eigenvector corresponding to the maximum eigenvalue of T (i.e., corresponding to $\lambda_1(T)$). Also, note that the above choice satisfies the constraints $\sum_k \gamma_k b_k = 0$, $\sum_k \gamma_k = 1$, and $\|b_k\|_2, \|a_k\|_2 \leq 1$. Then, we arrive at

$$\alpha_I(\mathcal{N}) = 1 - \lambda_1(T) = 1 - \|T\|_\infty, \tag{7.68}$$

concluding the proof by also recalling Lemma 7. □

Proposition 40. *Let $\mathcal{N}$ be a qubit channel given by the transformation $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$, and let $n \in \mathbb{N}$. Then, we have that, for all n-qubit states ρ and σ,*

$$T(\mathcal{N}^{\otimes n}(\rho), \mathcal{N}^{\otimes n}(\sigma)) \leq 4n \|T\|_\infty T(\rho, \sigma). \tag{7.69}$$

As such, we have that

$$\eta_{\mathrm{Tr}}(\mathcal{N}^{\otimes n}) \leq 4n \|T\|_\infty. \tag{7.70}$$

Proof. First, recall that all quantum channels have a quantum state that is a fixed point [11, Theorem 6.11]. With that, let τ be a state that is a fixed point of the qubit channel $\mathcal{N}$ (i.e.; $\mathcal{N}(\tau) = \tau$). With that, consider that, for an arbitrary qubit state ω,

$$\|\mathcal{N}(\omega) - \tau\|_1 = \|\mathcal{N}(\omega) - \mathcal{N}(\tau)\|_1 \tag{7.71}$$

$$\leq \|T\|_\infty \|\omega - \tau\|_1 \tag{7.72}$$

$$\leq 2 \|T\|_\infty, \tag{7.73}$$

where we use Lemma 7 for the first inequality.

By [75, Proposition 11], we have that

$$\|\mathcal{N}^{\otimes n}\|_{W_1} \leq \sup_{\rho_{RA}} \|(\mathrm{id}_R \otimes (\mathcal{N} - \mathcal{E}))(\rho_{RA})\|_1 \tag{7.74}$$

where $\mathcal{E}$ is a replacer channel that acts as $\mathcal{E}(X) = \operatorname{Tr}[X]\tau$, and for a channel $\mathcal{P}$,

$$\|\mathcal{P}\|_{W_1} := \sup_{\rho \neq \sigma} \frac{\|\mathcal{P}(\rho - \sigma)\|_{W_1}}{\|\rho - \sigma\|_{W_1}}, \tag{7.75}$$

where $\|\cdot\|_{W_1}$ is the quantum Wasserstein distance of order 1 [75, Definition 7]. Then using Lemma 8 with $d_A = d_B = 2$ therein, we arrive at

$$\left\|\mathcal{N}^{\otimes n}\right\|_{W_1} \leq 2 \sup_{\omega} \|(\mathcal{N} - \mathcal{E})(\omega)\|_1 . \tag{7.76}$$

See also [19, Proposition 33].

Combining (7.76) with the previous bound on the trace norm with the fixed state τ in (7.73), we see that

$$\left\|\mathcal{N}^{\otimes n}\right\|_{W_1} \leq 4 \|T\|_\infty . \tag{7.77}$$

Also recall that, from [75, Proposition 2], for a traceless Hermitian operator X,

$$\frac{1}{2} \|X\|_1 \leq \|X\|_{W_1} \leq \frac{n}{2} \|X\|_1 . \tag{7.78}$$

Utilizing this together with

$$\left\|\mathcal{N}^{\otimes n}(\rho - \sigma)\right\|_{W_1} \leq \left\|\mathcal{N}^{\otimes n}\right\|_{W_1} \|\rho - \sigma\|_{W_1} \leq 4 \|T\|_\infty \|\rho - \sigma\|_{W_1} , \tag{7.79}$$

we arrive at

$$\frac{1}{2} \left\|\mathcal{N}^{\otimes n}(\rho - \sigma)\right\|_1 \leq 4 \|T\|_\infty \cdot \frac{n}{2} \|\rho - \sigma\|_1 , \tag{7.80}$$

concluding the proof, together with $T(\rho, \sigma) = \|\rho - \sigma\|_1 / 2$ and the definition of $\eta_{\mathrm{Tr}}(\cdot)$. □

Lemma 8. *Let $\mathcal{N}$ and $\mathcal{M}$ be two quantum channels that map $\mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$, with d_A and d_B the dimensions of the input and the output systems, respectively. Then, we have that*

$$\sup_{\rho_{RA}} \|(\mathrm{id}_R \otimes (\mathcal{N} - \mathcal{M}))(\rho_{RA})\|_1 \leq \sqrt{d_A d_B} \sup_{\rho} \|(\mathcal{N} - \mathcal{M})(\rho)\|_1 , \tag{7.81}$$

where system R is isomorphic to system A (i.e., $d_R = d_A$).

Proof. Consider that

$$\begin{aligned}
\sup_{\rho_{RA}} \|(\mathrm{id}_R \otimes (\mathcal{N} - \mathcal{M}))(\rho_{RA})\|_1 &= \sup_{\rho_{RA}} \frac{\|(\mathrm{id}_R \otimes (\mathcal{N} - \mathcal{M}))(\rho_{RA})\|_1}{\|\rho_{RA}\|_1} && (7.82)\\
&\leq \sqrt{d_A d_B} \sup_{\rho_{RA}} \frac{\|(\mathrm{id}_R \otimes (\mathcal{N} - \mathcal{M}))(\rho_{RA})\|_2}{\|\rho_{RA}\|_1} && (7.83)\\
&= \sqrt{d_A d_B} \sup_{\rho} \frac{\|(\mathcal{N} - \mathcal{M})(\rho)\|_2}{\|\rho\|_1} && (7.84)\\
&\leq \sqrt{d_A d_B} \sup_{\rho} \frac{\|(\mathcal{N} - \mathcal{M})(\rho)\|_1}{\|\rho\|_1} && (7.85)\\
&= \sqrt{d_A d_B} \sup_{\rho} \|(\mathcal{N} - \mathcal{M})(\rho)\|_1 , && (7.86)
\end{aligned}$$

where the first inequality follows because $\|A\|_1 \leq \sqrt{d}\, \|A\|_2$ with $d = d_A \cdot d_B$ the dimension of A; the second equality follows from the stability of the superoperator norms induced by the two norm, by using [76, Theorem 4] with the choice $p = 2 \geq 2$ and $q = 1 \leq 2$ therein; and the last inequality again by the change of two norm to one norm using $\|A\|_2 \leq \|A\|_1$. □

Remark 10 (Comparison to Existing Work)**.** *For a qubit channel $\mathcal{N}$ specified by the transformation $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$, by applying the techniques used in the proof of [19, Proposition 3], one can obtain that*

$$T(\mathcal{N}^{\otimes n}(\rho), \mathcal{N}^{\otimes n}(\sigma)) \leq 8n \max_{p \in \{X,Y,Z\}} \left| \frac{D_p}{D_p - 1} \right| T(\rho, \sigma), \tag{7.87}$$

where $T = \operatorname{diag}(D_X, D_Y, D_Z)$. Comparing it with Proposition 40, we see that Proposition 40 is tighter since $\|T\|_\infty \leq \max_{p \in \{X,Y,Z\}} \left| \frac{D_p}{D_p - 1} \right|$. With that Proposition 40 provides a non-trivial worst-case bound for the trace-distance contraction for a larger set of noise channels than previously known. Moreover, we improve the contraction coefficient of Wasserstein distance of order 1 of $\mathcal{N}^{\otimes n}$ defined in (7.75) given in [19, Lemma 32] by establishing (7.77).

Next, we provide an analytical expression for the expansion coefficient in (6.53) for qubit channels.

Lemma 9 (Expansion Coefficient for Qubit Channels)**.** *Let $\mathcal{N}$ be a qubit channel specified by the transformation $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$. Then, we have that*

$$\breve{\eta}_{\mathrm{Tr}}(\mathcal{N}) = \sigma_{\min}(T), \tag{7.88}$$

where $\breve{\eta}_{\mathrm{Tr}}$ is defined in (6.53) and $\sigma_{\min}(A)$ is the minimum singular value of the operator A.

Proof. By recalling the definition of $\breve{\eta}_{\mathrm{Tr}}$ in (6.53), we have

$$\breve{\eta}_{\mathrm{Tr}}(\mathcal{N}) = \inf_{\rho_1, \rho_2 : T(\rho_1, \rho_2) \neq 0} \frac{T(\mathcal{N}(\rho_1), \mathcal{N}(\rho_2))}{T(\rho_1, \rho_2)}. \tag{7.89}$$

Also denote the qubit states as follows for $i \in \{1, 2\}$:

$$\rho_i = \frac{1}{2}(I + w_i \cdot \vec{\sigma}), \tag{7.90}$$

$$\mathcal{N}(\rho_i) = \frac{1}{2}\left(I + (t + Tw_i) \cdot \vec{\sigma}\right). \tag{7.91}$$

Then, we arrive at the following:

$$\breve{\eta}_{\mathrm{Tr}}(\mathcal{N}) = \inf_{w_1, w_2 : \|w_1 - w_2\|_2 \neq 0} \frac{\|T(w_1 - w_2)\|_2}{\|w_1 - w_2\|_2} \tag{7.92}$$

$$= \inf_{w' : \|w'\|_2 \neq 0} \frac{\|Tw'\|_2}{\|w'\|_2} \tag{7.93}$$

$$= \sigma_{\min}(T), \tag{7.94}$$

where the first equality follows by substituting qubit representations to the trace distance and using the fact that $T(\rho_1, \rho_2) = \|w_1 - w_2\|_2$. □

7.6 Generalized Amplitude-Damping Channels

Consider the generalized amplitude damping (GAD) channel $\mathcal{N}_{p,\eta}$, defined for $p \in [0,1]$ and $\eta \in [0,1]$ in terms of the following Kraus operators:

$$K_1 \coloneqq \sqrt{p}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\eta} \end{bmatrix}, \quad K_2 \coloneqq \sqrt{p}\begin{bmatrix} 0 & \sqrt{1-\eta} \\ 0 & 0 \end{bmatrix}, \tag{7.95}$$

$$K_3 \coloneqq \sqrt{1-p}\begin{bmatrix} \sqrt{\eta} & 0 \\ 0 & 1 \end{bmatrix}, \quad K_4 \coloneqq \sqrt{1-p}\begin{bmatrix} 0 & 0 \\ \sqrt{1-\eta} & 0 \end{bmatrix}, \tag{7.96}$$

such that $\mathcal{N}_{p,\eta}(\cdot) \coloneqq \sum_{i=1}^{4} K_i(\cdot)K_i^\dagger$. See [77, Sections III-C and III-D] for various interpretations and discussions of the generalized amplitude damping channel.

Lemma 10. *For $p \in [0,1]$ and $\eta \in [0,1]$, the Doeblin coefficient of the generalized amplitude damping channel is given by*

$$\alpha(\mathcal{N}_{p,\eta}) = (1-\sqrt{\eta})^2. \tag{7.97}$$

Proof. By definition of the Doeblin coefficient, i.e., using the expression in (3.1), it is sufficient to find a feasible Hermitian operator X in order to give a lower bound on the coefficient. We choose

$$X = \begin{bmatrix} p+(1-p)\eta-\sqrt{\eta} & 0 \\ 0 & p\eta+(1-p)-\sqrt{\eta} \end{bmatrix}. \tag{7.98}$$

For this choice of X to be feasible, the following matrix inequality should hold:

$$0 \leq \Gamma_{AB}^{\mathcal{N}} - I_A \otimes X_B = \begin{bmatrix} \sqrt{\eta} & 0 & 0 & \sqrt{\eta} \\ 0 & \sqrt{\eta}-\eta & 0 & 0 \\ 0 & 0 & \sqrt{\eta}-\eta & 0 \\ \sqrt{\eta} & 0 & 0 & \sqrt{\eta} \end{bmatrix}. \tag{7.99}$$

It can easily be seen to hold, because $0 \leq \eta \leq 1$. The Choi matrix for the generalized amplitude damping channel can be found in [77, Eq. (105)] and [14, Appendix C]. The following inequality thus holds:

$$\alpha(\mathcal{N}_{p,\eta}) \geq \mathrm{Tr}[X] = (1-\sqrt{\eta})^2. \tag{7.100}$$

To arrive at the opposite inequality, let us employ the dual program in (3.3) with the specific choice of

$$Y_{AB} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}. \tag{7.101}$$

Here, we have that $\mathrm{Tr}_A[Y_{AB}] = I$ and $Y_{AB} \geq 0$ (eigenvalues being 0 and 2), making it a feasible solution for the dual program. For that choice,

$$\mathrm{Tr}\left[\Gamma_{AB}^{\mathcal{N}_{p,\eta}} Y_{AB}\right] = p+(1-p)\eta-2\sqrt{\eta}+p\eta+(1-p) = 1-2\sqrt{\eta}+\eta = (1-\sqrt{\eta})^2 \tag{7.102}$$

so that

$$(1-\sqrt{\eta})^2 \geq \alpha(\mathcal{N}_{p,\eta}). \tag{7.103}$$

Combining (7.100) and (7.103), we conclude (7.97). □

Lemma 10 implies the following inequality:

$$\eta_{\mathrm{Tr}}(\mathcal{N}_{p,\eta}) \leq 1 - (1 - \sqrt{\eta})^2. \tag{7.104}$$

The right-hand side is strictly smaller than one for all $\eta < 1$, and it is independent of p.

Lemma 11. *For $p \in [0,1]$ and $\eta \in [0,1]$, the quantity $\alpha_{\pm}(\mathcal{N}_{p,\eta})$ for the generalized amplitude damping channel $\mathcal{N}_{p,\eta}$ is given by*

$$\alpha_{\pm}(\mathcal{N}_{p,\eta}) = \begin{cases} \frac{2p(1-2p)(1-\eta)^2}{\eta+2p(1-\eta)} & p \leq \frac{\sqrt{\eta}}{2(1+\sqrt{\eta})} \\ (1-\sqrt{\eta})^2 & \frac{\sqrt{\eta}}{2(1+\sqrt{\eta})} \leq p \leq 1 - \frac{\sqrt{\eta}}{2(1+\sqrt{\eta})} \\ \frac{2(1-p)(2p-1)(1-\eta)^2}{\eta+2(1-p)(1-\eta)} & p \geq 1 - \frac{\sqrt{\eta}}{2(1+\sqrt{\eta})} \end{cases}. \tag{7.105}$$

Proof. For the middle case, it is sufficient to show that the solution to $\alpha(\mathcal{N}_{p,\eta})$ also fulfills the stronger conditions here. For the first case, we start by showing that the given solution is a lower bound. Consider the matrix

$$X = \begin{bmatrix} -p(1-\eta) & 0 \\ 0 & p\eta + 1 - p - \frac{\eta}{\eta+2p(1-\eta)} \end{bmatrix}. \tag{7.106}$$

For this to be a feasible solution we need to show that,

$$0 \leq \Gamma^{\mathcal{N}}_{AB} - I_A \otimes X_B, \tag{7.107}$$

$$0 \leq \Gamma^{\mathcal{N}}_{AB} + I_A \otimes X_B, \tag{7.108}$$

which can again be done by checking positivity by explicit calculation. The bound is then exactly given by $\mathrm{Tr}\, X$. For the third case, the lower bound follows similarly by choosing

$$X = \begin{bmatrix} (1-p)\eta + p - \frac{\eta}{\eta+2(1-p)(1-\eta)} & 0 \\ 0 & -(1-p)(1-\eta) \end{bmatrix}. \tag{7.109}$$

Now we move on to prove the converse. For the middle case, the upper bound is established by choosing the same Y in (7.101) as a specific choice for $Y_2 = Y$ and $Y_1 = 0$ in the dual SDP of $\alpha_{\pm}(\mathcal{N})$ presented in (3.18).

For the third case, consider [6]

$$Y_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & x_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad Y_2 = \begin{bmatrix} 1 & 0 & 0 & x_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ x_2 & 0 & 0 & 1 + x_1 \end{bmatrix}, \tag{7.110}$$

with

$$x_1 \coloneqq \frac{(1-\eta)\left(\eta(1-2p)^2 - 4(1-p)^2\right)}{\left((2p-1)\eta + 2(1-p)\right)^2}, \tag{7.111}$$

[6] Note that the following choices were made with the support of numerical simulations and the respective optimizers.

$$x_2 \coloneqq \frac{-\sqrt{\eta}}{(2p-1)\eta + 2(1-p)}. \tag{7.112}$$

The chosen Y_1 and Y_2 are feasible choices for the dual SDP since $Y_1, Y_2 \geq 0$ and $\mathrm{Tr}_A[Y_2 - Y_1] = I_B$. This follows by the fact that $x_1 \geq 0$ for $1 \geq p \geq 1 - (\sqrt{\eta} - \eta)/(2(1-\eta))$ and $1 + x_1 \geq 0$ with $(1 + x_1) - x_2^2 \geq 0$. Furthermore, we have

$$\mathrm{Tr}[\Gamma^{\mathcal{N}}_{AB}(Y_1 + Y_2)] = \frac{2(1-p)(2p-1)(1-\eta)^2}{(2p-1)\eta + 2(1-p)} \geq b(\mathcal{N}_{p,\eta}). \tag{7.113}$$

For the first case to obtain the desired upper bound, choose

$$Y_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & x_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad Y_2 = \begin{bmatrix} 1 + x_1 & 0 & 0 & x_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ x_2 & 0 & 0 & 1 \end{bmatrix} \tag{7.114}$$

and the proof follows by replacing p with $(1-p)$. □

Remark 11 (Multiplicativity)**.** *Due to multiplicativity of $\alpha_{\mp}(\mathcal{N})$ (Proposition 25), we also conclude from Lemma 11 and Corollary 6 that $\alpha\left(\mathcal{N}_{p,\eta}^{\otimes n}\right) = (\alpha(\mathcal{N}_{p,\eta}))^n$ whenever*

$$\frac{\sqrt{\eta}}{2(1+\sqrt{\eta})} \leq p \leq 1 - \frac{\sqrt{\eta}}{2(1+\sqrt{\eta})} \tag{7.115}$$

since $\alpha(\mathcal{N}_{p,\eta}) = \alpha_{\mp}(\mathcal{N}_{p,\eta})$ in that regime.

For comparison, we also give exact expressions for the contraction and expansion coefficients.

Lemma 12 (Contraction Coefficient for the GAD Channel)**.** *Let $p \in [0,1]$ and $\eta \in [0,1]$. Then, the contraction coefficient and expansion coefficient of the GAD channel is characterized as follows:*

$$\eta_{\mathrm{Tr}}(\mathcal{N}_{p,\eta}) = \sqrt{\eta}, \tag{7.116}$$
$$\check{\eta}_{\mathrm{Tr}}(\mathcal{N}_{p,\eta}) = \eta. \tag{7.117}$$

Proof. Note that for the GAD channel with parameters p and η, $T = \mathrm{diag}(\sqrt{\eta}, \sqrt{\eta}, \eta)$ (recall the qubit channel mapping $w_0 I + w \cdot \vec{\sigma} \to w_0 I + (t + Tw) \cdot \vec{\sigma}$.). The, the proof follows by directly applying Lemma 7 and Lemma 9 since $\|T\|_\infty = \sqrt{\eta}$ and $\sigma_{\min}(T) = \eta$. □

8 Applications of Quantum Doeblin Coefficients

In this section, we discuss how quantum Doeblin coefficients and their properties can be utilized in a diverse range of applications, including limitations on noise-induced barren plateaus, error mitigation protocols, noisy hypothesis testing, fairness of quantum learning models, and mixing times of time-inhomogeneous quantum Markov chains. Our analysis in terms of quantum Doeblin coefficients provides insights over previous works, and we obtain efficiently computable limitations in all of these applications when the systems involved are noisy.

8.1 Noise-Induced Barren Plateaus

Variational quantum algorithms and quantum neural networks have been considered in the context of near-term quantum computing [78], with the possibility of using quantum and classical resources in tandem for learning and inference tasks. There, the quantum computational resources are (mostly) used to estimate a cost function and then classical computational resources are used to optimize the parameters of the variational quantum circuits by minimizing the cost function estimated. The barren plateau phenomenon occurs when the optimization landscape becomes extremely flat and thus the parameters of a quantum circuit become untrainable when using gradient-based optimizers [79, 80].

In practice, near-term quantum devices are noisy. As such, it is a natural question to understand how well one can train parameterized quantum circuits in the presence of noisy gates. This was originally studied under a model in which Pauli qubit noise or depolarizing noise acts on each qubit before and after each unitary in a quantum circuit [16]. The authors of [16] showed that, when noise is present, it is harder to train the parameters because the partial derivatives of the cost function with respect to the parameters decay exponentially with the depth of the circuit. This phenomenon was termed "noise-induced barren plateaus". Then, Ref. [17] explored this phenomenon when the parameter shift rule is used as the gradient estimation technique and when the noisy channel $\mathcal{N}$ has a non-zero trace distance contraction coefficient (i.e., $\eta_{\mathrm{Tr}}(\mathcal{N}) < 1$). Moreover, Ref. [17] showed that noise-induced barren plateaus occur when the parameterized unitary gates in each layer form a (global) unitary two-design (i.e., when $\mathcal{U}_i(\vec{\theta}_i)$ in (8.1) realizes a unitary two-design, for all i). Ref. [19] extended the study of noise-induced barren plateaus with non-unital noise, for random quantum circuits that form local unitary two-designs and such that qubit local noise channels are interleaved with unitary gates. Note that [19] focused on the average case and showed that non-unital noise may help in avoiding vanishing variance of a (local) cost function under the selected random circuit structure; they arrived at the same conclusions on the trainability of the parameters with the depth of the circuit for cost functions with global observables.

However, previous works focused only on a subset of circuit architectures and noise models (e.g., random quantum circuits, circuits with Pauli qubit noise) and most of those bounds are hard to compute efficiently (e.g., estimating $\eta_{\mathrm{Tr}}(\mathcal{N})$ for noise channels). In this work, we extend the study of the impact of noise on the trainability of quantum circuits to more general noise models and architectures, and we devise efficiently computable bounds using Doeblin coefficients.

Let us consider the following model for the quantum channel realized by a noisy parameterized quantum circuit:

$$\mathcal{M}_{(\vec{\theta}_1,\ldots,\vec{\theta}_D)} \coloneqq \mathrm{id}_R \otimes \left(\mathcal{N}_D \circ \mathcal{U}_D(\vec{\theta}_D) \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{U}_1(\vec{\theta}_1)\right), \tag{8.1}$$

where each $\mathcal{N}_i$ is a noisy channel and each $\mathcal{U}_i$ is a parameterized unitary channel for $i \in \{1, \ldots, D\}$. Also, suppose that each unitary channel $\mathcal{U}_i(\vec{\theta}_i) = U_i(\vec{\theta}_i)(\cdot)U_i(\vec{\theta}_i)^\dagger$ is comprised of the following unitary operators:

$$U_i(\vec{\theta}_i) = \prod_{j=1}^{J_i} \exp\left(-\frac{i\theta_i^j H_i^j}{2}\right), \tag{8.2}$$

with each H_i^j a Hermitian operator satisfying $\|H_i^j\|_\infty \leq 1$. Also, note that our analysis applies

to the case in which these unitaries act on n qudits (our results boil down to the qubit case simply by setting $d = 2$).

Let the cost function to be evaluated be as follows:

$$C(\vec{\theta}_1, \ldots, \vec{\theta}_D) \coloneqq \operatorname{Tr}\left[O\mathcal{M}_{(\vec{\theta}_1,\ldots,\vec{\theta}_D)}(\rho_0)\right], \tag{8.3}$$

for an observable O and an initial state ρ_0 acting on the Hilbert space $\mathcal{H}_R \otimes \mathcal{H}_A$ of the joint system of R and A.

Next, we establish an upper bound on the partial derivatives of the cost function with respect to the circuit parameters (e.g., rotation angles) given a noisy quantum circuit taking the form of (8.1). In the proof of Proposition 41, we make use of some techniques from [17,19] (trace distance contraction and Taylor expansions of partial derivatives around a point), together with the contraction bounds based on Doeblin coefficients, as studied in our paper.

Proposition 41 (Barren Plateaus with General Noisy Circuits and Gradient Based Optimization)**.** *Fix $D \in \mathbb{N}$, and let $i \in \{1, \ldots, D\}$ and $j \in \{1, \ldots, J_i\}$. Then, we have the following upper bound on the partial derivatives of the cost function in (8.3):*

$$\left|\frac{\partial C(\vec{\theta}_1, \ldots, \vec{\theta}_D)}{\partial \theta_i^j}\right| \leq C_1 \left\|O\right\|_\infty \left(\prod_{k=i}^{D} (1 - \alpha(\mathcal{N}_k))\right)^{2/3}, \tag{8.4}$$

where

$$C_1 \coloneqq \left(\frac{8}{3}\right)^{1/3} + \frac{4}{3}\left(\frac{3}{8}\right)^{2/3} \leq 2.0801. \tag{8.5}$$

For the setting in which $\mathcal{N}_i = \mathcal{N}$ for all $i \in \{1, \ldots, D\}$, we have that

$$\left|\frac{\partial C(\vec{\theta}_1, \ldots, \vec{\theta}_D)}{\partial \theta_i^j}\right| \leq C_1 \left\|O\right\|_\infty (1 - \alpha(\mathcal{N}))^{2(D-i+1)/3}. \tag{8.6}$$

Proof. Recall from [19, Eq. (C32)] that

$$\left|\frac{\partial f(x)}{\partial x}\right| \leq \left|\frac{f(x+h) - f(x-h)}{2h}\right| + \frac{h^2}{6} \sup_{y \in (x-h, x+h)} \left|\frac{\partial^3 f(y)}{\partial^3 x}\right|. \tag{8.7}$$

Fix $i \in \{1, \ldots, D\}$ and $j \in \{1, \ldots, J_i\}$. Then also pick two values for θ_i^j, given by $\theta_i^j = \theta + h$ and $\theta_i^j = \theta - h$, where $\theta \in \mathbb{R}$ and $h > 0$. With that, choose $\vec{\theta_i^+}$ and $\vec{\theta_i^-}$ such that all other entries are the same except for θ_i^j, which takes on the fixed two values. Define the shorthand

$$\mathcal{P}_{[D,i]}^{j,\theta_i^j} \coloneqq \mathrm{id}_R \otimes \left(\mathcal{N}_D \circ \mathcal{U}_D(\vec{\theta}_D) \circ \cdots \circ \mathcal{N}_i \circ \mathcal{U}_i(\vec{\theta}_i)\right), \tag{8.8}$$

$$\mathcal{P}_{[k,1]} \coloneqq \mathrm{id}_R \otimes \left(\mathcal{N}_k \circ \mathcal{U}_k(\vec{\theta}_k) \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{U}_1(\vec{\theta}_1)\right), \tag{8.9}$$

and

$$\rho_{i-1} \coloneqq \mathcal{P}_{[i-1,1]}(\rho_0) \tag{8.10}$$

for $i > 1$, with ρ_0 being the initial input state, which lies in $\mathcal{D}(\mathcal{H}_R \otimes \mathcal{H}_A)$. Note that we have removed the dependence on the circuit parameters for brevity.

Now consider that

$$\begin{aligned}
&\operatorname{Tr}\left[O\left(\mathcal{P}^{j,\theta+h}_{[D,i]}(\rho_i)-\mathcal{P}^{j,\theta-h}_{[D,i]}(\rho_i)\right)\right] \\
&\leq \|O\|_\infty \left\|\mathcal{P}^{j,\theta+h}_{[D,i]}(\rho_i)-\mathcal{P}^{j,\theta-h}_{[D,i]}(\rho_i)\right\|_1 && (8.11)\\
&\leq \|O\|_\infty \prod_{k=i}^{D}(1-\alpha(\mathcal{N}_k))\left\|\left(\mathrm{id}_R\otimes\mathcal{U}_i\left(\vec{\theta_i^+}\right)\right)(\rho_{i-1})-\left(\mathrm{id}_R\otimes\mathcal{U}_i\left(\vec{\theta_i^-}\right)\right)(\rho_{i-1})\right\|_1 && (8.12)\\
&\leq 2\|O\|_\infty \prod_{k=i}^{D}(1-\alpha(\mathcal{N}_k)), && (8.13)
\end{aligned}$$

where the first inequality follows from Hölder's inequality; the second from the repetitive application of the complete contraction of trace distance under each noise layer until the i-th layer in (6.16), together with the unitary invariance property of trace distance (recall that the marginal state at the subsystem R is $\operatorname{Tr}_A[\rho_0]$ for all states considered); and the last inequality follows because $\|\rho-\sigma\|_1\leq 2$ for all states ρ and σ.

Also, from [19, Lemma 45], we have that

$$\left|\frac{\partial^3 C(\vec{\theta}_1,\ldots,\vec{\theta}_D)}{\partial^3\theta_i^j}\right| \leq 8\|O\|_\infty \left\|I_R\otimes H_{\theta_i^j}\right\|_\infty^3 = 8\|O\|_\infty \left\|H_{\theta_i^j}\right\|_\infty^3. \tag{8.14}$$

With that and together with the assumption that $\left\|H_{\theta_i^j}\right\|_\infty\leq 1$, we arrive at

$$\left|\frac{\partial^3 C(\vec{\theta}_1,\ldots,\vec{\theta}_D)}{\partial^3\theta_i^j}\right| \leq 8\|O\|_\infty. \tag{8.15}$$

Recalling (8.7), consider that

$$\begin{aligned}
&\left|\frac{\partial C(\vec{\theta}_1,\ldots,\vec{\theta}_D)}{\partial\theta_i^j}\right| \\
&\leq \frac{1}{2h}\left|\operatorname{Tr}\left[O\left(\mathcal{P}^{j,\theta+h}_{[D,i]}(\rho_i)-\mathcal{P}^{j,\theta-h}_{[D,i]}(\rho_i)\right)\right]\right| + \frac{h^2}{6}\sup_{\theta_i^j}\left|\frac{\partial^3 C(\vec{\theta}_1,\ldots,\vec{\theta}_D)}{\partial^3\theta_i^j}\right| && (8.16)\\
&\leq \frac{1}{2h}2\|O\|_\infty\prod_{k=i}^{D}(1-\alpha(\mathcal{N}_k)) + \frac{8h^2}{6}\|O\|_\infty && (8.17)\\
&= \|O\|_\infty\left(\frac{8}{3}\right)^{1/3}\left(\prod_{k=i}^{D}(1-\alpha(\mathcal{N}_k))\right)^{2/3} + \frac{4}{3}\left(\frac{3}{8}\prod_{k=i}^{D}(1-\alpha(\mathcal{N}_k))\right)^{2/3}\|O\|_\infty && (8.18)\\
&\leq C_1\|O\|_\infty\left(\prod_{k=i}^{D}(1-\alpha(\mathcal{N}_k))\right)^{2/3}, && (8.19)
\end{aligned}$$

where the second inequality follows by (8.13) and (8.15); the equality follows by substituting

$$h=\left(\frac{3}{8}\prod_{k=i}^{D}(1-\alpha(\mathcal{N}_k))\right)^{1/3}, \tag{8.20}$$

which minimizes the right-hand side. This concludes the proof of (8.4).

The inequality in (8.6) follows by substituting $\mathcal{N}_k=\mathcal{N}$ in (8.4), concluding the proof. □

Corollary 12 (Barren Plateaus with Product Noisy Channels)**.** *Let $i \in \{1, \ldots, D\}$ and $j \in \{1, \ldots, J_i\}$. Then, we have the following upper bound on the partial derivatives of the cost function given in (8.3):*

$$\left|\frac{\partial C(\vec{\theta}_1, \ldots, \vec{\theta}_D)}{\partial \theta_i^j}\right| \leq C_1 \left\|O\right\|_\infty \left(\prod_{k=i}^{D} \left(1 - (\alpha_{\mp}(\mathcal{P}_k))^\ell\right)\right)^{2/3}, \tag{8.21}$$

where $\mathcal{N}_i = \mathcal{P}_i^{\otimes \ell}$ and

$$C_1 \coloneqq \left(\frac{8}{3}\right)^{1/3} + \frac{4}{3}\left(\frac{3}{8}\right)^{2/3} \leq 2.0801. \tag{8.22}$$

For the setting in which $\mathcal{N}_i = \mathcal{N} = \mathcal{P}^{\otimes \ell}$, we have that

$$\left|\frac{\partial C(\vec{\theta}_1, \ldots, \vec{\theta}_D)}{\partial \theta_i^j}\right| \leq C_1 \left\|O\right\|_\infty \left(1 - (\alpha_{\mp}(\mathcal{P}))^\ell\right)^{2(D-i+1)/3}. \tag{8.23}$$

Proof. The proof follows by adapting together Proposition 41 with $\alpha(\mathcal{N}) \geq \alpha_{\mp}(\mathcal{N})$ and $\alpha_{\mp}(\mathcal{P}^{\otimes \ell}) = (\alpha_{\mp}(\mathcal{P}))^\ell$ by the multiplicativity of $\alpha_{\mp}(\mathcal{N})$ (Proposition 25). □

Remark 12 (Gradient Computation Using Parameter Shift Rule)**.** *Similar to the approach used in [17] for the case when the parameter shift rule [81–84] can be used as the gradient estimation technique, one can arrive at a stronger statement in terms of the depth of the circuit compared to the results obtained in Proposition 41. In particular, we find that*

$$\left|\frac{\partial C(\vec{\theta}_1, \ldots, \vec{\theta}_D)}{\partial \theta_i^j}\right| \leq C' \left\|O\right\|_\infty \prod_{k=i}^{D} \left(1 - \alpha(\mathcal{N}_k)\right), \tag{8.24}$$

where C' is a constant depending on the parameter shift rule used.

Next, we elaborate on the impact of the unital noise channels (i.e., a channel $\mathcal{N}$ such that $\mathcal{N}(I) = I$) on the partial derivatives of the cost function.

Proposition 42 (Barren Plateaus with Unital Noise)**.** *Fix $D \in \mathbb{N}$, and let $i \in \{1, \ldots, D\}$ and $j \in \{1, \ldots, J_i\}$. Then, we have the following upper bound on the partial derivatives of the cost function given in (8.3):*

$$\left|\frac{\partial C(\vec{\theta}_1, \ldots, \vec{\theta}_D)}{\partial \theta_i^j}\right| \leq 3\left(\frac{4}{3}\right)^{1/3} \left\|O\right\|_\infty \left(\prod_{k=1}^{D} \left(1 - \alpha(\mathcal{N}_k)\right)\right)^{2/3}. \tag{8.25}$$

Furthermore,

$$\left|C(\vec{\theta}_1, \ldots, \vec{\theta}_D) - \frac{1}{|R|d^n} \operatorname{Tr}[O]\right| \leq 2 \left\|O\right\|_\infty \left(\prod_{k=1}^{D} \left(1 - \alpha(\mathcal{N}_k)\right)\right), \tag{8.26}$$

where $|R|$ is the size of the system R.

For the setting for which $\mathcal{N}_i = \mathcal{N}$ for all $i \in \{1, \ldots, D\}$, we have that

$$\left|\frac{\partial C(\vec{\theta}_1, \ldots, \vec{\theta}_D)}{\partial \theta_i^j}\right| \leq 3\left(\frac{4}{3}\right)^{1/3} \left\|O\right\|_\infty \left(\left(1 - \alpha(\mathcal{N})\right)\right)^{2D/3}. \tag{8.27}$$

Proof. For unital noise channels (i.e., satisfying $\mathcal{N}_i(I) = I$), consider the following, starting from (8.12):

$$\begin{aligned}
&\operatorname{Tr}\left[O\left(\mathcal{P}^{j,\theta+h}_{[D-i+1,i]}(\rho_i) - \mathcal{P}^{j,\theta-h}_{[D-i+1,i]}(\rho_i)\right)\right] \\
&\leq \|O\|_\infty \prod_{k=i}^{D} (1-\alpha(\mathcal{N}_k)) \left\| \left(\mathrm{id}_R \otimes \mathcal{U}_i\left(\vec{\theta_i^+}\right)\right)(\rho_{i-1}) - \left(\mathrm{id}_R \otimes \mathcal{U}_i\left(\vec{\theta_i^-}\right)\right)(\rho_{i-1}) \right\|_1 \qquad (8.28) \\
&\leq \|O\|_\infty \prod_{k=i}^{D} (1-\alpha(\mathcal{N}_k)) \left(\left\| \left(\mathrm{id}_R \otimes \mathcal{U}_i\left(\vec{\theta_i^+}\right)\right)(\rho_{i-1}) - \frac{I}{|R|d^n} \right\|_1 + \right. \\
&\qquad \left. \left\| \left(\mathrm{id}_R \otimes \mathcal{U}_i\left(\vec{\theta_i^-}\right)\right)(\rho_{i-1}) - \frac{I}{|R|d^n} \right\|_1 \right) \qquad (8.29) \\
&\leq \|O\|_\infty \prod_{k=i}^{D} (1-\alpha(\mathcal{N}_k)) \left(\left\| \left(\mathrm{id}_R \otimes \mathcal{U}_i\left(\vec{\theta_i^+}\right)\right)(\rho_{i-1}) - \frac{I}{|R|d^n} \right\|_1 + \right. \\
&\qquad \left. \left\| \left(\mathrm{id}_R \otimes \mathcal{U}_i\left(\vec{\theta_i^+}\right)\right)(\rho_{i-1}) - \left(\mathrm{id}_R \otimes \mathcal{U}_i\left(\vec{\theta_i^-}\right)\right)\left(\frac{I}{|R|d^n}\right) \right\|_1 \right) \qquad (8.30) \\
&\leq \|O\|_\infty \prod_{k=i}^{D} (1-\alpha(\mathcal{N}_k)) \prod_{k=i-1}^{1} (1-\alpha(\mathcal{N}_k)) \left(\left\| \rho_0 - \frac{I}{|R|d^n} \right\|_1 + \left\| \rho_0 - \frac{I}{|R|d^n} \right\|_1 \right) \qquad (8.31) \\
&\leq 4 \|O\|_\infty \prod_{k=1}^{D} (1-\alpha(\mathcal{N}_k)) , \qquad (8.32)
\end{aligned}$$

where the second inequality follows from the triangular inequality of the trace norm; the first equality follows because the noise channels $\mathcal{N}$ are unital by assumption and all other channels are unitary; the third inequality follows by considering the complete contraction of trace distance under the rest of the noise channels after the i-th layer in (6.16), together with the unitary invariance of the trace norm and recalling $\rho_{i-1} \coloneqq \mathcal{P}_{[i-1,1]}(\rho_0)$; and finally the last inequality follows because $\|\rho - \sigma\|_1 \leq 2$ for all states ρ and σ.

We conclude the inequality (8.25) on the partial derivative for the case of unital channels by combining (8.32) together with the steps followed after (8.16), with the choice

$$h = \left(\frac{3}{4} \prod_{k=i}^{D} (1-\alpha(\mathcal{N}_k)) \right)^{1/3} . \qquad (8.33)$$

The second inequality (8.26) follows similarly to the proof of the previous claim in (8.32) by the application of Hölder's inequality, complete contraction of the trace distance together with unitary invariance of the trace norm. In particular, it proceeds as above by considering that

$$\left| C(\vec{\theta}_1, \ldots, \vec{\theta}_D) - \frac{1}{|R|d^n} \operatorname{Tr}[O] \right| = \left| \operatorname{Tr}\left[O \left(\mathcal{P}_{[D,1]}(\rho_0) - \mathcal{P}_{[D,i]}\left(\frac{I}{|R|d^n} \right) \right) \right] \right| \qquad (8.34)$$

The last inequality (8.27) follows by substituting $\mathcal{N}_i = \mathcal{N}$ for all $i \in \{1, \ldots, D\}$, concluding the proof. □

Discussion Our results presented in Proposition 41 demonstrate that the learnability of the circuit parameters decreases with the depth of the corresponding parameterized unitary from the end of the circuit. In fact, the first few layers are not trainable whenever $\alpha(\mathcal{N}_i) > 0$ for all $i \in \{1, \ldots, D\}$. To the best of our knowledge, our finding here is the first to demonstrate this phenomenon even with *noiseless subsystems* available within the circuit. In previous works [16,17,19], all qubits are considered to be impacted by noise channels. Thus, our work shows that it is not necessary for all the qubits to be noisy in order to observe this trainability issue in terms of the decaying of partial derivatives of the cost function. These noiseless subsystems can even be understood as corresponding to the existence of a noiseless feedback system. Furthermore, our results hold regardless of the size of the noiseless subsystem R. To this end, considering the trivial subsystem R, our results recover previously known results from [16,17]. Also, note that our results can be extended to the setting where there exists a fixed unitary in the noiseless subsystem R.

The noise models used in our analysis are quite general, satisfying only $\alpha(\mathcal{N}) > 0$ or $\alpha_{\mp}(\mathcal{N}) > 0$, without any assumption on the specific structure of noise or the gradient-optimization technique used. For instance, Ref. [16] analysed noise-induced barren plateaus under a special class of i.i.d. qubit Pauli noise channels, while Ref. [19] worked with qubit noise channels (both unital and non-unital). For general noise channels (without noiseless subsystems), Ref. [17] provided analytical upper bounds on the partial derivatives for noise channels satisfying $\eta_{\mathrm{Tr}}(\mathcal{N}) < 1$, when the parameter shift rule is used as the gradient estimation technique (when the generators H_i^j of the parameterized unitary operators in (8.2) have only two unique eigenvalues or unitary operators can be decomposed into product of operators satisfying the said property [84]), while Ref. [18] has considered noise channels where the Hilbert–Schmidt norm is contractive (one norm in $\eta_{\mathrm{Tr}}(\mathcal{N})$ replaced by the two nrom). A possible downside of those approaches is that the estimation of $\eta_{\mathrm{Tr}}(\mathcal{N})$ or the two-norm contraction are computationally difficult. However, our bounds obtained in Proposition 41 and Corollary 12 are efficiently computable by SDPs. With that, it is possible to compute the bounds in order to obtain insights about how exactly the depth of the circuit and the noise channels impact the decaying of the derivative magnitudes and the trainability of the circuit parameters.

Moreover, our results hold for all parameterized quantum circuits, regardless of how the initialization of parameters is carried out. Some previous works [17,19] on noise-induced barren plateaus have focused on random quantum circuits (e.g., global and two-qubit unitary two-designs) and analysed the impact on the variance of the cost function to arrive at conclusions similar to ours. Also, in [19], it was shown that, with non-unital noise, a lower bound on the variance can be obtained, which in turn may avoid barren plateaus for such circuits for local cost functions. Note that these results hold on average. However, in practice there are use-cases where the parameters are not initialized randomly and the circuit structures are decided based on the prior knowledge and domain expertise. In those settings, results obtained under the assumption of random circuit initializations are not applicable. Also, it is suspected that smart initializations through warm starts and pre-training of parameters may serve as a strategy to overcome barren plateaus [85]. However, our results show a negative result here *when the circuits are noisy*. In fact, we show that, even if one uses clever initialization procedures to select parameters and unitary circuit structures, with noisy circuits we cannot harness the suspected advantage. This further highlights that, regardless of the expressivity of the circuit (e.g., unitary two-designs have higher expressivity), noise within

the parameterized quantum circuits leads to barren plateaus, hindering trainability.

From Proposition 41 and Corollary 12 for general noise channels, we deduce that even though the last few layers can be potentially trained, the first few layers are very hard to train whenever $\alpha(\mathcal{N}_i) > 0$ or $\alpha_{\mp}(\mathcal{P}_i) > 0$ for $i \in \{1, \ldots, D\}$. This also suggests that one can only (potentially) train the last few layers and fix the circuit parameters of the first few layers without optimizing them. In contrast, if the noise channels are unital, Proposition 42 demonstrates that the trainability of the circuit parameters decays exponentially with the overall depth of the circuit regardless of the location of the unitary in the circuit. Our upper bounds on the partial derivatives from Proposition 42 provide insights for selecting the depth of the circuit for learning and inference tasks that require optimization of parameters, when the noise is unital.

Furthermore, we observe that some noisy circuits may be classically simulable. With reference to (8.26) in Proposition 42, we see that, with increasing depth D of the circuit and $\alpha(\mathcal{N}_i) > 0$ for all $i \in \{1, \ldots, D\}$, the cost function value $C(\vec{\theta}_1, \ldots, \vec{\theta}_D)$ concentrates around $\frac{1}{|R|d^n} \operatorname{Tr}[O]$. To this end, whenever $\operatorname{Tr}[O] = 0$, one can estimate the cost function value to be zero. In that case, it is possible to classically simulate the cost function, which is an expectation value of the observable O with a state dependent on the circuit parameters. Moreover, it is possible to obtain a confidence interval for the cost function value as follows: consider the scenario in which $\mathcal{N}_i = \mathcal{N}$ for all $i \in \{1, \ldots, D\}$; then by rewriting (8.26) we arrive at

$$\frac{1}{|R|d^n} \operatorname{Tr}[O] - 2 \|O\|_\infty (1 - \alpha(\mathcal{N}))^D \leq C(\vec{\theta}_1, \ldots, \vec{\theta}_D) \leq \frac{1}{|R|d^n} \operatorname{Tr}[O] + 2 \|O\|_\infty (1 - \alpha(\mathcal{N}))^D . \tag{8.35}$$

Note that (8.26) is valid for unital noise channels. For qubit noise channels, Ref. [19, Proposition 35] developed a procedure to classically simulate the cost function when O is replaced by a Pauli operator, up to an absolute error $\varepsilon > 0$ with a high probability over random quantum circuits (on average). We leave further exploration on the classical simulability of the cost functions under general noise channels and general circuit architectures for future work.

In summary, our study of noise-induced barren plateaus extends to new noisy circuits with the possibility of having noiseless subsystems and circuit architectures independent of the parameter initialization and the gradient estimation technique. Also, we provide efficiently computable limits on the scaling of partial derivatives of the cost function and the overall concentration of the cost function under noisy parametrized quantum circuits.

8.2 Error Mitigation

The goal of error mitigation is to learn expectation values of observables while mitigating the impact of errors caused by noise [20]. Here, we specifically discuss the task of estimating the expectation value $\operatorname{Tr}[A\rho]$ of an observable A up to some error and with high probability, by mitigating the impact of the errors caused by noise in the quantum circuits.

Definition 14 (Error Mitigation Protocol)**.** *Fix $\varepsilon > 0$ and $\delta \in (0, 1)$. Fix $\mathcal{S}$ to be a set of states and $\mathcal{O}$ to be a set of observables. Let $A \in \mathcal{O}$ be a Hermitian operator. An error mitigation protocol $\mathcal{P}_A$ (as depicted in Fig. 2) takes as input N noisy quantum states, $\bigotimes_{i=1}^{N} \mathcal{Q}_i(\rho)$, where each $\mathcal{Q}_i$ is a noisy quantum channel and $\rho \in \mathcal{S}$ is the original state of interest. Then, the channel $\mathcal{P}_A$ produces an estimate of $\operatorname{Tr}[A\rho]$, which we denote by $\hat{E}(\rho, A, N)$, as an output.*

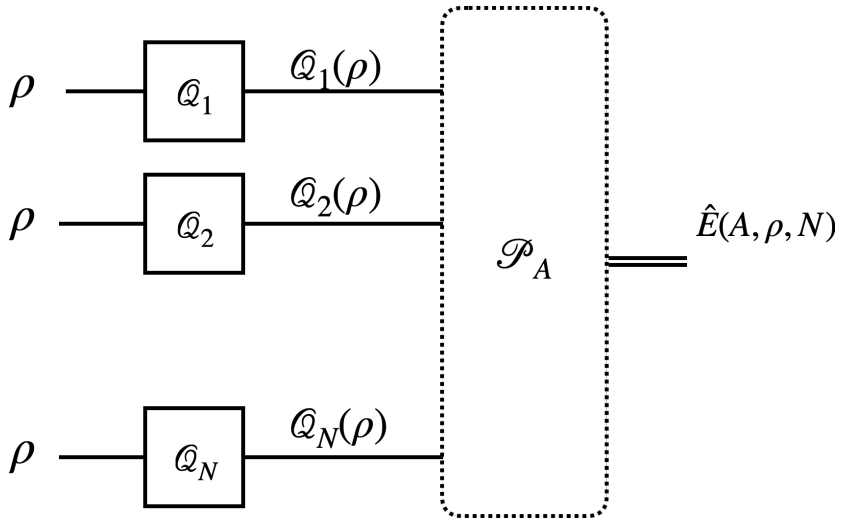


Figure 2: Error mitigation protocol $\mathcal{P}_A$: The goal of the protocol is to provide an estimate for $\mathrm{Tr}[A\rho]$ for an observable A and all states ρ in a certain class $\mathcal{S}$. In particular, $\mathcal{P}_A$ takes as input the N distorted input states $\bigotimes_{i=1}^{N} \mathcal{Q}_i(\rho)$, where each $\mathcal{Q}_i$ is a quantum channel, and outputs an estimate for $\mathrm{Tr}[A\rho]$ denoted as $\hat{E}(A,\rho,N)$ such that $\left|\mathrm{Tr}[A\rho] - \hat{E}(\rho,A,N)\right| \leq \varepsilon$ with probability not less than $1-\delta$ for $\varepsilon > 0$ and $\delta \in (0,1)$.

We say that the error mitigation protocol is successful if

$$\Pr\Big(\left|\mathrm{Tr}[A\rho] - \hat{E}(\rho,A,N)\right| \leq \varepsilon\Big) \geq 1-\delta, \tag{8.36}$$

for all $\rho \in \mathcal{S}$ and $A \in \mathcal{O}$.

Note that $\mathcal{P}_A$ in Fig. 2 includes many error mitigation protocols known and also protocols for which classical post-processing is allowed (see [21] and references therein).

Also define,

$$D_{\mathcal{O}}(\rho,\sigma) \coloneqq \sup_{A\in\mathcal{O}} \left|\mathrm{Tr}[A(\rho-\sigma)]\right|. \tag{8.37}$$

Using the proof of Theorem 1 and Theorem 3 in [21] along with our bounds on (complete) contraction of trace distance, we next analyse the error mitigation capabilities and limitations of a broader class of noisy quantum channels where the *noisy channels are not necessarily unital*. Also note that our bounds represent fundamental limitations that are independent of which particular error mitigation protocol is used to obtain the estimate $\hat{E}(\rho,A,N)$.

Let the noise channels $\mathcal{Q}_i$ in Fig. 2 for all $i \in \{1,\ldots,N\}$ be characterized as follows:

$$\mathcal{Q}_i = \mathrm{id}_R \otimes (\mathcal{N}_D \circ \mathcal{U}_D \circ \cdots \circ \mathcal{N}_j \circ \mathcal{U}_j \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{U}_1), \tag{8.38}$$

where each $\mathcal{N}_j$ is a noisy channel and each $\mathcal{U}_j$ is a unitary channel, for $j \in \{1,\ldots,D\}$.

Theorem 6 (Sample Complexity of Error Mitigation)**.** *Suppose that an error-mitigation strategy in Fig. 2 satisfies (8.36) for some $\varepsilon > 0$ and $0 \leq \delta \leq 1/2$. Also fix $\mathcal{Q}_i = \mathcal{Q}_k$ for all $i,k \in \{1,\ldots,N\}$, and suppose that $\mathcal{Q}_i$ is given by (8.38). If there exist $\rho_{RA}, \sigma_{RA} \in \mathcal{S}$ such*

that $D_{\mathcal{O}}(\rho_{RA}, \sigma_{RA}) \geq 2\varepsilon$ and $\rho_R = \sigma_R$, then an error mitigation protocol requires N noisy states such that the following inequality should hold

$$N \geq \frac{(1-2\delta)}{\prod_{i=1}^{D}(1-\alpha(\mathcal{N}_i))} \tag{8.39}$$

in order to satisfy (8.36) for all $\rho_{RA} \in \mathcal{S}$ and $A_{RA} \in \mathcal{O}$.

Proof. The proof follows by starting in a similar fashion to the proof of Theorem 1 and Theorem 3 in [21], with the assumption that there exist $\rho_{RA}, \sigma_{RA} \in \mathcal{S}$ such that $D_{\mathcal{O}}(\rho_{RA}, \sigma_{RA}) \geq 2\varepsilon$.

Consider Eq. (A5) of [21], where $(\mathcal{Q}_i)_{i=1}^{N}$ denotes the tuple of noisy channels, each of which is applied on one of the N copies of the state. Let us suppose that the same noisy channel (with D noisy layers) is applied on each copy of the given state. Then

$$1 - 2\delta \leq T\left(\bigotimes_{i=1}^{N} \mathcal{Q}_i(\rho_{RA}), \bigotimes_{i=1}^{N} \mathcal{Q}_i(\sigma_{RA})\right) \tag{8.40}$$

$$\leq N\, T(\mathcal{Q}(\rho_{RA}), \mathcal{Q}(\sigma_{RA}))\,, \tag{8.41}$$

where the last inequality holds by the subadditivity of trace distance and choosing $\mathcal{Q}_i = \mathcal{Q}$.

Note that $\mathcal{Q} = \mathrm{id} \otimes (\mathcal{N}_D \circ \mathcal{U}_D \circ \cdots \circ \mathcal{N}_j \circ \mathcal{U}_j \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{U}_1)$, where each $\mathcal{U}_i$, for all $i \in \{1, \ldots, D\}$, is a unitary circuit applied in between each noisy channel. Also, let us define the following shorthand for $i \in \{1, \ldots, D\}$:

$$\mathcal{P}_i \coloneqq \mathrm{id}_R \otimes (\mathcal{N}_i \circ \mathcal{U}_i \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{U}_1). \tag{8.42}$$

Then, we have that

$$\begin{aligned}
&T(\mathcal{Q}(\rho_{RA}), \mathcal{Q}(\sigma_{RA})) \\
&\leq (1-\alpha(\mathcal{N}_D))\, T((\mathrm{id}_R \otimes \mathcal{U}_D) \circ \mathcal{P}_{D-1}(\rho_{RA}), (\mathrm{id}_R \otimes \mathcal{U}_D) \circ \mathcal{P}_{D-1}(\sigma_{RA})) &(8.43)\\
&= (1-\alpha(\mathcal{N}_D))\, T(\mathcal{P}_{D-1}(\rho_{RA}), \mathcal{P}_{D-1}(\sigma_{RA}))\,, &(8.44)
\end{aligned}$$

where the first inequality follows from applying Proposition 30 for the channel $\mathcal{N}_D$, along with the equality $\mathrm{Tr}_A[\mathcal{P}_{D-1}(\rho_{RA})] = \mathrm{Tr}_A[\mathcal{P}_{D-1}(\sigma_{RA})]$ due to $\rho_R = \sigma_R$; and the last equality follows by unitary invariance of the trace distance.

Next, by the repetitive application of the Doeblin coefficient bound in order to bound the complete contraction of trace distance, due to $D-1$ noisy channels and by using unitary invariance of trace distance, we have that

$$T(\mathcal{Q}(\rho_{RA}), \mathcal{Q}(\sigma_{RA})) \leq \prod_{i=1}^{D}(1-\alpha(\mathcal{N}_i))\; T(\rho_{RA}, \sigma_{RA}) \tag{8.45}$$

$$\leq \prod_{i=1}^{D}(1-\alpha(\mathcal{N}_i)), \tag{8.46}$$

where the second inequality follows because $T(\rho_{RA}, \sigma_{RA}) \leq 1$. Finally, putting everything together, we arrive at the desired result. □

The lower bound in Theorem 6 is SDP-computable. Thus, when we are given the description of the set of noisy channels, it can be evaluated efficiently when the dimension of the systems are small. However, in general, the computational complexity increases exponentially with an increasing number of qubits. Instead, for a channel $\mathcal{N} = \mathcal{P}^{\otimes n}$, one can use $\alpha_{\overline{+}}(\mathcal{P})$ to provide an efficiently computable bound by employing multiplicativity of $\alpha_{\overline{+}}$ (Proposition 25), as presented next.

Remark 13 (Error Mitigation Limits with Local Noisy Channels)**.** *By considering local noisy qubit channels instead of global noisy channels as in Theorem 6, we need*

$$N \geq \frac{(1-2\delta)}{\prod_{i=1}^{D}\left(1-\alpha(\mathcal{P}_i^{\otimes n})\right)} \tag{8.47}$$

$$\geq \frac{(1-2\delta)}{\prod_{i=1}^{D}\left(1-\alpha_{\overline{+}}(\mathcal{P}_i)^n\right)}, \tag{8.48}$$

where $\alpha_{\overline{+}}(\mathcal{P}_i)$ *is defined in* (3.13)*.*

Discussion Error mitigation under deep noisy quantum circuits has been studied in [21,22]. Those works showed that there is an exponential dependence of the depth of the circuit on the number of noisy data samples required for error mitigation. This suggests that, without access to an exponential number of noisy data samples, it is not possible to mitigate errors when estimating the expectation values of observables (see Definition 14 for the success criterion of an error-mitigation protocol).

In particular, [21, Theorem 3] showed this exponential scaling of the required number of noisy input samples when the noisy quantum channels are depolarizing noise channels (i.e., consider $\mathcal{Q}_i$ in Fig. 2 as given by (8.1) with R system being the trivial system and each $\mathcal{N}_i$ being a depolarizing channel with flip parameter p). In [22], the study was extended to general unital and non-unital noise channels with the assumption of random unitaries in the circuit (i.e., consider $\mathcal{Q}_i$ in Fig. 2 as given by (8.1) with R system being the trivial system and each $\mathcal{U}_i$ being (independently) randomly sampled from the set of unitaries that form a unitary two-design). In this setting with random unitaries and qubit noise channels, Ref. [22] showed that there exists an exponential overhead with respect to both the depth and the width (number of qubits) of the circuit.

However, in practice, the unitaries are not always random quantum circuits (i.e., they can be fixed or chosen from a different sampling or optimization approach), and the noise channels that can arise in quantum devices are far beyond depolarizing noise or even qubit noise channels. In our work, using Doeblin coefficients, we extend the analysis of the resource requirements for successful error mitigation to a large class of noisy channels and quantum circuit architectures. Theorem 6 demonstrates that this exponential dependence holds for a large class of noise channels whenever $\alpha(\mathcal{N}_i) > 0$.

Let us consider an example setting in which each $\mathcal{N}_i = \mathcal{N}_{p,\eta}$ for $p \in [0,1]$ and $\eta \in (0,1)$. Then, using the previously known bound from [21, Eq. (I22)] (by choosing $k = 1$ therein), we have the following lower bound

$$N \geq \frac{(1-2\delta)^2 \min\{p, 1-p\}\lambda_{\min}(\Gamma^{N_{p,\eta}})}{\eta^D 8\ln(2)}, \tag{8.49}$$

where

$$\lambda_{\min}(\Gamma^{\mathcal{N}_{p,\eta}}) = \frac{1+\eta}{2} - \sqrt{\eta + (1-\eta)^2(1-2p)^2}. \tag{8.50}$$

This then raises the question whether the exponential dependence of the depth D of the circuit on the sample complexity of error mitigation techniques also vanishes in the case of $p \in \{0, 1\}$. However, our result in Theorem 6 shows that there is still an exponential dependence on D, as long as $\eta \in (0, 1)$ for all $p \in [0, 1]$, because $\alpha(\mathcal{N}_{p,\eta}) = (1 - \sqrt{\eta})^2$. Also, note that $p = 1$ is a practically relevant noise model, known as the amplitude damping channel. This highlights a practically relevant scenario where our results provide useful insights.

Our results in Theorem 6 hold even if there is a noiseless subsystem (denoted by R in (8.1)), showcasing that there exists an exponential scaling of the required resources, characterized by $\alpha(\mathcal{N}_i)$. Thus, our result is valid for a large class of noise models. One example towards this is the gate-based noise models, where the noise of the circuit is significant when only non-identity unitary circuits are applied. In this scenario, we can assume that in the noiseless system R, there are no non-identity unitaries in operation. Most, importantly this conveys that having a noiseless subsystem is not sufficient to circumvent the exponential resource requirement for error mitigation.

Another advantage of our results in Theorem 6 is that the lower bound given there is SDP computable. This allows one to get a sense of how the required number of noisy data samples scales for an error mitigation protocol to achieve a desired accuracy depending on the depth of the circuit and the noise channels related to specific applications of interest. When there are local noise channels in the system (i.e., $\mathcal{N}_i = \mathcal{P}_i^{\otimes n}$), we see that the exponential scaling depends on $\alpha(\mathcal{P}_i^{\otimes n})$ as given in Remark 13. By Remark 7, the lower bound in (8.47) can be computed with the time complexity scaling as $O(\text{poly}(n))$ by utilizing permutation symmetry of $\mathcal{P}_i^{\otimes n}$. Furthermore, a somewhat weaker bound given in (8.48) can be efficiently computed by using $\alpha_{\text{王}}(\mathcal{N})$, without any computational complexity scaling with n.

In summary, our results provide efficiently computable limits on the sample complexity of a broad range of error-mitigation protocols under various noise models satisfying $\alpha(\mathcal{N}) > 0$ or $\alpha_{\text{王}}(\mathcal{N}) > 0$ and circuit architectures with general quantum circuits (not necessarily random circuits) and circuits with noiseless subsystems.

8.3 Noisy Quantum Hypothesis Testing

In this section, we study the impact of noise on the hypothesis testing of two quantum states. In particular, this refers to the scenario where we need to guess whether a given state is ρ or σ, when we have access to only multiple samples of the states $\mathcal{N}(\rho)$ or $\mathcal{N}(\sigma)$ (finite), where $\mathcal{N}$ is a noisy quantum channel. To this end, first let us recall the noiseless scenario, for which $\mathcal{N}$ is the identity channel, and then utilize those results together with contraction coefficients and Doeblin coefficients to quantify the impact of having access only to noisy samples of quantum data.

Sample Complexity of Noiseless Quantum Hypothesis Testing *Problem setup*: Suppose that there are two states ρ and σ, $\rho^{\otimes n}$ is selected with probability $\beta \in (0, 1)$, and $\sigma^{\otimes n}$ is selected with probability $1 - \beta$. The sample complexity is equal to the minimum value of n needed to reach a constant error probability in deciding which state was selected. To define

this quantity formally, let us recall that the Helstrom–Holevo theorem [86,87] states that the optimal error probability $p_e(\rho,\sigma,\beta)$ of symmetric quantum hypothesis testing is as follows:

$$p_e(\rho,\sigma,\beta) \coloneqq \min_{M_1,M_2\geq 0} \{\beta \operatorname{Tr}[M_2\rho] + (1-\beta)\operatorname{Tr}[M_1\sigma] : M_1 + M_2 = I\} \tag{8.51}$$

$$= \frac{1}{2}\left(1 - \|\beta\rho - (1-\beta)\sigma\|_1\right). \tag{8.52}$$

With this in mind, we are assuming in this paradigm that there is a constant $\varepsilon \in [0,1]$, and our goal is to determine the minimum value of n such that

$$p_e(\rho^{\otimes n},\sigma^{\otimes n},\beta) \coloneqq \frac{1}{2}\left(1 - \|\beta\rho^{\otimes n} - (1-\beta)\sigma^{\otimes n}\|_1\right) \leq \varepsilon. \tag{8.53}$$

To this end, let us define

$$\mathrm{SC}_{(\rho,\sigma)}(\varepsilon,\beta) \coloneqq \min\{n \in \mathbb{N} : p_e(\rho^{\otimes n},\sigma^{\otimes n},\beta) \leq \varepsilon\}, \tag{8.54}$$

and recall Theorem 7 and Corollary 8 of [88].

Theorem 7 (Sample Complexity of Symmetric Hypothesis Testing (Theorem 7 and Corollary 8 of [88]), [89, Theorem 2])**.** *Let ρ and σ be states. Fix $\varepsilon \in [0,1]$ and $\beta \in (0,1)$. The following statements hold:*

1. *If $\rho \perp \sigma$ (i.e., $\rho\sigma = 0$), $\varepsilon \in [1/2,1]$, or $\exists\, s \in [0,1]$ such that $\varepsilon \geq \beta^s(1-\beta)^{1-s}$, then*

$$\mathrm{SC}_{(\rho,\sigma)}(\varepsilon,\beta) = 1. \tag{8.55}$$

2. *If $\rho = \sigma$ and $\min\{\beta, 1-\beta\} > \varepsilon \in [0,1/2)$, then*

$$\mathrm{SC}_{(\rho,\sigma)}(\varepsilon,\beta) = +\infty. \tag{8.56}$$

3. *If the conditions mentioned above are excluded, then*

$$\max\left\{\frac{\ln\left(\frac{\beta(1-\beta)}{\varepsilon(1-\varepsilon)}\right)}{-\ln F(\rho,\sigma)}, \frac{1-\frac{\varepsilon(1-\varepsilon)}{\beta(1-\beta)}}{[d_{\mathrm{B}}(\rho,\sigma)]^2}\right\} \leq \mathrm{SC}_{(\rho,\sigma)}(\varepsilon,\beta) \leq \left\lceil \frac{2\ln\left(\frac{\sqrt{\beta(1-\beta)}}{\varepsilon}\right)}{-\ln F(\rho,\sigma)} \right\rceil, \tag{8.57}$$

where $F(\rho,\sigma) \coloneqq \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$ and $[d_B(\rho,\sigma)]^2 \coloneqq 2\left(1 - \sqrt{F(\rho,\sigma)}\right)$.

In [88], the sample complexity of hypothesis testing of noiseless quantum data has been studied. Following that, the sample complexity of hypothesis testing with access to only privatized (noisy) quantum states was studied in [67,90]. In this case, privacy constraints are imposed by quantum local differential privacy [66]. See [91,92] for further details on ensuring privacy for quantum data. The contraction coefficients of quantum divergences under private (noisy) channels have been used as one of the key technical tools in the analysis of private quantum hypothesis testing.

In this work, we extend the study of the sample complexity of noisy quantum hypothesis testing when we are given noisy channel outputs of the original input states to discriminate. We utilize quantum Doeblin coefficients and the tools developed in this work to do so.

Let ρ and σ be states. Define, from [65],

$$H_{1/2}(\rho\|\sigma) \coloneqq \frac{1}{2}\int_1^\infty \left[E_\gamma(\rho\|\sigma) + E_\gamma(\sigma\|\rho)\right]\gamma^{-3/2}\,\mathsf{d}\gamma. \tag{8.58}$$

Proposition 43 (Sample Complexity of Noisy Quantum Hypothesis Testing)**.** *Let ρ and σ be states, and let $\mathcal{N}$ be a quantum channel. Fix $\varepsilon \in [0,1]$ and $\beta \in (0,1)$. The following statements hold:*

1. *If $\mathcal{N}(\rho) \perp \mathcal{N}(\sigma)$ (i.e., $\mathcal{N}(\rho)\mathcal{N}(\sigma) = 0$), $\varepsilon \in [1/2, 1]$, or $\exists\, s \in [0,1]$ such that $\varepsilon \geq \beta^s(1-\beta)^{1-s}$, then*

$$\mathrm{SC}_{(\mathcal{N}(\rho),\mathcal{N}(\sigma))}(\varepsilon,\beta) = 1. \tag{8.59}$$

2. *If $\mathcal{N}(\rho) = \mathcal{N}(\sigma)$ and $\min\{\beta, 1-\beta\} > \varepsilon \in [0, 1/2)$, then*

$$\mathrm{SC}_{(\mathcal{N}(\rho),\mathcal{N}(\sigma))}(\varepsilon,\beta) = +\infty. \tag{8.60}$$

3. *If the first two conditions are excluded, then*

$$\begin{aligned} \frac{1 - \frac{\varepsilon(1-\varepsilon)}{\beta(1-\beta)}}{(1-\alpha(\mathcal{N}))\, H_{1/2}(\rho\|\sigma)} &\leq \mathrm{SC}_{(\mathcal{N}(\rho),\mathcal{N}(\sigma))}(\varepsilon,\beta) \\ &\leq \left\lceil 2\ln\left(\frac{\sqrt{\beta(1-\beta)}}{\varepsilon}\right)\left(\frac{1}{(1-\check{\alpha}(\mathcal{N}))\, T(\rho,\sigma)}\right)^2 \right\rceil, \end{aligned} \tag{8.61}$$

   where $\alpha(\mathcal{N})$ and $\check{\alpha}(\mathcal{N})$ are defined in (3.1) and (6.54), respectively [7]*.*

4. *If the first two conditions are not satisfied and*

$$\lambda_{\rho,\sigma,\mathcal{N}} \coloneqq \max\left\{\lambda_{\min}(\mathcal{N}(\rho)), \lambda_{\min}(\mathcal{N}(\sigma))\right\} > 0, \tag{8.62}$$

   then

$$\mathrm{SC}_{(\mathcal{N}(\rho),\mathcal{N}(\sigma))}(\varepsilon,\beta) \geq \frac{\lambda_{\rho,\sigma,\mathcal{A}}}{4} \ln\left(\frac{\beta(1-\beta)}{\varepsilon(1-\varepsilon)}\right)\left(\frac{1}{(1-\alpha(\mathcal{N}))\, T(\rho,\sigma)}\right)^2. \tag{8.63}$$

Proof. Items (1) and (2) follow directly by Theorem 7.

For the lower bound in Item (3), consider the following:

$$\begin{aligned} [d_B(\mathcal{N}(\rho),\mathcal{N}(\sigma))]^2 &= 2\left(1 - \sqrt{F}(\mathcal{N}(\rho),\mathcal{N}(\sigma))\right) && (8.64)\\ &\leq H_{1/2}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) && (8.65)\\ &\leq \eta_{\mathrm{Tr}}(\mathcal{N}) H_{1/2}(\rho\|\sigma) && (8.66)\\ &\leq (1-\alpha(\mathcal{N}))\, H_{1/2}(\rho\|\sigma)\,, && (8.67) \end{aligned}$$

where the first inequality follows by [65, Eq. (5.50)], the second inequality by identifying that $H_{1/2}$ is an f-divergence and its contraction coefficient is upper bounded by the contraction coefficient of the trace distance; and the last inequality because $\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha(\mathcal{N})$. Substituting the above in the lower bound with the denominator $[d_{\mathrm{B}}(\rho,\sigma)]^2$ in Item 3 of Theorem 7 concludes the lower bound.

[7] Note that $1-\alpha(\mathcal{N})$ can be replaced by $\eta_{\mathrm{Tr}}(\mathcal{N})$ for a tighter lower bound and $1-\check{\alpha}(\mathcal{N})$ by $\check{\eta}_{\mathrm{Tr}}(\mathcal{N})$ for a tighter upper bound, whenever they are analytically computable. For an example, consider the GAD channel (Lemma 12).

For the upper bound in Item (3), consider that

$$-\ln \sqrt{F}(\mathcal{N}(\rho),\mathcal{N}(\sigma)) \geq \frac{1}{2}\left[T(\mathcal{N}(\rho),\mathcal{N}(\sigma))\right]^2, \tag{8.68}$$

which follows because

$$-\ln \sqrt{F}(\rho,\sigma) \geq \frac{1}{2}[d_{\mathrm{B}}(\rho,\sigma)]^2 \geq \frac{1}{2}[T(\rho,\sigma)]^2. \tag{8.69}$$

The above inequalities follow because $-\ln(x) \geq 1-x$ for $x>0$, and the last one due to the well known Fuchs-van-de-Graaf inequalities. Next, by employing the bound in (6.53) involving the reverse Doeblin coefficient, we have that

$$T(\mathcal{N}(\rho),\mathcal{N}(\sigma)) \geq (1-\breve{\alpha}(\mathcal{N}))\, T(\rho,\sigma). \tag{8.70}$$

Combining (8.68) and (8.70), together with the upper bound in Item (3) of Theorem 7, completes the proof of the upper bound.

Now, let us prove the claim in Item (4). First recall that

$$-\ln F(\rho,\sigma) \leq D(\rho\|\sigma). \tag{8.71}$$

Also recall the assumption $\lambda_{\rho,\sigma,\mathcal{N}} \coloneqq \max\{\lambda_{\min}(\mathcal{N}(\rho)), \lambda_{\min}(\mathcal{N}(\sigma))\} > 0$. Without loss of generality, suppose that $\lambda_{\rho,\sigma,\mathcal{N}} = \lambda_{\min}(\mathcal{N}(\sigma))$ (when it is otherwise, interchange ρ and σ). With that, we have that

$$\begin{aligned}
-\ln F(\mathcal{N}(\rho),\mathcal{N}(\sigma)) &\leq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) && (8.72)\\
&\leq \frac{4\left[T(\mathcal{N}(\rho),\mathcal{N}(\sigma))\right]^2}{\lambda_{\min}(\mathcal{N}(\sigma))} && (8.73)\\
&\leq \frac{4\left(1-\alpha(\mathcal{N})\right)^2}{\lambda_{\min}(\mathcal{N}(\sigma))}\left[T(\rho,\sigma)\right]^2, && (8.74)
\end{aligned}$$

where the second inequality follows by [93, Theorem 2] and the monotonicity of norms such that $\|A\|_2 \leq \|A\|_1$; and the last inequality by $\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1-\alpha(\mathcal{N})$. Plugging the above inequality in the lower bound with the denominator $-\ln F(\rho,\sigma)$ in Item (3) of Theorem 7 completes the proof. □

Remark 14 (Noisy Hypothesis Testing of Quantum States with Reference Systems)**.** *Let ρ_{RA} and σ_{RA} be quantum states such that $\rho_R = \sigma_R$. Let noise channel $\mathcal{N}$ act on system A. Then, using the same proof techniques in Proposition 43, we can obtain the following upper and lower bounds as follows:*

$$\begin{aligned}
\frac{1-\frac{\varepsilon(1-\varepsilon)}{\beta(1-\beta)}}{(1-\alpha_+(\mathcal{N}))\, H_{1/2}(\rho_{RA}\|\sigma_{RA})} &\leq \mathrm{SC}_{(\mathrm{id}_R\otimes\mathcal{N}(\rho_{RA}),\mathrm{id}_R\otimes\mathcal{N}(\sigma_{RA}))}(\varepsilon,\beta)\\
&\leq \left\lceil 2\ln\left(\frac{\sqrt{\beta(1-\beta)}}{\varepsilon}\right)\left(\frac{1}{(1-\breve{\alpha}(\mathcal{N}))\, T(\rho_{RA},\sigma_{RA})}\right)^2\right\rceil, \qquad (8.75)
\end{aligned}$$

where α_+ is defined in Eq. (3.25), and $\breve{\alpha}$ in (6.54). The main technical tools employed in this claim are:

$$H_{1/2}(\mathrm{id}_R\otimes\mathcal{N}(\rho_{RA})\|\mathrm{id}_R\otimes\mathcal{N}(\sigma_{RA})) \leq (1-\alpha_+(\mathcal{N}))\, H_{1/2}(\rho_{RA}\|\sigma_{RA}), \tag{8.76}$$

which follows from Corollary 7 by the choice $D_f = H_{1/2}$ in (8.58), and

$$T(\mathrm{id}_R \otimes \mathcal{N}(\rho_{RA}) \| \mathrm{id}_R \otimes \mathcal{N}(\sigma_{RA})) \geq (1 - \breve{\alpha}(\mathcal{N}))\, T(\rho_{RA} \| \sigma_{RA}), \tag{8.77}$$

which follows due to Proposition 33.

Corollary 13 (Generalized Amplitude Damping Channels)**.** *Let ρ and σ be qubit states, and let $\mathcal{N} \equiv \mathcal{N}_{p,\eta}$ be a generalized damping channel with parameters $\eta \in (0,1)$ and $p \in [0,1]$. Fix $\varepsilon \in (0,1)$ and $\beta \in (0,1)$. For the cases where the first two conditions of Proposition 43 do not hold, we have the following upper and lower bound on the sample complexity:*

$$\frac{1 - \frac{\varepsilon(1-\varepsilon)}{\beta(1-\beta)}}{\sqrt{\eta} H_{1/2}(\rho \| \sigma)} \leq \mathrm{SC}_{(\mathcal{N}(\rho), \mathcal{N}(\sigma))}(\varepsilon, \beta) \leq \left\lceil 2 \ln\left(\frac{\sqrt{\beta(1-\beta)}}{\varepsilon} \right) \left(\frac{1}{\eta T(\rho, \sigma)} \right)^2 \right\rceil. \tag{8.78}$$

Proof. The proof follows by substituting $1 - \breve{\alpha}(\mathcal{N}_{p,\eta}) = \eta$ for the generalized damping channels ([14, Proposition C.1]) and $\eta_{\mathrm{Tr}}(\mathcal{N}_{p,\eta}) = \sqrt{\eta}$ (Lemma 12). For the lower bound, recall that the lower bounds in Proposition 43 are tighter with the use of $\eta_{\mathrm{Tr}}(\mathcal{N})$ instead of $1 - \alpha(\mathcal{N})$ since $\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha(\mathcal{N})$. □

Corollary 13 provides an example of a noisy channel for which the sample complexity of noisy hypothesis testing of the states ρ and σ can be finite with $\eta < 1$, whenever $\mathcal{N}_{p,\eta}(\rho) \neq \mathcal{N}_{p,\eta}(\sigma)$.

Remark 15 (Another Lower Bound for the special setting of GAD channel)**.** *It is possible to obtain a lower bound that inversely depends on $[T(\rho, \sigma)]^2$ as in Item 4 of Proposition 43 considering the minimum eigenvalues of the Choi operator of GAD channel:*

$$\lambda_{\min}\left(\Gamma^{\mathcal{N}_{p,\eta}} \right) = \frac{1+\eta}{2} - \sqrt{\eta + (1-\eta)^2 (1-2p)^2}. \tag{8.79}$$

This will lead to $\lambda_{\min}(\mathcal{N}(\rho)) \geq \lambda_{\min}\left(\Gamma^{\mathcal{N}_{p,\eta}} \right)$. This will be strictly greater than 1 when $p \notin \{0,1\}$. In particular for $p \in (0,1)$, we have that

$$\mathrm{SC}_{(\mathcal{N}_{p,\eta}(\rho), \mathcal{N}_{p,\eta}(\sigma))}(\varepsilon, \beta) \geq \frac{\lambda_{\min}\left(\Gamma^{\mathcal{N}_{p,\eta}} \right)}{4} \ln\left(\frac{\beta(1-\beta)}{\varepsilon(1-\varepsilon)} \right) \left(\frac{1}{\sqrt{\eta} T(\rho, \sigma)} \right)^2. \tag{8.80}$$

8.4 Fairness of Noisy Quantum Models

Decision-making models are prone to unfair treatments to individuals based on different characteristics including their gender, race, etc. With the development of quantum learning models, there is also a risk of implicit biases and unfair treatments to quantum data generated by various sources. Thus, it is important to quantify fairness in quantum learning models. Most current quantum devices are noisy, and so we focus here on how fair are noisy quantum channels when used in a learning model.

Quantum fairness refers to treating all input states equally, meaning that all pairs of input states that are close in some distance metric (e.g., close in normalized trace distance) should yield similar outcomes when processed by a quantum channel [94]. Define $\mathcal{A} \coloneqq \mathcal{M} \circ \mathcal{E}$, which is a quantum-to-classical channel in which a quantum channel $\mathcal{E}$ is followed by a measurement channel comprised of a POVM $(M_i)_{i \in \mathcal{O}}$. With that, quantum fairness is defined in [94] as follows.

Definition 15 ((γ,β)-Fairness [94])**.** *Let $\mathcal{A} = \mathcal{M} \circ \mathcal{E}$, and let $\hat{D}(\cdot\|\cdot)$ and $d(\cdot\|\cdot)$ be distance metrics on $\mathcal{D}(\mathcal{H})$ and $\mathcal{D}(\mathcal{O})$, respectively. Fix $0 < \gamma, \beta \leq 1$. Then the decision model $\mathcal{A}$ is (γ,β) fair if for all $\rho,\sigma \in \mathcal{D}(\mathcal{H})$ such that $\hat{D}(\rho\|\sigma) \leq \gamma$,*

$$d(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \beta. \tag{8.81}$$

We next show that noisy quantum channels provide fairness to quantum data when used as a quantum learning model.

Proposition 44 (Fairness Guarantee from Noisy Channels)**.** *Suppose that $\hat{D}(\rho\|\sigma) = \frac{1}{2}\|\rho-\sigma\|_1$ and $d(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) = \frac{1}{2}\sum_i |\mathrm{Tr}[M_i\mathcal{E}(\rho-\sigma)]|$. We have that $\mathcal{A} = \{\mathcal{E},\{M_i\}_{i\in\mathcal{O}}\}$ is (γ,β_γ)-fair for all $\rho,\sigma \in \mathcal{D}(\mathcal{H})$ such that $\hat{D}(\rho\|\sigma) \leq \gamma$, where*

$$\beta_\gamma \coloneqq \gamma\left(1-\alpha(\mathcal{A})\right), \tag{8.82}$$

with $\alpha(\mathcal{N})$ defined in (3.1).

Proof. The proof follows analogously to the proof of [67, Proposition 9] by bounding the contraction of trace distance under the noisy channels in terms of the Doeblin coefficient of the channel $\mathcal{A}$ (in particular by $1-\alpha(\mathcal{A})$). □

8.5 Mixing, Indistinguishability, and Decoupling Times of Quantum Processes

By definition, contraction coefficients capture how a measure of dissimilarity that satisfies the data-processing inequality, i.e., a generalized divergence $\mathbb{D}$, is altered after the action of a channel. It is for this reason that they place fundamental limits on tasks that require distinguishing the outputs of a channel. As quantum channels are fundamentally descriptions of dynamical systems, contraction coefficients also capture at what speed, if at all, dynamical systems converge to a fixed point or become independent of the input. In this section, we show that the quantum Doeblin coefficient can be used to bound the mixing time and acts as a quantum generalization of a 'coefficient of ergodicity.' Moreover, as the induced quantum Doeblin coefficient also captures these quantities, we introduce the 'decoupling time,' a quantum generalization of mixing time that is specifically captured by complete contraction coefficients.

For clarity, we begin by setting certain terminology for discrete-time quantum Markov chains. A quantum *homogeneous* Markov chain is a process where, at each time step, the same quantum channel $\mathcal{N}_{A\to A}$ occurs; i.e., for all $n \in \mathbb{N}$, the total process is as follows:

$$\mathcal{N}^n \coloneqq \bigcirc_{i\in[n]}\mathcal{N} = \underbrace{\mathcal{N}\circ\cdots\circ\mathcal{N}}_{n \text{ times}}. \tag{8.83}$$

We thus will talk of a 'homogeneous Markov chain $\mathcal{N}_{A\to A}$' as this completely specifies the Markov chain. A quantum *inhomogeneous* Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is identified by a sequence of quantum channels $(\mathcal{N}^i_{A_i\to A_{i+1}})_{i\in\mathbb{N}}$ where, for all $n \in \mathbb{N}$, $\mathcal{G}_n \coloneqq \bigcirc_{i\in[n]}\mathcal{N}^i$. Note that a homogeneous Markov chain is clearly a special case of an inhomogeneous one. These definitions directly generalize the notion of homogeneous and non-homogeneous Markov chains in the classical literature (see, e.g., [95,96]). Finally, given an inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ defined via $(\mathcal{N}^i_{A_i\to A_{i+1}})_{i\in\mathbb{N}}$, for all $k \leq m \in \mathbb{N}$, we let $\mathcal{G}_{k:m} \coloneqq \bigcirc_{i=k}^m \mathcal{N}^i$, which allows us to select sub-processes of the inhomogeneous Markov chain.

8.5.1 Mixing Processes and Mixing Times

We begin by defining notions of convergence, or 'mixing,' for quantum Markov chains. We then show how mixing times can be bounded by contraction coefficients.

Definition 16. *([97, 98]) We say a homogeneous Markov chain $\mathcal{N}_{A\to A}$ is mixing if it has a unique fixed point state $\omega \in \mathcal{D}(A)$ and*

$$\lim_{n\to\infty} \|\mathcal{N}^n(\rho) - \omega\|_1 = 0 \quad \forall \rho \in \mathcal{D}(A) \ . \tag{8.84}$$

We say it is strongly mixing if ω is furthermore full rank, which is also known as being primitive [11].

We remark that value of strongly mixing is that it often allows to establish particularly powerful claims about convergence to the fixed point; see, e.g., [11, Theorem 6.7], [98, Theorem 50 & Theorem 72], [99, Theorem 8]. We can generalize these definitions straightforwardly to inhomogeneous Markov chains.

Definition 17. *For an inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$, we say it is mixing if there exists $\omega \in \mathcal{D}(A)$ such that*

$$\lim_{n\to\infty} \|\mathcal{G}_n(\rho) - \omega\|_1 = 0 \quad \forall \rho \in \mathcal{D}(A) \ . \tag{8.85}$$

We say it is strongly mixing if ω is furthermore full rank.

Note that a mixing Markov chain destroys information about the initial state, but also accomplishes something stronger—namely, converges to a fixed state. We define the following weaker notions of mixing by considering only the former convergence property of destroying information about the initial state.

Definition 18. *An inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is weakly mixing if*

$$\lim_{n\to\infty} \|\mathcal{G}_n(\rho) - \mathcal{G}_n(\sigma)\|_1 = 0 \quad \forall \rho, \sigma \in \mathcal{D}(A) \ . \tag{8.86}$$

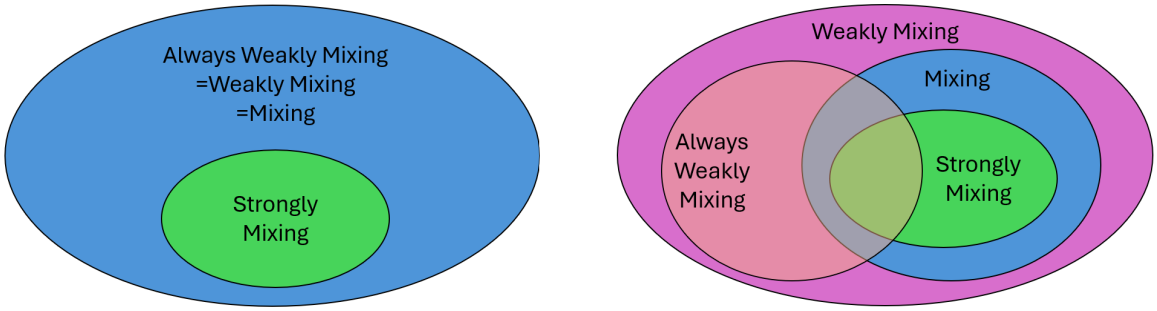
Moreover, we say an inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is always weakly mixing if for all $k \geq 1$,

$$\lim_{n\to\infty} \|\mathcal{G}_{k:n}(\rho) - \mathcal{G}_{k:n}(\sigma)\|_1 = 0 \quad \forall \rho, \sigma \in \mathcal{D}(A) \ , \tag{8.87}$$

where $\mathcal{G}_{k:n} \coloneqq \bigcirc_{i=k}^{n} \mathcal{N}^i$ is a sub-process of the inhomogeneous Markov chain.

We remark that being 'weakly mixing' matches the original *operational* notion of being 'weakly ergodic' [100], and being 'always weakly mixing' operationally aligns with the stricter notion of being 'weakly ergodic' introduced subsequently [101]. However, neither formally generalizes the notion of weakly ergodic, which has to do with the convergence of the entries of the transition matrix describing the Markov chain (see [95] for an in-depth study of weak ergodicity). In effect, the value of being 'always weakly mixing' is that it guarantees the process does not cease to be weakly mixing at some finite time.

We also remark that the definitions of mixing measure "strength" in terms of the convergence guarantees. We provide a complete characterization of the containments between the different notions of mixing (see Fig. 3 for a depiction).



(a) Homogeneous Markov Chains (b) Inhomogeneous Markov Chains

Figure 3: The diagram of containments of notions of mixing for (a) homogeneous and (b) inhomogeneous Markov chains as established in Proposition 45.

Proposition 45. *For homogeneous Markov chains, we have the following set containments for the types of mixing:*

$$\text{Strongly Mixing} \subsetneq \text{Mixing} = \text{Weakly Mixing} = \text{Always Weakly Mixing} \ . \tag{8.88}$$

For inhomogeneous Markov chains,

$$\begin{gathered} \text{Strongly Mixing} \subsetneq \text{Mixing} \subsetneq \text{Weakly Mixing} \\ \text{Always Weakly Mixing} \subsetneq \text{Weakly Mixing} \\ \text{Mixing} \not\subseteq \text{Always Weakly Mixing} \not\subseteq \text{Mixing} \\ \text{Strongly Mixing} \not\subseteq \text{Always Weakly Mixing} \not\subseteq \text{Strongly Mixing} \\ \text{Always Weakly Mixing} \cap \text{Mixing} \neq \emptyset \\ \text{Always Weakly Mixing} \cap \text{Strongly Mixing} \neq \emptyset \end{gathered} \ . \tag{8.89}$$

Proof. As the proof is long but straightforward, we relegate it to Section F.1. □

The above notions of mixing are all asymptotic. In practical applications, namely, sampling algorithms, one is generally interested in how many iterations of a mixing channel $\mathcal{N}$ with fixed point ω are necessary for every input to become δ-close to ω with respect to $\mathbb{D}$, which is known as a mixing time.

Definition 19. *Given a divergence $\mathbb{D}$, a channel $\mathcal{N}$ with a unique fixed point ω, and $\delta \geq 0$, the $\mathbb{D}$-mixing time is defined as*

$$t^{\mathbb{D}}_{\text{mix}}(\mathcal{N}, \delta) \coloneqq \inf\{n \in \mathbb{N} : \mathbb{D}(\mathcal{N}^n(\rho), \omega) \leq \delta \quad \forall \rho \in \mathcal{D}(A)\} \ , \tag{8.90}$$

which is equal to $+\infty$ when no such $n \in \mathbb{N}$ exists.

It is known that the mixing time is generally controlled by the input-dependent contraction coefficient (Definition 12) [12, 98, 99]. However, this can be further relaxed to the contraction coefficient and the complete contraction coefficient via the same proof method.

Proposition 46. *Let $\mathcal{N}_{A\to A}$ be a mixing channel with fixed point ω. Then for a divergence $\mathbb{D}$ and $\delta \in (0, C_{\mathbb{D}}(\omega)]$,*

$$t^{\mathbb{D}}_{\text{mix}}(\mathcal{N}, \delta) \leq \left\lceil \frac{\log(C_{\mathbb{D}}(\omega)/\delta)}{\log(1/\eta_{\mathbb{D}}(\mathcal{N}))} \right\rceil , \tag{8.91}$$

where $C_{\mathbb{D}}(\omega) \coloneqq \sup_{\rho\in\mathcal{D}(A)} \mathbb{D}(\rho\|\omega)$.

Proof. Consider that

$$\begin{aligned}\mathbb{D}(\mathcal{N}^n(\rho)\|\omega) = \mathbb{D}(\mathcal{N}^n(\rho)\|\mathcal{N}^n(\omega)) &\leq \eta_{\mathbb{D}}(\mathcal{N})\,\mathbb{D}(\mathcal{N}^{n-1}(\rho)\|\mathcal{N}^{n-1}(\omega)) \\ &\leq [\eta_{\mathbb{D}}(\mathcal{N})]^n\,\mathbb{D}(\rho\|\omega),\end{aligned} \tag{8.92}$$

where the equality is due to the assumption that ω is a fixed point, the first and second inequalities use the definition of contraction coefficient from (6.3), and the third inequality follows from Proposition 29. As it has to hold for all states, we apply the bound $\mathbb{D}(\rho\|\omega) \leq C_{\mathbb{D}}(\omega)$ in order to obtain an input-independent bound. Thus, it suffices to solve $[\eta_{\mathbb{D}}(\mathcal{N})]^n C_{\mathbb{D}}(\omega) \leq \delta$ for n. By algebraic manipulations, we find that $n \geq \log(C_{\mathbb{D}}(\omega)/\delta) \,/\, \log\big((\eta_{\mathbb{D}}(\mathcal{N}))^{-1}\big)$. Taking the ceiling function such that n is a natural number completes the proof. □

We remark that the above is somewhat limited, as often $C_{\mathbb{D}}(\omega) = +\infty$, although in many cases this can be handled if ω is full rank so that a reverse Pinsker inequality can be applied [98,99].[8] However, for the case of the trace distance, we obtain the following corollary, which needs no rank constraint.

Corollary 14. *Let $\mathcal{N}_{A\to A}$ be a mixing channel with fixed point ω. Then for all $\delta \in (0,1)$,*

$$t^T_{\mathrm{mix}}(\mathcal{N},\delta) \leq \left\lceil \frac{\log(\delta^{-1})}{\log([1-\alpha_I(\mathcal{N})]^{-1})} \right\rceil \leq \left\lceil \frac{\log(\delta^{-1})}{\log([1-\alpha(\mathcal{N})]^{-1})} \right\rceil . \tag{8.93}$$

Proof. This follows from applying Proposition 46, as $C_T(\omega) \leq 1$ for every state ω. Then one obtains the first inequality from (6.34) and the second then follows from Proposition 5. □

We remark that, while the final upper bound in (8.93) is likely generally loose, it can still be useful. First, Theorem 5 implies that for time-homogeneous Markov chains defined via a qubit channel, either the system does not contract or Corollary 14 obtains a non-trivial bound, so there exist clear use cases. Second, while concurrent work [99] finds an efficient method for obtaining bounds with input-dependent contraction coefficients, it requires the fixed point to be full rank. While under those conditions it is possible that there are cases where that work's method obtains tighter mixing times due to the hierarchy of quantum contraction coefficients, Corollary 14 applies for channels with non-full rank fixed points and thus includes cases that cannot be addressed by previous works. As a concrete example, any generalized amplitude damping channel with $p = 1$, $\mathcal{N}_{1,\eta}$, has the fixed point $|0\rangle\langle 0|$, which has unit rank, but $\alpha(\mathcal{N}_{1,\eta}) > 0$ unless $\eta = 0$ by Lemma 10.

8.5.2 Indistinguishability Times

As noted previously, it is known that mixing times are naturally addressed by input-dependent contraction coefficients. It is thus natural to ask what the contraction coefficient more directly captures. In the classical literature, contraction coefficients are identified as 'coefficients of ergodicity' [95] as they capture the finite-time notion of being weakly mixing (Definition 18).[9]

[8] Alternatively, if one is willing to make the stronger assumption that their sampling procedure is guaranteed to be initialized with a specific ρ, they could simply compute $\mathbb{D}(\rho\|\omega)$ to avoid the issue altogether.

[9] In fact, it is claimed in [102] that Doeblin announced his coefficient was relevant to a variant of weak ergodicity, but he never published the results. Ref. [102] conjectures this could be because Doeblin died prematurely during World War II.

We briefly and formally show this to be the case in the quantum setting, as it will help motivate decoupling times, which are introduced in Section 8.5.3.

Definition 20. *Let $\mathbb{D}$ be a divergence and $\mathcal{N}$ a quantum channel. Then the indistinguishability time is defined for $\delta \geq 0$ as*

$$t^{\mathbb{D}}_{\text{ind}}(\mathcal{N}, \delta) \coloneqq \inf\{n \in \mathbb{N} : \mathbb{D}(\mathcal{N}^n(\rho)\|\mathcal{N}^n(\sigma)) \leq \delta \quad \forall \rho, \sigma \in \mathcal{D}(A)\} \ , \tag{8.94}$$

which is equal to $+\infty$ when no such $n \in \mathbb{N}$ exists.

Note this definition is more demanding than the mixing time (Definition 19) as it does not require the reference state to be fixed.

Proposition 47. *For an arbitrary inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ defined by channels $(\mathcal{N}^i_{A_i\to A_{i+1}})_{i\in\mathbb{N}}$ and for all $\rho, \sigma \in \mathcal{D}(A_1)$,*

$$\mathbb{D}(\mathcal{G}_n(\rho)\|\mathcal{G}_n(\sigma)) \leq \mathbb{D}(\rho\|\sigma) \prod_{i\in[n]} \eta_{\mathbb{D}}(\mathcal{N}^i) \leq \mathbb{D}(\rho\|\sigma) \prod_{i\in[n]} \eta^c_{\mathbb{D}}(\mathcal{N}^i) \ . \tag{8.95}$$

In particular, for a homogeneous Markov chain $\mathcal{N}_{A\to A}$ and $\delta \in (0, C_{\mathbb{D}}]$,

$$t^{\mathbb{D}}_{\text{ind}}(\mathcal{N}, \delta) \leq \left\lceil \frac{\log(C_{\mathbb{D}}/\delta)}{\log(1/\eta_{\mathbb{D}}(\mathcal{N}))} \right\rceil \leq \left\lceil \frac{\log(C_{\mathbb{D}}/\delta)}{\log(1/\eta^c_{\mathbb{D}}(\mathcal{N}))} \right\rceil \ , \tag{8.96}$$

where $C_{\mathbb{D}} \coloneqq \sup_{\rho,\sigma\in\mathcal{D}(A)} \mathbb{D}(\rho\|\sigma)$.

Proof. By the same argument as in (8.92),

$$\mathbb{D}(\mathcal{G}_n(\rho)\|\mathcal{G}_n(\sigma)) \leq \mathbb{D}(\rho\|\sigma) \prod_{i\in[n]} \eta_{\mathbb{D}}(\mathcal{N}^i) \leq \mathbb{D}(\rho\|\sigma) \prod_{i\in[n]} \eta^c_{\mathbb{D}}(\mathcal{N}^i) \ . \tag{8.97}$$

In the case the Markov chain is homogeneous, the product simplifies to $[\eta_{\mathbb{D}}(\mathcal{N})]^n$ as before. Using $\mathbb{D}(\rho\|\sigma) \leq C_{\mathbb{D}}$ and following the same argument as in the proof of Proposition 46 completes the proof. □

While the issue is again that $C_{\mathbb{D}}$ can often be $+\infty$, in the case of trace distance we obtain the following tractable corollary involving the induced Doeblin coefficient α_I from Proposition 8.

Corollary 15. *For an arbitrary inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ defined by channels $\left(\mathcal{N}^i_{A_i\to A_{i+1}}\right)_{i\in\mathbb{N}}$ and for all $\rho, \sigma \in \mathcal{D}(A_1)$,*

$$T(\mathcal{G}_n(\rho)\|\mathcal{G}_n(\sigma)) \leq \prod_{i\in[n]} (1 - \alpha_I(\mathcal{N}^i)) \leq \prod_{i\in[n]} (1 - \alpha(\mathcal{N}^i)) \ . \tag{8.98}$$

Let $\mathcal{N}_{A\to A}$ be a time-homogeneous Markov chain. Then for all $\delta \in (0, 1)$,

$$t^T_{\text{ind}}(\mathcal{N}, \delta) \leq \left\lceil \frac{\log(\delta^{-1})}{\log([1 - \alpha_I(\mathcal{N})]^{-1})} \right\rceil \leq \left\lceil \frac{\log(\delta^{-1})}{\log([1 - \alpha(\mathcal{N})]^{-1})} \right\rceil \ . \tag{8.99}$$

Proof. The proof follows similarly to the proof of Corollary 14. □

We also observe the following.

Proposition 48. *An inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ defined by channels $\left(\mathcal{N}^i_{A_i\to A_{i+1}}\right)_{i\in\mathbb{N}}$ is always weakly mixing (Definition 18) if for all $k\in\mathbb{N}$, $\prod_{i=k}^{\infty}(1-\alpha_I(\mathcal{N}^i))=0$.*

Proof. If for all $k\in\mathbb{N}$, $\prod_{i=k}^{\infty}(1-\alpha_I(\mathcal{N}^i))=0$, then for all $\rho,\sigma\in\mathcal{D}(A)$,

$$\lim_{n\to\infty}\|\mathcal{G}_{k:n}(\rho)-\mathcal{G}_{k:n}(\sigma)\|_1\leq 2\cdot\lim_{n\to\infty}\prod_{i=k}^{n}(1-\alpha_I(\mathcal{N}^i))=0\ , \tag{8.100}$$

which is the definition of always weakly mixing. □

Thus, by Proposition 48, we conclude that (input-independent) contraction coefficients capture the notion of being weakly mixing.

8.5.3 Decoupling Times

So far, we have shown how to use the quantum Doeblin coefficient $\alpha(\mathcal{N})$ to bound the speed of mixing and indistinguishability. However, we have also shown that these are captured by the induced quantum Doeblin coefficient $\alpha_I(\mathcal{N})$. This would suggest that the quantum Doeblin coefficient should be relevant for an even stronger demand, in an analogous manner to how indistinguishability time is a stronger demand than mixing time and operationally separates input-dependent and input-independent contraction coefficients. Here we show this is indeed the case as the complete contraction coefficient, and thus quantum Doeblin coefficient, capture the notion of being 'weakly decoupling' and can bound the 'decoupling time.'

We begin with natural definitions that strengthen each version of a channel from being mixing to being decoupling.

Definition 21. *A homogeneous Markov chain $\mathcal{N}_{A\to A}$ that is mixing with fixed point ω is decoupling if for every finite Hilbert space R,*

$$\lim_{n\to\infty}\|(\mathrm{id}_R\otimes\mathcal{N}^n)(\rho)-\rho_R\otimes\omega_A\|_1=0\quad\forall\rho\in\mathcal{D}(R\otimes A)\ . \tag{8.101}$$

Moreover, we say it is strongly decoupling or completely primitive if ω is full rank.

It is easy to see that the above definition captures decoupling as it makes the R and A systems independent, which is the key idea in decoupling [103–105] (see also [106–108]). However, like being mixing, it includes the stronger assumption that it converges to a fixed state. Thus, we can define the weaker notion of weakly decoupling.

Definition 22. *An inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is weakly decoupling if for every finite Hilbert space R and all $\rho_{RA},\sigma_{RA}\in\mathcal{D}(R\otimes A)$ such that $\rho_R=\sigma_R$,*

$$\lim_{n\to\infty}\|(\mathrm{id}_R\otimes\mathcal{G}_n)(\rho)-(\mathrm{id}_R\otimes\mathcal{G}_n)(\sigma)\|_1=0\ . \tag{8.102}$$

Moreover, we say an inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is always weakly decoupling if for all $k\geq 1$, for every finite Hilbert space R, and for all $\rho_{RA},\sigma_{RA}\in\mathcal{D}(R\otimes A)$ such that $\rho_R=\sigma_R$,

$$\lim_{n\to\infty}\|(\mathrm{id}_R\otimes\mathcal{G}_{k:n})(\rho)-(\mathrm{id}_R\otimes\mathcal{G}_{k:n})(\sigma)\|_1=0\quad\forall\rho,\sigma\in\mathcal{D}(A)\ , \tag{8.103}$$

where $\mathcal{G}_{k:n}:=\bigcirc_{i=k}^{n}\mathcal{N}^i$ is a sub-process of the inhomogeneous Markov chain.

As in the mixing case, we remark that being 'always weakly decoupling' precludes that the process is decoupling only for some finite time.

We remark that the given definition for weakly decoupling does not make the decoupling explicit. Here, we show that it is equivalent to a definition that makes the decoupling explicit.

Proposition 49. *An inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ being weakly decoupling is equivalent to the following condition, holding for every finite Hilbert space R and all $\rho_{RA}, \sigma_{RA} \in \mathcal{D}(R\otimes A)$ such that $\rho_R = \sigma_R$,*

$$\lim_{n\to\infty} \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho_{RA}) - \sigma_R \otimes \mathcal{G}_n(\sigma_A)\|_1 = 0 \ , \tag{8.104}$$

and similarly for always weakly decoupling.

Proof. Eq. (8.102) implies (8.104) by letting ρ_{RA} be arbitrary and choosing $\sigma_{RA} = \sigma_R \otimes \sigma_A = \rho_R \otimes \sigma_A$. To prove the reverse implication, suppose that (8.104) holds for every finite Hilbert space R, and all $\hat{\rho}_{RA}, \hat{\sigma}_{RA} \in \mathcal{D}(R\otimes A)$ such that $\hat{\rho}_R = \hat{\sigma}_R$. Let $\rho_{RA}, \sigma_{RA} \in \mathcal{D}(R\otimes A)$ be such that $\rho_R = \sigma_R$. Then

$$\begin{aligned}
&\|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho_{RA}) - (\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma_{RA})\|_1 \\
&= \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho_{RA}) - \rho_R \otimes \mathcal{G}_n(\rho_A) + \rho_R \otimes \mathcal{G}_n(\rho_A) - (\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma_{RA})\|_1 && (8.105)\\
&\leq \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho_{RA}) - \rho_R \otimes \mathcal{G}_n(\rho_A)\|_1 + \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma_{RA}) - \rho_R \otimes \mathcal{G}_n(\rho_A)\|_1 \ , && (8.106)
\end{aligned}$$

where we have used the triangle inequality. Then,

$$\begin{aligned}
&\lim_{n\to\infty} \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho_{RA}) - (\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma_{RA})\|_1 \\
&\leq \lim_{n\to\infty} [\|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho_{RA}) - \rho_R \otimes \mathcal{G}_n(\rho_A)\|_1 + \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma_{RA}) - \rho_R \otimes \mathcal{G}_n(\rho_A)\|_1] && (8.107)\\
&= 0 + 0 \ , && (8.108)
\end{aligned}$$

where we have used the assumption that the limit exists for each term individually and is equal to zero (in the first case by letting $\hat{\rho} = \rho_{RA}$ and $\hat{\sigma} = \rho_R \otimes \rho_A$ and in the second case by letting $\hat{\rho} = \sigma_{RA}$ and $\hat{\sigma}_{RA} = \rho_R \otimes \rho_A$). The same argument works for always weakly decoupling. This completes the proof. □

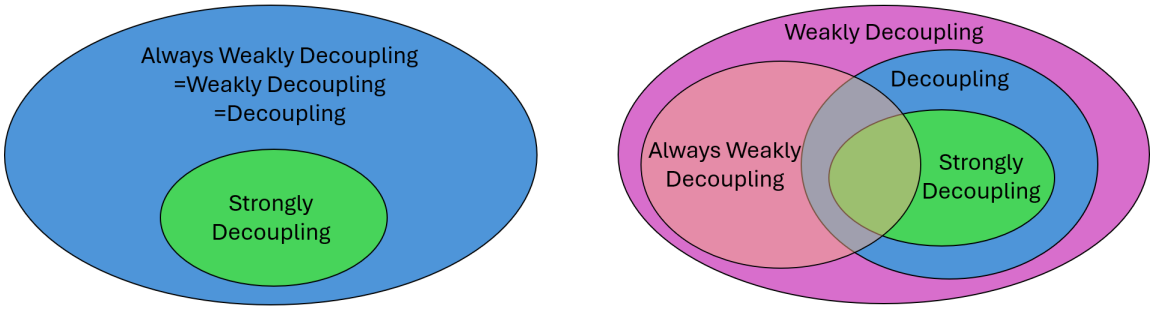
We end our discussion on the notions of decoupling by establishing the equivalent of Proposition 45 for various definitions of decoupling, as shown in Fig. 4.

Proposition 50. *For homogeneous Markov chains, we have the following set containments of the types of decoupling:*

$$\text{Strongly Decoupling} \subsetneq \text{Decoupling} = \text{Weakly Decoupling} = \text{Always Weakly Decoupling} \ .$$

For inhomogeneous Markov chains,

$$\begin{gathered}
\text{Strongly Decoupling} \subsetneq \text{Decoupling} = \text{Weakly Decoupling} \\
\text{Always Weakly Decoupling} \subsetneq \text{Weakly Decoupling} \\
\text{Decoupling} \not\subseteq \text{Always Weakly Decoupling} \not\subseteq \text{Decoupling} \\
\text{Strongly Decoupling} \not\subseteq \text{Always Weakly Decoupling} \not\subseteq \text{Strongly Decoupling} \\
\text{Always Weakly Decoupling} \cap \text{Decoupling} \neq \emptyset \\
\text{Always Weakly Decoupling} \cap \text{Strongly Decoupling} \neq \emptyset
\end{gathered} \ . \tag{8.109}$$



(a) Homogeneous Markov Chains (b) Inhomogeneous Markov Chains

Figure 4: The diagram of containments of notions of decoupling for (a) homogeneous and (b) inhomogeneous Markov chains as established in Proposition 50. As may be seen by comparing to Fig. 3, mixing and decoupling have analogous containments.

Proof. See Section F.2. □

With the taxonomy of decoupling channels addressed, we define the notion of decoupling time and decoupled indistinguishability time for a divergence $\mathbb{D}$.

Definition 23. *Let $\mathbb{D}$ be a divergence, and let $\mathcal{N}_{A\to A}$ be a quantum channel with unique fixed point $\omega \in \mathcal{D}(A)$. Then for all $\delta \geq 0$, the decoupling time of the homogeneous Markov chain $\mathcal{N}$ is defined as*

$$t^{\mathbb{D}}_{\mathrm{dec}}(\mathcal{N},\delta) \coloneqq \inf\left\{ n\in\mathbb{N} : \begin{array}{c} \mathbb{D}((\mathrm{id}_R\otimes\mathcal{N}^n)(\rho_{RA})\|\rho_R\otimes\omega_A)\leq\delta\,, \\ \forall R\,,\,\forall \rho_{RA}\in\mathcal{D}(R\otimes A) \end{array} \right\}\,, \tag{8.110}$$

which is equal to $+\infty$ when no such $n\in\mathbb{N}$ exists.

Similarly, the decoupled indistinguishability time is defined as

$$t^{\mathbb{D}}_{\text{d-i}}(\mathcal{N},\delta) \coloneqq \inf\left\{ n\in\mathbb{N} : \begin{array}{c} \mathbb{D}((\mathrm{id}_R\otimes\mathcal{N}^n)(\rho_{RA})\|(\mathrm{id}_R\otimes\mathcal{N}^n)(\sigma_{RA}))\leq\delta\,, \\ \forall R\,,\,\forall \rho_{RA},\sigma_{RA}\in\mathcal{D}(R\otimes A)\text{ s.t. }\rho_R=\sigma_R \end{array} \right\}\,, \tag{8.111}$$

which is equal to $+\infty$ when no such $n\in\mathbb{N}$ exists.

Note these conditions are more demanding than mixing time or indistinguishability time, respectively.

We then obtain the following bounds in terms of complete contraction coefficients. We begin with the decoupling time.

Proposition 51. *Let $\mathbb{D}$ be a divergence, and let $\mathcal{N}_{A\to A}$ be a quantum channel with unique fixed state $\omega\in\mathcal{D}(A)$. Then, for all $\delta\in\left(0,\hat{C}_{\mathbb{D}}(\omega)\right]$,*

$$t^{\mathbb{D}}_{\mathrm{dec}}(\mathcal{N},\delta) \leq \left\lceil \frac{\log\left(\hat{C}_{\mathbb{D}}(\omega)/\delta\right)}{\log(1/\eta^c_{\mathbb{D}}(\mathcal{N},\omega))} \right\rceil \leq \left\lceil \frac{\log\left(\hat{C}_{\mathbb{D}}(\omega)/\delta\right)}{\log(1/\eta^c_{\mathbb{D}}(\mathcal{N}))} \right\rceil, \tag{8.112}$$

where $\hat{C}_{\mathbb{D}}(\omega)\coloneqq\sup_{\rho\in\mathcal{D}(RA)}\mathbb{D}(\rho_{RA}\|\rho_R\otimes\omega_A)$.

Proof. The proof is effectively the same as those for mixing and indistinguishability times, but we provide it here, as the R system is now relevant. Let R be an arbitrary finite-dimensional Hilbert space, and let $\rho_{RA} \in \mathcal{D}(R \otimes A)$. Then,

$$\begin{aligned}
\mathbb{D}((\mathrm{id}_R \otimes \mathcal{N}^n)(\rho) \| \rho_R \otimes \omega) &= \mathbb{D}((\mathrm{id}_R \otimes \mathcal{N}^n)(\rho) \| (\mathrm{id}_R \otimes \mathcal{N}^n)(\rho_R \otimes \omega)) && (8.113)\\
&\leq \eta_{\mathbb{D}}^c(\mathcal{N}, \omega)\, \mathbb{D}((\mathrm{id}_R \otimes \mathcal{N}^{n-1})(\rho) \| (\mathrm{id}_R \otimes \mathcal{N}^{n-1})(\rho_R \otimes \omega_A)) && (8.114)\\
&\leq [\eta_{\mathbb{D}}^c(\mathcal{N}, \omega)]^n\, \mathbb{D}(\rho_{RA} \| \rho_R \otimes \omega_A)\ , && (8.115)
\end{aligned}$$

where the equality follows from the assumption that ω is a fixed point of $\mathcal{N}$, the first inequality follows because ρ_{RA} and $\rho_R \otimes \omega_A$ have the same marginal on the R system so one can use the definition of input-dependent complete contraction coefficient (Definition 12) and the final inequality follows from applying the same argument $n-1$ times. Then, we use the upper bound $\mathbb{D}(\rho_{RA} \| \rho_R \otimes \omega_A) \leq \hat{C}_{\mathbb{D}}(\omega)$. As this holds for an arbitrary ρ_{RA}, to obtain a bound on decoupling time, it suffices to solve the inequality $[\eta_{\mathbb{D}}^c(\mathcal{N}, \omega)]^n\, \hat{C}_{\mathbb{D}}(\omega) \leq \delta$ for n. Using properties of the logarithm to do so, similarly as done in the proof of Proposition 46, completes the proof for the input-dependent complete contraction coefficient, and then one may relax it to the input-independent variant. □

For the case of the trace distance, we obtain the following corollary:

Corollary 16. *Let $\mathcal{N}_{A\to A}$ be a quantum channel with unique fixed point $\omega \in \mathcal{D}(A)$. Then*

$$t_{\mathrm{dec}}^T(\mathcal{N}, \delta) \leq \left\lceil \frac{\log(\delta^{-1})}{\log([1 - \alpha(\mathcal{N})]^{-1})} \right\rceil\ . \tag{8.116}$$

Proof. This follows from Proposition 51, that $\hat{C}_T(\omega) \leq 1$, and Proposition 30. □

Note that this means not only have we introduced the decoupling time, but we have also established a method for being able to efficiently compute an upper bound on it.

We now consider the decoupled indistinguishability time.

Proposition 52. *For an inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ defined by channels $(\mathcal{N}^i_{A_i\to A_{i+1}})_{i\in\mathbb{N}}$, for every finite dimensional Hilbert space R and $\rho, \sigma \in \mathcal{D}(R \otimes A_1)$ such that $\rho_R = \sigma_R$, the following inequality holds:*

$$\mathbb{D}((\mathrm{id}_R \otimes \mathcal{G}_n)(\rho) \| (\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma)) \leq \mathbb{D}(\rho \| \sigma) \prod_{i\in[n]} \eta_{\mathbb{D}}^c(\mathcal{N}^i)\ . \tag{8.117}$$

In particular, for a homogeneous Markov chain $\mathcal{N}_{A\to A}$ and $\delta \in (0, \hat{C}_{\mathbb{D}}]$,

$$t_{\text{d-i}}^{\mathbb{D}}(\mathcal{N}, \delta) \leq \left\lceil \frac{\log\left(\hat{C}_{\mathbb{D}}/\delta\right)}{\log(1/\eta_{\mathbb{D}}^c(\mathcal{N}))} \right\rceil\ , \tag{8.118}$$

where $\hat{C}_{\mathbb{D}} \coloneqq \sup_{\rho,\sigma\in\mathcal{D}(R\otimes A):\rho_R=\sigma_R} \mathbb{D}(\rho\|\sigma)$.

Proof. The proof is the same as that for Proposition 47, except one needs to handle the R system in the same manner as in Proposition 51, which results in using the complete contraction coefficient. □

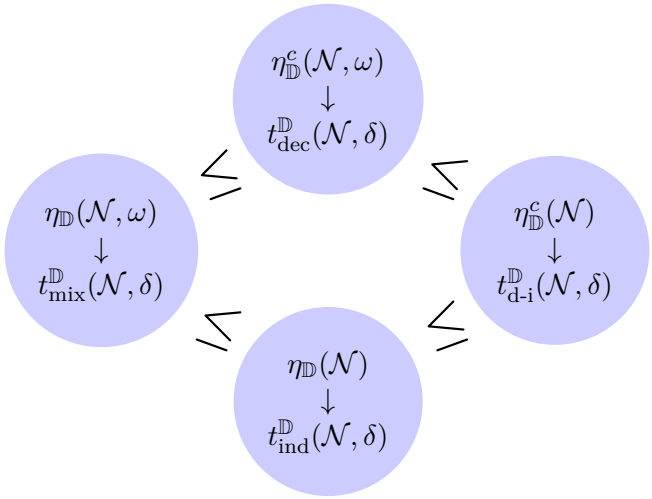


Figure 5: Depiction of relations between different tasks related to the convergence of quantum dynamical systems and the contraction coefficient, as established in this section. The $\leq$ sign denotes the inequality between the operational quantities as well as the inequality between the corresponding contraction coefficients. The symbol $\downarrow$ denotes that the given contraction coefficient controls the finite-time convergence of the corresponding task. Here, $\eta_{\mathbb{D}}(\mathcal{N},\omega), \eta_{\mathbb{D}}(\mathcal{N}), \eta^c_{\mathbb{D}}(\mathcal{N},\omega)$, and $\eta^c_{\mathbb{D}}(\mathcal{N})$ are defined in (6.6), (6.3), (6.7), and (6.4), respectively. Also, $t^{\mathbb{D}}_{\mathsf{mix}}(\mathcal{N},\delta)$, $t^{\mathbb{D}}_{\mathsf{ind}}(\mathcal{N},\delta)$, $t^{\mathbb{D}}_{\mathsf{dec}}(\mathcal{N},\delta)$, and $t^{\mathbb{D}}_{\text{d-i}}(\mathcal{N},\delta)$ are defined in Definition 19, Definition 20, (23), and (8.111), respectively.

Again, we obtain various special cases for the trace distance. We state the ones that we believe should be of most interest. The proofs follow by the same arguments that we used for other cases and are thus omitted.

Corollary 17. *Let $\mathcal{N}_{A\to A}$ be define a time-homogeneous Markov chain. Then for all $\delta \in (0,1)$,*

$$t^T_{\text{d-i}}(\mathcal{N},\delta) \leq \left\lceil \frac{\log(\delta^{-1})}{\log([1-\alpha(\mathcal{N})]^{-1})} \right\rceil . \tag{8.119}$$

Proposition 53. *An inhomogeneous Markov chain $(\mathcal{G}_n)_{n\in\mathbb{N}}$ defined by channels $(\mathcal{N}^i_{A\to A_{i+1}})_{i\in\mathbb{N}}$ is always weakly decoupling (Definition 22) if for all $k\in\mathbb{N}$, $\prod_{i=k}^{\infty}(1-\alpha_I(\mathcal{N}^i))=0$.*

In summary, what we have seen in this section is that for every divergence $\mathbb{D}$, channel $\mathcal{N}$, and fixed $\delta \geq 0$, we have the ordering on operational tasks and identification of which contraction coefficient controls them (see Fig. 5). Moreover, we have shown that in the case $\mathbb{D} = T$, we can control the finite-time convergence of each task using the quantum Doeblin coefficient, which is efficiently computable, and this is nontrivial as long as $\alpha(\mathcal{N}) < 1$.

9 Conclusion

9.1 Summary

In this paper, we comprehensively studied quantum Doeblin coefficients, two of which had already been defined and considered [11, 14] and others which we defined here. We showed that

these coefficients satisfy several desirable properties including efficient computability using SDPs, concatenation, and, for one of them, multiplicativity. We also developed various interpretations of the two main quantum Doeblin coefficients (i.e., $\alpha(\mathcal{N})$ and $\alpha_I(\mathcal{N})$), including as minimal singlet fractions, exclusion values, reverse max-mutual and oveloH informations, reverse robustnesses, and hypothesis testing reverse mutual and oveloH informations. Furthermore, we studied the complete contraction coefficient of quantum divergences and showed how Doeblin coefficients provide efficient computable bounds on them. We also analytically characterized the Doeblin coefficients of various practically relevant channels and extensively studied them for qubit channels, while utilizing the structure of qubit channels in terms of the Stokes parameterization. Finally, we showcased how the technical tools developed in this work find use in applications ranging from identifying the limitations on the trainability of quantum learning algorithms using parameterized quantum circuits when the circuits are noisy, the overhead of error mitigation protocols to learn expected values of observables with noisy data samples, the impact of noise in the hypothesis testing of quantum states, the fairness offered by noisy learning algorithms, and mixing, indistinguishability, and decoupling times of time-varying channels. In all of these applications, we provided efficiently computable bounds while expanding what was previously known, in terms of the circuit architectures, noise models, and so on.

9.2 Open Questions

There are various open questions that arise from our work. Here we list them without any particular ordering.

1. In Proposition 31, we showed that $\eta_\gamma^c(\mathcal{N}) \leq 1-\alpha_+(\mathcal{N})$. It is an interesting open question to determine whether $\eta_\gamma^c(\mathcal{N}) \leq 1 - \alpha(\mathcal{N})$. Due to the Hermitian operator X present in the optimization of $\alpha(\mathcal{N})$, our proof technique in establishing the former cannot be used as is in the aforementioned problem.

2. In this work, we showed that some Doeblin coefficients satisfy multiplicativity while others do not. It would be interesting to explore other possible Doeblin coefficients that might satisfy these desirable properties.

3. We analysed Doeblin coefficients of various classes of channels and special channels that are of practical relevance. Studying how the symmetry of a given channel helps in obtaining analytical expressions for their Doeblin coefficients and their tensor product is of relevance (i.e., $\mathcal{N}$ and $\mathcal{N}^{\otimes n}$ for $n \in \mathbb{N}$).

4. All applications of the quantum Doeblin coefficients in this work are related to the fact that they provide upper bounds on trace-distance contraction coefficients. It would be interesting to find more direct applications of the quantum Doeblin coefficients, in the sense that they do not rely upon this kind of relation to trace-distance contraction coefficients. A recent such example was put forward in [109], in which the Doeblin information in Proposition 11 plays a role in characterizing the retrocausal capacity of a quantum channel.

Acknowledgements

We acknowledge helpful discussions with Nana Liu on noise-induced barren plateaus and the parameter shift rule. We also thank Hemant Mishra for helpful discussions. Part of this work was completed during the conference "Beyond IID in Information Theory," held at the University of Illinois Urbana-Champaign from July 29 to August 2, 2024, and supported by NSF Grant No. 2409823. Additional parts were completed during a visit of CH to the School of Electrical and Computer Engineering at Cornell University, whose admin staff he thanks for hospitality, and at the 2025 Quantum Resource Workshop in Jeju, Korea. We thank the organizers of both events, Beyond IID and Quantum Resources, for organizing these events and including time for research discussions during them.

TN and MMW acknowledge support from NSF Grant No. 2329662. IG thanks Eric Chitambar for providing a copy of [49]. IG is supported by the Ministry of Education, Singapore, through grant T2EP20124-0005. CH received funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 550206990. TN acknowledges support from the Dieter Schwarz Exchange Programme on Quantum Communication and Security at the Centre for Quantum Technologies, National University of Singapore, Singapore, during her research visit, when the second version of the arXiv post was finalized.

References

[1] W. Doeblin. "Sur les proprietes asymptotiques de mouvement régis par certains types de chaines simples". Bulletin mathematique de la Societe Roumaine des Sciences **39**, 57–115 (1937). url: http://www.jstor.org/stable/43769809.

[2] Anuran Makur and Japneet Singh. "Doeblin coefficients and related measures". IEEE Transactions on Information Theory **70**, 4667–4692 (2024).

[3] Hao Chen, Jiacheng Tang, and Abhishek Gupta. "Change detection of Markov kernels with unknown pre and post change kernel". In 2022 IEEE 61st Conference on Decision and Control (CDC). Pages 4814–4820. (2022).

[4] Vrettos Moulos. "Finite-time analysis of round-robin Kullback-Leibler upper confidence bounds for optimal adaptive allocation with multiple plays and Markovian rewards". In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems. Volume 33, pages 7863–7874. Curran Associates, Inc. (2020). url: https://proceedings.neurips.cc/paper_files/paper/2020/file/597c7b407a02cc0a92167e7a371eca25-Paper.pdf.

[5] Jeffrey S. Rosenthal. "Minorization conditions and convergence rates for Markov chain Monte Carlo". Journal of the American Statistical Association **90**, 558–566 (1995).

[6] Jeffrey M. Alden and Robert L. Smith. "Rolling horizon procedures in nonhomogeneous Markov decision processes". Operations Research **40**, S183–S194 (1992).

[7] Jacob Steinhardt and Percy Liang. "Learning fast-mixing models for structured prediction". In Francis Bach and David Blei, editors, Proceedings of the 32nd International Conference on Machine Learning. Volume 37 of Proceedings of Machine Learning Research, pages 1063–1072. Lille, France (2015). PMLR. url: https://proceedings.mlr.press/v37/steinhardtb15.html.

[8] Somshubhro Bandyopadhyay, Rahul Jain, Jonathan Oppenheim, and Christopher Perry. "Conclusive exclusion of quantum states". Physical Review A **89**, 022336 (2014).

[9] Hemant K. Mishra, Michael Nussbaum, and Mark M. Wilde. "On the optimal error exponents for classical and quantum antidistinguishability". Letters in Mathematical Physics **114**, 76 (2024).

[10] R. L. Dobrushin. "Central limit theorem for nonstationary Markov chains. I". Theory of Probability & Its Applications **1**, 65–80 (1956).

[11] Michael M. Wolf. "Quantum channels and operations—guided tour" (2012). Available at `https://mediatum.ub.tum.de/doc/1701036/document.pdf`.

[12] Maxim Raginsky. "Strong data processing inequalities and Φ-Sobolev inequalities for discrete channels". IEEE Transactions on Information Theory **62**, 3355–3389 (2016).

[13] Stephen Chestnut and Manuel E. Lladser. "Occupancy distributions in Markov chains via Doeblin's ergodicity coefficient". Discrete Mathematics & Theoretical Computer Science (2010).

[14] Christoph Hirche. "Quantum Doeblin coefficients: A simple upper bound on contraction coefficients" (2024). arXiv:2405.00105v2.

[15] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph. "On the reality of the quantum state". Nature Physics **8**, 475–478 (2012).

[16] Samson Wang, Enrico Fontana, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. "Noise-induced barren plateaus in variational quantum algorithms". Nature Communications **12**, 6961 (2021).

[17] Marco Schumann, Frank K. Wilhelm, and Alessandro Ciani. "Emergence of noise-induced barren plateaus in arbitrary layered noise models". Quantum Science and Technology **9**, 045019 (2024).

[18] Phattharaporn Singkanipa and Daniel A. Lidar. "Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and limit sets". Quantum **9**, 1617 (2025).

[19] Antonio Anna Mele, Armando Angrisani, Soumik Ghosh, Sumeet Khatri, Jens Eisert, Daniel Stilck França, and Yihui Quek. "Noise-induced shallow circuits and absence of barren plateaus" (2024). arXiv:2403.13927v2.

[20] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. "Quantum error mitigation". Reviews of Modern Physics **95**, 045005 (2023).

[21] Ryuji Takagi, Hiroyasu Tajima, and Mile Gu. "Universal sampling lower bounds for quantum error mitigation". Physical Review Letters **131**, 210602 (2023).

[22] Yihui Quek, Daniel Stilck França, Sumeet Khatri, Johannes Jakob Meyer, and Jens Eisert. "Exponentially tighter bounds on limitations of quantum error mitigation". Nature Physics **20**, 1648–1658 (2024).

[23] Mark M. Wilde. "Quantum information theory". Cambridge University Press. (2017). Second edition.

[24] Masahito Hayashi. "Quantum information theory: Mathematical foundation". Springer. (2017). Second edition.

[25] John Watrous. "The theory of quantum information". Cambridge University Press. (2018).

[26] Alexander S. Holevo. "Quantum systems, channels, information: A mathematical introduction". Volume 16. Walter de Gruyter. (2019). Second edition.

[27] Sumeet Khatri and Mark M. Wilde. "Principles of quantum communication theory: A modern approach" (2024). arXiv:2011.04672v2.

[28] Naresh Sharma and Naqueeb Ahmad Warsi. "On the strong converses for the quantum channel capacity theorems" (2012). arXiv:1205.1712.

[29] Francesco Buscemi and Nilanjana Datta. "The quantum capacity of channels with arbitrarily correlated noise". IEEE Transactions on Information Theory **56**, 1447–1460 (2010). arXiv:0902.0158.

[30] Fernando G. S. L. Brandao and Nilanjana Datta. "One-shot rates for entanglement manipulation under non-entangling maps". IEEE Transactions on Information Theory **57**, 1754–1760 (2011). arXiv:0905.2673.

[31] Ligong Wang and Renato Renner. "One-shot classical-quantum capacity and hypothesis testing". Physical Review Letters **108**, 200501 (2012). arXiv:1007.5456.

[32] Kaiyuan Ji, Hemant K. Mishra, Milán Mosonyi, and Mark M. Wilde. "Barycentric bounds on the error exponents of quantum hypothesis exclusion". IEEE Transactions on Information Theory **72**, 2391–2423 (2026). arXiv:2407.13728.

[33] Xin Wang and Mark M. Wilde. "Resource theory of asymmetric distinguishability". Physical Review Research **1**, 033170 (2019).

[34] Nilanjana Datta. "Min- and max-relative entropies and a new entanglement monotone". IEEE Transactions on Information Theory **55**, 2816–2826 (2009).

[35] John Watrous. "Advanced topics in quantum information theory, lecture 1: Conic programming". `https://web.archive.org/web/20240429005322/https://johnwatrous.com/wp-content/uploads/2023/08/QIT-notes.01.pdf` (2020).

[36] Stephen P. Boyd and Lieven Vandenberghe. "Convex optimization". Cambridge University Press. (2004).

[37] Nathaniel Johnston. "Norms and cones in the theory of quantum entanglement". PhD thesis. University of Guelph. (2012). arXiv:1207.1479.

[38] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. "Complete family of separability criteria". Physical Review A **69**, 022308 (2004).

[39] Caleb McIrvin, Ankith Mohan, and Jamie Sikora. "Quantum state exclusion through offset measurement". Physical Review A **110**, 042211 (2024).

[40] Vojtěch Havlíček and Jonathan Barrett. "Simple communication complexity separation from quantum state antidistinguishability". Physical Review Research **2**, 013326 (2020).

[41] Matthew Leifer and Cristhiano Duarte. "Noncontextuality inequalities from antidistinguishability". Physical Review A **101**, 062113 (2020).

[42] Roope Uola, Tom Bullock, Tristan Kraft, Juha-Pekka Pellonpää, and Nicolas Brunner. "All quantum resources provide an advantage in exclusion tasks". Physical Review Letters **125**, 110402 (2020).

[43] Vincent Russo and Jamie Sikora. "Inner products of pure states and their antidistinguishability". Physical Review A **107**, L030202 (2023).

[44] Nathaniel Johnston, Vincent Russo, and Jamie Sikora. "Tight bounds for antidistinguishability and circulant sets of pure quantum states". Quantum **9**, 1622 (2025). arXiv:2311.17047.

[45] Armando Angrisani, Mina Doosti, and Elham Kashefi. "Differential privacy amplification in quantum and quantum-inspired algorithms" (2023). arXiv:2203.03604.

[46] Xin Wang, Kun Fang, and Runyao Duan. "Semidefinite programming converse bounds for quantum communication". IEEE Transactions on Information Theory **65**, 2583–2592 (2019). arXiv:1709.00200.

[47] Xin Wang. "Semidefinite optimization for quantum information". PhD thesis. University of Technology Sydney (Australia). (2018). url: https://www.proquest.com/openview/01f199b857646e73d07ef7e3df9a9fb5/.

[48] Eric Chitambar, Ian George, Brian Doolittle, and Marius Junge. "The communication value of a quantum channel". IEEE Transactions on Information Theory **69**, 1660–1679 (2022).

[49] Eric Chitambar. "Part I: Fundamental principles of quantum information processing" (2021). Lecture Notes, Unpublished.

[50] Ian George and Eric Chitambar. "Cone-restricted information theory". Journal of Physics A: Mathematical and Theoretical **57**, 265302 (2024).

[51] Robert König, Renato Renner, and Christian Schaffner. "The operational meaning of min-and max-entropy". IEEE Transactions on Information Theory **55**, 4337–4347 (2009).

[52] Leonid Gurvits and Howard Barnum. "Largest separable balls around the maximally mixed bipartite quantum state". Physical Review A **66**, 062311 (2002).

[53] Maciej Lewenstein and Anna Sanpera. "Separability and entanglement of composite quantum systems". Physical Review Letters **80**, 2261–2264 (1998).

[54] Bartosz Regula. "Probabilistic transformations of quantum resources". Physical Review Letters **128**, 110505 (2022).

[55] Bartosz Regula. "Tight constraints on probabilistic convertibility of quantum states". Quantum **6**, 817 (2022).

[56] Michal Horodecki, Peter W. Shor, and Mary Beth Ruskai. "Entanglement breaking channels". Reviews in Mathematical Physics **15**, 629–641 (2003). arXiv:quant-ph/0302031.

[57] Shalev Ben-David and Eric Blais. "A new minimax theorem for randomized algorithms". Journal of the ACM **70**, 1–58 (2023). arXiv:2002.10802.

[58] Peter W Shor. "Additivity of the classical capacity of entanglement-breaking quantum channels". Journal of Mathematical Physics **43**, 4334–4340 (2002).

[59] Christopher King. "Maximal p-norms of entanglement breaking channels". Quantum Information and Computation **3**, 186–190 (2003). arXiv:quant-ph/0212057.

[60] Yury Polyanskiy and Sergio Verdú. "Arimoto channel coding converse and Rényi divergence". In 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton). Pages 1327–1333. (2010).

[61] Fumio Hiai and Mary Beth Ruskai. "Contraction coefficients for noisy quantum channels". Journal of Mathematical Physics **57**, 015211 (2015).

[62] Christoph Hirche, Cambyse Rouzé, and Daniel Stilck França. "On contraction coefficients, partial orders and approximation of capacities for quantum channels". Quantum **6**, 862 (2022).

[63] Li Gao and Cambyse Rouzé. "Complete entropic inequalities for quantum Markov chains". Archive for Rational Mechanics and Analysis **245**, 183–238 (2022).

[64] Andrew Lesniewski and Mary Beth Ruskai. "Monotone Riemannian metrics and relative entropy on noncommutative probability spaces". Journal of Mathematical Physics **40**, 5702–5724 (1999).

[65] Christoph Hirche and Marco Tomamichel. "Quantum Rényi and f-divergences from integral representations". Communications in Mathematical Physics **405**, 208 (2024).

[66] Christoph Hirche, Cambyse Rouzé, and Daniel Stilck França. "Quantum differential privacy: An information theory perspective". IEEE Transactions on Information Theory **69**, 5771–5787 (2023). arXiv:2202.10717.

[67] Theshani Nuradha and Mark M. Wilde. "Contraction of private quantum channels and private quantum hypothesis testing". IEEE Transactions on Information Theory **71**, 1851–1873 (2025).

[68] H. Yuen, R. Kennedy, and M. Lax. "Optimum testing of multiple hypotheses in quantum detection theory". IEEE Transactions on Information Theory **21**, 125–134 (1975).

[69] Omar Fawzi, Ala Shayeghi, and Hoang Ta. "A hierarchy of efficient bounds on quantum capacities exploiting symmetry". IEEE Transactions on Information Theory **68**, 7346–7360 (2022).

[70] Vishal Singh, Theshani Nuradha, and Mark M. Wilde. "Extendible quantum measurements and limitations on classical communication". In 2025 IEEE International Symposium on Information Theory (ISIT). Page 1–6. IEEE (2025). arXiv:2412.18556v2.

[71] Navneeth Ramakrishnan, Raban Iten, Volkher B. Scholz, and Mario Berta. "Computing quantum channel capacities". IEEE Transactions on Information Theory **67**, 946–960 (2020).

[72] Nicholas Laracuente and Graeme Smith. "Information fragility or robustness under quantum channels" (2023). arXiv:2312.17450.

[73] Paula Belzig, Li Gao, Graeme Smith, and Peixue Wu. "Reverse-type data processing inequality". Communications in Mathematical Physics **406**, 295 (2025). arXiv:2411.19890.

[74] Christopher King and Mary Beth Ruskai. "Minimal entropy of states emerging from noisy quantum channels". IEEE Transactions on Information Theory **47**, 192–209 (2001). arXiv:quant-ph/9911079.

[75] Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd. "The quantum Wasserstein distance of order 1". IEEE Transactions on Information Theory **67**, 6627–6643 (2021).

[76] John Watrous. "Notes on super-operator norms induced by Schatten norms". Quantum Information & Computation **5**, 58–68 (2005). arXiv:quant-ph/0411077.

[77] Sumeet Khatri, Kunal Sharma, and Mark M. Wilde. "Information-theoretic aspects of the generalized amplitude-damping channel". Physical Review A **102**, 012401 (2020).

[78] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. "Variational quantum algorithms". Nature Reviews Physics **3**, 625–644 (2021).

[79] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. "Barren plateaus in quantum neural network training landscapes". Nature Communications **9**, 4812 (2018).

[80] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J Coles, Lukasz Cincio, Jarrod R McClean, Zoë Holmes, and Marco Cerezo. "Barren plateaus in variational quantum computing". Nature Reviews Physics **7**, 174–189 (2025). arXiv:2405.00781.

[81] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang-Pu Sun. "Hybrid quantum-classical approach to quantum optimal control". Physical Review Letters **118**, 150503 (2017).

[82] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. "Quantum circuit learning". Physical Review A **98**, 032309 (2018).

[83] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. "Evaluating analytic gradients on quantum hardware". Physical Review A **99**, 032331 (2019).

[84] Gavin E. Crooks. "Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition" (2019). arXiv:1905.13311.

[85] Manuel S. Rudolph, Jacob Miller, Danial Motlagh, Jing Chen, Atithi Acharya, and Alejandro Perdomo-Ortiz. "Synergistic pretraining of parametrized quantum circuits via tensor networks". Nature Communications **14**, 8367 (2023).

[86] Carl W. Helstrom. "Detection theory and quantum mechanics". Information and Control **10**, 254–291 (1967).

[87] Alexander S. Holevo. "Statistical decision theory for quantum systems". Journal of Multivariate Analysis **3**, 337–394 (1973).

[88] Hao-Chung Cheng, Nilanjana Datta, Nana Liu, Theshani Nuradha, Robert Salzmann, and Mark M. Wilde. "An invitation to the sample complexity of quantum hypothesis testing". npj Quantum Information **11**, 94 (2025). arXiv:2403.17868.

[89] Theshani Nuradha and Mark M Wilde. "Query complexity of classical and quantum channel discrimination". Quantum Science and Technology **10**, 045075 (2025).

[90] Hao-Chung Cheng, Christoph Hirche, and Cambyse Rouzé. "Sample complexity of locally differentially private quantum hypothesis testing" (2024). arXiv:2406.18658.

[91] Li Zhou and Mingsheng Ying. "Differential privacy in quantum computation". In Proceedings of IEEE Computer Security Foundations Symposium (CSF). Pages 249–262. IEEE (2017).

[92] Theshani Nuradha, Ziv Goldfeld, and Mark M. Wilde. "Quantum pufferfish privacy: A flexible privacy framework for quantum systems". IEEE Transactions on Information Theory **70**, 5731–5762 (2024).

[93] Koenraad M. R. Audenaert and Jens Eisert. "Continuity bounds on the quantum relative entropy". Journal of Mathematical Physics **46**, 102104 (2005).

[94] Ji Guan, Wang Fang, and Mingsheng Ying. "Verifying fairness in quantum machine learning". In Proceedings of International Conference on Computer Aided Verification. Pages 408–429. Springer (2022).

[95] Eugene Seneta. "Non-negative matrices and Markov chains". Springer Science & Business Media. (2006).

[96] David A. Levin and Yuval Peres. "Markov chains and mixing times". Volume 107. American Mathematical Society. (2017).

[97] Daniel Burgarth, Giulio Chiribella, Vittorio Giovannetti, Paolo Perinotti, and Kazuya Yuasa. "Ergodic and mixing quantum channels in finite dimensions". New Journal of Physics **15**, 073045 (2013).

[98] Ian George, Alice Zheng, and Akshay Bansal. "Divergence inequalities with applications in ergodic theory" (2024). arXiv:2411.17241.

[99] Ian George and Marco Tomamichel. "A unified approach to quantum contraction and correlation coefficients" (2025). arXiv:2505.15281.

[100] John Hajnal. "The ergodic properties of non-homogeneous finite Markov chains". In Mathematical Proceedings of the Cambridge Philosophical Society. Volume 52, pages 67–77. Cambridge University Press (1956).

[101] John Hajnal and Maurice S. Bartlett. "Weak ergodicity in non-homogeneous Markov chains". In Mathematical Proceedings of the Cambridge Philosophical Society. Volume 54, pages 233–246. Cambridge University Press (1958).

[102] Eugene Seneta. "On the historical development of the theory of finite inhomogeneous Markov chains". In Mathematical Proceedings of the Cambridge Philosophical Society. Volume 74, pages 507–513. Cambridge University Press (1973).

[103] Benjamin Schumacher and Michael D. Westmoreland. "Approximate quantum error correction". Quantum Information Processing **1**, 5–12 (2002).

[104] Patrick Hayden, Michał Horodecki, Andreas Winter, and Jon Yard. "A decoupling approach to the quantum capacity". Open Systems & Information Dynamics **15**, 7–19 (2008).

[105] Frédéric Dupuis. "The decoupling approach to quantum information theory" (2010). arXiv:1004.1641.

[106] Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. "Quantum communication using coherent rejection sampling". Physical Review Letters **119**, 120506 (2017).

[107] Pau Colomer and Andreas Winter. "Decoupling by local random unitaries without simultaneous smoothing, and applications to multi-user quantum information tasks". Communications in Mathematical Physics **405**, 281 (2024).

[108] Hao-Chung Cheng, Li Gao, and Mario Berta. "Quantum broadcast channel simulation via multipartite convex splitting". Communications in Mathematical Physics **406**, 36 (2025).

[109] Kaiyuan Ji, Seth Lloyd, and Mark M. Wilde. "Retrocausal capacity of a quantum channel" (2025). arXiv:2509.08965.

[110] Michael A. Nielsen. "Conditions for a class of entanglement transformations". Physical Review Letters **83**, 436 (1999).

[111] Xin Wang and Mark M. Wilde. "α-logarithmic negativity". Physical Review A **102**, 032416 (2020).

A Proof of Theorem 2

The proof is largely identical to that of [50, Theorem 10], which itself is a rather direct generalization of the proof of [49, Theorem 6]. For completeness, we present it in full. Let $\mathcal{K}(B \otimes B')$ be an arbitrary convex cone such that $I \in \text{relint}(\mathcal{K})$ and it is closed under local completely-positive (CP) maps on the B' (i.e., $\text{id}_B \otimes \psi_{B' \to A'}(X) \in \mathcal{K}(A' \otimes B')$ if $X \in \mathcal{K}(B \otimes B')$ and Φ, ψ are completely-positive. We note the B' space will ultimately correspond to the A space of the cone.

Let $|\varphi\rangle_{A'B'}$ be an arbitrary pure state. Without loss of generality we can take $|\varphi\rangle$ to be maximally entangled as Nielsen's Theorem guarantees the existence of an LOCC transformation such that $|\Phi_d\rangle_{A'\widetilde{A}'} \to |\varphi\rangle_{A'B'}$, where we recall $|\Phi_d\rangle_{A'\widetilde{A}'}$ is the maximally entangled state where $A' \cong \widetilde{A}'$ and here we take $d := |A'|$. Specifically, Nielsen's theorem [110] uses a measurement on Bob's side with Kraus operators $\{M^k\}_k$ and correcting unitaries $\{U^k\}_k$ on Alice's side such that $U^k_{A'} \otimes M^k_{\widetilde{A}' \to B'} |\Phi_d\rangle_{A'\widetilde{A}'} = \sqrt{p(k)} |\varphi^k\rangle_{A'B'}$. Consider the channel state exclusion game winning probability for cone $\mathcal{K}(B \otimes B')$. That is,

$$\sum_x P(x|x) := \sum_x \sum_k \text{Tr}\left[\Lambda^x_{BB'} \left(\mathcal{N}_{A\to B} \circ \mathcal{E}^x_{A'\to A} \otimes \text{id}^{B'}\right)\left(p(k)|\varphi^k\rangle\langle\varphi^k|\right)\right], \tag{A.1}$$

where by construction

$$p(k)|\varphi^k\rangle\langle\varphi^k| = [(U^k \otimes M^k)\Phi_d(U^k \otimes M^k)^\dagger],$$

and by assumption $\Lambda^y_{BB'} \in \mathcal{K}(B \otimes B')$ for all y. Note $\{(I \otimes M^k)^\dagger \Lambda^x (I \otimes M^k)\}_{k,y}$ forms a POVM on $B\widetilde{A}'$. This follows from the fact that $\{M^k\}_k$ are Kraus operators for a CPTP map, and so the adjoint map, $\sum_k M^{k\dagger}(\cdot)M^k$ is a unital map. Similarly, note that $\mathcal{U}^k(\cdot) := U^k(\cdot)U^{k\dagger}$ denotes a unitary channel. Therefore, the collection $\{\mathcal{E}^x \circ \mathcal{U}^k\}_{x,k}$ is a family of encoders. Therefore, we can re-express (A.1) as

$$\sum_z P(z|z) = \sum_z \text{Tr}\left[\hat{\Lambda}^z_{B\widetilde{A}'}\left(\mathcal{N}_{A\to B} \circ \hat{\mathcal{E}}^z_{\widetilde{A}'\to A} \otimes \text{id}_{\widetilde{A}'}[\Phi_{d_{A'\widetilde{A}'}}]\right)\right], \tag{A.2}$$

where the $\hat{\mathcal{E}}^z$ and $\hat{\Lambda}^z$ are the concatenated encoders and decoder, i.e. we define $\mathcal{Z} = \mathcal{X} \times \mathcal{K}$. Note that as $\mathcal{K}(B \otimes \widetilde{A}')$ is closed under local CP maps, $\hat{\Lambda}^z_{B\widetilde{A}'} \in \mathcal{K}(B \otimes \widetilde{A}')$ for all z. This shows that we can restrict our attention just to a shared maximally entangled state as we built a new strategy achieving the same value. Furthermore, without loss of generality, we can assume that $d_{A'} \geq d_A$, because the transformation $|\Phi_d\rangle_{A''\widetilde{A}''} \to |\varphi\rangle_{A'B'}$ is always possible for any $d_{A''} \geq d_{A'}$ because we could have just as well used the same argument with system A'' and arrived at $\Phi_{d_{A''\widetilde{A}''}}$ in (A.2).

We next take Kraus-operator decompositions $\hat{\mathcal{E}}^z(\cdot) = \sum_k N^{z,k}(\cdot)N^{z,k\dagger}$ with each $N^{z,k} : \mathbb{C}^{d_{A'}} \to \mathbb{C}^{d_A}$. Since $d_{A'} \geq d_A$, we can use the transpose trick/ricochet property $N^{z,k} \otimes I|\Phi_d\rangle_{A'\widetilde{A}'} = I \otimes N^{z,k^T}|\Phi_d\rangle_{A\widetilde{A}}$ to obtain

$$\sum_x P(x|x) = \frac{1}{d_{A'}} \sum_z \sum_k \mathrm{Tr}\Big[\widetilde{P}^{z,k} \left(\mathcal{N}_{A\to B} \otimes \mathrm{id}_{\widetilde{A}}\right)(\Phi_{d_A})\Big] = \mathrm{Tr}[\Omega^{AB}\Gamma^{\mathcal{N}}_{AB}], \tag{A.3}$$

where $\widetilde{P}^{z,k} \coloneqq (I_B \otimes \overline{N}^{z,k})\hat{\Lambda}^z_{B\widetilde{A}'}(I_B \otimes N^{z,k^T})$ and we have swapped the ordering of the systems to match earlier notation, and

$$\Omega_{AB} = \frac{1}{d_{A'}} \sum_z \sum_k (\overline{N}^{z,k} \otimes I_B)\hat{\Lambda}^z_{A'B}(N^{z,k^T} \otimes I_B) \tag{A.4}$$

$$\coloneqq \frac{1}{d_{A'}} \sum_z \widetilde{\mathcal{E}}_{z\,A'\to A} \otimes \mathrm{id}_B \left(\hat{\Lambda}^{A'B}_z\right), \tag{A.5}$$

in which $\widetilde{\mathcal{E}}_z(\cdot) \coloneqq \sum_k \overline{N}_{z,k}(\cdot)N^T_{z,k}$. Note, $\widetilde{\mathcal{E}}_{z\,A'\to A} \otimes \mathrm{id}_B(\hat{\Lambda}^z_{A'B}) \in \mathcal{K}(A \otimes B)$ as it is again a local CP map $\widetilde{A}' \to A$ and the cone structure was invariant under CP maps on $\widetilde{A}'$ by assumption.[10] As $\mathcal{K}$ is a cone, sums of its elements are contained, so it follows that $\Omega_{AB} \in \mathcal{K}(A \otimes B)$. Moreover, since each $\widetilde{\mathcal{E}}_z$ is trace preserving, we have

$$\mathrm{Tr}_A\, \Omega_{AB} = \frac{1}{d_{A'}} \sum_z \mathrm{Tr}_A\Big[\widetilde{\mathcal{E}}_{z\,A'\to A} \otimes \mathrm{id}^B \left(\hat{\Lambda}^z_{A'B}\right)\Big] \tag{A.6}$$

$$= \frac{1}{d_{A'}} \sum_z \mathrm{Tr}_{A'} \left(\hat{\Lambda}^z_{A'B}\right) \tag{A.7}$$

$$= \frac{1}{d_{A'}} \mathrm{Tr}_{A'} \left(I_{A'} \otimes I_B\right) = I_B. \tag{A.8}$$

Hence $\mathrm{Tr}_A\, \Omega_{AB} = I_B$ and $\Omega \in \mathcal{K}$ is a necessary condition on the operator Ω_{AB} such that $\sum_x P(x|x) = \mathrm{Tr}[\Omega^{AB}\Gamma^{\mathcal{N}}_{AB}]$. That is, any solution induces a feasible solution to (3.51), so it is a necessary condition. Now we make sure it is also sufficient.

Consider an arbitrary positive semidefinite operator Ω_{AB} such that $\mathrm{Tr}_A\, \Omega_{AB} = I_B$ and $\Omega_{AB} \in \mathcal{K}(A \otimes B)$. Consider the discrete Weyl operators on A, explicitly given by $U_{m,n} = \sum_{k=0}^{d_A-1} e^{imk2\pi/d_A}|m \oplus n\rangle\langle m|$, where $m, n = 0, \ldots, d_A - 1$ and addition is taken modulo d_A. It's well known that twirling the discrete Weyl operators results in the completely depolarizing map, e.g.,

$$\Delta(X) \coloneqq \frac{1}{d_A} \sum_{m,n} U_{m,n}(X)U^\dagger_{m,n} = \mathrm{Tr}[X]\mathbb{I}\ . \tag{A.9}$$

[10] Note we re-ordered the spaces so we before had $\mathcal{K}(B \otimes \widetilde{A}')$, which was then $\mathcal{K}(\widetilde{A}' \otimes B)$, which is still invariant under the same space, the ordering merely changed.

Hence,

$$\Delta_A \otimes \mathrm{id}_B[\Omega_{AB}] = \mathbb{I}_A \otimes \mathrm{Tr}_A\, \Omega_{AB} = \mathbb{I}_A \otimes \mathbb{I}_B. \tag{A.10}$$

This implies that the set $\{\mu(m,n) \coloneqq \mathcal{U}^A_{m,n} \otimes \mathrm{id}^B(\Omega^{AB})\}_{m,n}$ forms a valid POVM on AB. Therefore, we can construct an entanglement-assisted protocol as follows. Let Alice and Bob share a maximally entangled state $|\tau_{d_A}\rangle_{\widetilde{A}A}$. Alice then applies the $\mathcal{U}^T_{m,n}(\cdot) \coloneqq U^T_{m,n}(\cdot)U^*_{m,n}$ to the A system and sends it through the channel $\mathcal{N}$. Bob then performs the POVM $\{\mu(m,n)\}_{m,n}$ just described on systems $\widetilde{A}B$. The obtained score is

$$\sum_{m,n} P(m,n|m,n) = \frac{1}{d_A}\sum_{m,n} \mathrm{Tr}\left[\left(\mathcal{U}^{\widetilde{A}}_{m,n} \otimes \mathrm{id}^B\left[\Omega_{\widetilde{A}B}\right]\right)\left(\mathrm{id}_{\widetilde{A}} \otimes \mathcal{N}\right)\left(\mathrm{id}_{\widetilde{A}} \otimes \mathcal{U}^T_{m,n}\left[\tau_{d_A}\right]\right)\right] \tag{A.11}$$

$$= \frac{1}{(d_A)^2}\sum_{m,n} \mathrm{Tr}\left[\Omega^{\widetilde{A}B}\left(\mathrm{id}^{\widetilde{A}} \otimes \mathcal{N}\left[\tau_{d_A}\right]\right)\right] \tag{A.12}$$

$$= \mathrm{Tr}[\Omega\Gamma^{\mathcal{N}}]. \tag{A.13}$$

The key idea in these equalities is that the unitary encoding $U_{m,n}$ performed on Alice's side is canceled by exactly one POVM element on Bob's side.

We therefore have necessary and sufficient conditions for $\mathcal{G}^{\mathcal{K}}_{\mathrm{QSE}}(\mathcal{N})$ to be given by $\max\left\{\mathrm{Tr}[\Omega_{AB}\Gamma^{\mathcal{N}}] : \mathrm{Tr}_A[\Omega] = I_B\,,\ \Omega_{AB} \in \mathcal{K}\right\}$, which is the conic program given in (3.51). Noting that we assumed $I \in \mathrm{relint}(\mathcal{K})$ means that strong duality holds, this completes the proof.

B Proof of Proposition 13

The proof given here is similar to the proof of [32, Proposition 7]. Consider that the desired equality in (4.40) is equivalent to

$$\frac{1}{d^2}\alpha(\mathcal{N}) = \exp\left(-\inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}\left(\pi_d \otimes \tau \| \Phi^{\mathcal{N}}\right)\right). \tag{B.1}$$

Now observe that

$$\exp\left(-\inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}\left(\pi_d \otimes \tau \| \Phi^{\mathcal{N}}\right)\right)$$

$$= \sup_{\tau\in\mathrm{aff}(\mathcal{D})} \exp\left(-D_H^{1-\frac{1}{d^2}}\left(\pi_d \otimes \tau \| \Phi^{\mathcal{N}}\right)\right) \tag{B.2}$$

$$= \sup_{\tau\in\mathrm{aff}(\mathcal{D})} \exp\left(-\left[-\ln \inf_{\Lambda\geq 0}\left\{\mathrm{Tr}[\Lambda\Phi^{\mathcal{N}}] : \mathrm{Tr}[\Lambda(\pi_d \otimes \tau)] \geq \frac{1}{d^2},\ \Lambda \leq I\right\}\right]\right) \tag{B.3}$$

$$= \sup_{\tau\in\mathrm{aff}(\mathcal{D})} \inf_{\Lambda\geq 0}\left\{\mathrm{Tr}[\Lambda\Phi^{\mathcal{N}}] : \mathrm{Tr}[\Lambda(\pi_d \otimes \tau)] \geq \frac{1}{d^2},\ \Lambda \leq I\right\}. \tag{B.4}$$

We first prove the inequality

$$\exp\left(-\inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}\left(\pi_d \otimes \tau \| \Phi^{\mathcal{N}}\right)\right) \leq \frac{1}{d^2}\alpha(\mathcal{N}). \tag{B.5}$$

Recall from Proposition 3 that the dual of $\alpha(\mathcal{N})$ is given by

$$\alpha(\mathcal{N}) = \inf_{Y_{AB} \geq 0} \left\{ \operatorname{Tr}[Y_{AB} \Gamma^{\mathcal{N}}_{AB}] : \operatorname{Tr}_A[Y_{AB}] = I_B \right\}, \tag{B.6}$$

which implies that

$$\frac{1}{d^2}\alpha(\mathcal{N}) = \inf_{Y_{AB} \geq 0} \left\{ \operatorname{Tr}\left[\frac{Y_{AB}}{d} \Phi^{\mathcal{N}}_{AB} \right] : \operatorname{Tr}_A[Y_{AB}] = I_B \right\}. \tag{B.7}$$

Let Y_{AB} be such that $\operatorname{Tr}_A[Y_{AB}] = I_B$, and let $\tau \in \operatorname{aff}(\mathcal{D})$; then

$$\operatorname{Tr}\left[\frac{Y_{AB}}{d} (\pi_d \otimes \tau) \right] = \frac{1}{d^2} \operatorname{Tr}[\operatorname{Tr}_A[Y_{AB}] \tau] \tag{B.8}$$

$$= \frac{1}{d^2} \operatorname{Tr}[I_B \tau] \tag{B.9}$$

$$= \frac{1}{d^2}. \tag{B.10}$$

Also, consider that

$$\frac{Y_{AB}}{d} \leq I_{AB} \tag{B.11}$$

because

$$\frac{1}{d^2} \sum_{i=1}^{d^2} U^i_A \frac{Y_{AB}}{d} \left(U^i_A \right)^\dagger = \pi_A \otimes \operatorname{Tr}_A\left[\frac{Y_{AB}}{d} \right] \tag{B.12}$$

$$= \frac{1}{d} \pi_A \otimes I_B, \tag{B.13}$$

where $\{U^i_A\}_{i=1}^{d^2}$ is the set of Heisenberg–Weyl operators. This implies that

$$\frac{Y_{AB}}{d} \leq \sum_{i=1}^{d^2} U^i_A \frac{Y_{AB}}{d} \left(U^i_A \right)^\dagger \tag{B.14}$$

$$= d^2 \frac{\pi_A}{d} \otimes I_B \tag{B.15}$$

$$= I_{AB}. \tag{B.16}$$

Thus, $\frac{Y_{AB}}{d}$ is a legitimate measurement operator, and it satisfies the constraint

$$\operatorname{Tr}\left[\frac{Y_{AB}}{d} (\pi_d \otimes \tau) \right] \geq \frac{1}{d^2}. \tag{B.17}$$

Then we find that

$$\inf_{\Lambda \geq 0} \left\{ \operatorname{Tr}[\Lambda \Phi^{\mathcal{N}}] : \operatorname{Tr}[\Lambda(\pi_d \otimes \tau)] \geq \frac{1}{d^2},\ \Lambda \leq I \right\} \leq \operatorname{Tr}\left[\frac{Y_{AB}}{d} \Phi^{\mathcal{N}}_{AB} \right]. \tag{B.18}$$

Taking an infimum over all $Y_{AB} \geq 0$ satisfying the constraint $\operatorname{Tr}_A[Y_{AB}] = I_B$ then implies that

$$\inf_{\Lambda \geq 0} \left\{ \operatorname{Tr}[\Lambda \Phi^{\mathcal{N}}] : \operatorname{Tr}[\Lambda(\pi_d \otimes \tau)] \geq \frac{1}{d^2},\ \Lambda \leq I \right\} \leq \frac{1}{d^2} \alpha(\mathcal{N}). \tag{B.19}$$

Since the inequality holds for all $\tau \in \mathrm{aff}(\mathcal{D})$, we can take the supremum over all such τ to conclude that

$$\exp\left(-\inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}(\pi_d\otimes\tau\|\Phi^{\mathcal{N}})\right) \leq \frac{1}{d^2}\alpha(\mathcal{N}). \tag{B.20}$$

To show the opposite inequality, consider that

$$\begin{aligned}
\frac{1}{d^2}\alpha(\mathcal{N}) &= \frac{1}{d^2}\sup_{X\in\mathrm{Herm}}\left\{\mathrm{Tr}[X] : I\otimes X \leq \Gamma^{\mathcal{N}}\right\} && \text{(B.21)}\\
&= \frac{1}{d^2}\sup_{X\in\mathrm{Herm}}\left\{\mathrm{Tr}[X] : \pi_d\otimes X \leq \Phi^{\mathcal{N}}\right\} && \text{(B.22)}\\
&= \frac{1}{d^2}\sup_{\lambda\geq 0,\tau\in\mathrm{aff}(\mathcal{D})}\left\{\lambda : \lambda(\pi_d\otimes\tau) \leq \Phi^{\mathcal{N}}\right\} && \text{(B.23)}\\
&= \sup_{\lambda\geq 0,\tau\in\mathrm{aff}(\mathcal{D})}\left\{\frac{\lambda}{d^2} : \lambda(\pi_d\otimes\tau) \leq \Phi^{\mathcal{N}}\right\} && \text{(B.24)}\\
&= \sup_{\lambda\geq 0,\tau\in\mathrm{aff}(\mathcal{D})}\inf_{\Lambda\geq 0}\left\{\frac{\lambda}{d^2} + \mathrm{Tr}[\Lambda\left(\Phi^{\mathcal{N}} - \lambda(\pi_d\otimes\tau)\right)]\right\} && \text{(B.25)}\\
&= \sup_{\lambda\geq 0,\tau\in\mathrm{aff}(\mathcal{D})}\inf_{\Lambda\geq 0}\left\{\mathrm{Tr}[\Lambda\Phi^{\mathcal{N}}] + \lambda\left(\frac{1}{d^2} - \mathrm{Tr}[\Lambda(\pi_d\otimes\tau)]\right)\right\} && \text{(B.26)}\\
&\leq \sup_{\tau\in\mathrm{aff}(\mathcal{D})}\inf_{\Lambda\geq 0}\sup_{\lambda\geq 0}\left\{\mathrm{Tr}[\Lambda\Phi^{\mathcal{N}}] + \lambda\left(\frac{1}{d^2} - \mathrm{Tr}[\Lambda(\pi_d\otimes\tau)]\right)\right\} && \text{(B.27)}\\
&= \sup_{\tau\in\mathrm{aff}(\mathcal{D})}\inf_{\Lambda\geq 0}\left\{\mathrm{Tr}[\Lambda\Phi^{\mathcal{N}}] : \frac{1}{d^2} \leq \mathrm{Tr}[\Lambda(\pi_d\otimes\tau)]\right\} && \text{(B.28)}\\
&\leq \sup_{\tau\in\mathrm{aff}(\mathcal{D})}\inf_{\Lambda\geq 0}\left\{\mathrm{Tr}[\Lambda\Phi^{\mathcal{N}}] : \frac{1}{d^2} \leq \mathrm{Tr}[\Lambda(\pi_d\otimes\tau)],\quad \Lambda\leq I\right\} && \text{(B.29)}\\
&= \sup_{\tau\in\mathrm{aff}(\mathcal{D})}\exp\left(-D_H^{1-\frac{1}{d^2}}\left(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}\right)\right) && \text{(B.30)}\\
&= \exp\left(-\inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}\left(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}\right)\right). && \text{(B.31)}
\end{aligned}$$

This concludes the proof.

C Proof of Proposition 28

As stated in the main text, the proof effectively follows the same steps as used to establish Proposition 19.

Proof of Lemma 4. Starting from the definition,

$$\begin{aligned}
\alpha_{\mp}(\mathcal{N}) &= \max_{X_B\in\mathrm{Herm}}\left\{\mathrm{Tr}[X_B] : -\Gamma^{\mathcal{N}}_{AB} \leq I_A\otimes X_B \leq \Gamma^{\mathcal{N}}_{AB}\right\} && \text{(C.1)}\\
&= \max_{\substack{\lambda\in\mathbb{R},\\ \tau\in\mathrm{aff}(\mathcal{D})}}\left\{\lambda : -\mathcal{N} \leq \lambda\mathcal{R}_\tau \leq \mathcal{N}\right\} && \text{(C.2)}
\end{aligned}$$

where the second equality follows from the same argument that established Eq. (4.24). Using that the complete positive ordering on channels means $\mathcal{M} \leq \mathcal{N} \iff \mathcal{M} = \mathcal{N} + \mathcal{G}$ for some completely positive map $\mathcal{G}$, we re-express the previous optimization as

$$\alpha_{\text{Ŧ}}(\mathcal{N}) = \max_{\substack{\lambda\in\mathbb{R},\\ \tau\in\text{aff}(\mathcal{D}),\\ \mathcal{M}',\mathcal{G}'\in\text{CP}}} \left\{\lambda : \lambda\mathcal{R}_\tau + \mathcal{M}' = \mathcal{N}\,,\, \mathcal{N} = \mathcal{G}' - \lambda\mathcal{R}_\tau\right\}\,. \tag{C.3}$$

Noting that the first constraint is the same as in Proposition 12, by the same argument as around Eq. (4.39), we conclude that $\lambda \in [0,1]$ and $\lambda\mathcal{R}_\tau + \mathcal{M}' = \mathcal{N}$ can be re-expressed as $\lambda\mathcal{R}_\tau + (1-\lambda)\,\mathcal{M}$ for $\mathcal{M}$ being a quantum channel:

$$\alpha_{\text{Ŧ}}(\mathcal{N}) = \max_{\substack{\lambda\in[0,1],\\ \tau\in\text{aff}(\mathcal{D}),\\ \mathcal{M}\in\text{CPTP},\\ \mathcal{G}'\in\text{CP}}} \left\{\lambda : \lambda\mathcal{R}_\tau + (1-\lambda)\mathcal{M} = \mathcal{N}\,,\, \mathcal{N} = \mathcal{G}' - \lambda\mathcal{R}_\tau\right\}\,. \tag{C.4}$$

Using the second constraint, for every quantum state ρ,

$$1 = \text{Tr}[\mathcal{N}(\rho)] = \text{Tr}[\mathcal{G}'(\rho)] - \lambda \Rightarrow \text{Tr}[\mathcal{G}'(\rho)] = (1+\lambda) \in [1,2]\,. \tag{C.5}$$

Thus, $\frac{1}{1+\lambda}\mathcal{G}'$ is always a CPTP map and we conclude

$$\alpha_{\text{Ŧ}}(\mathcal{N}) = \max_{\substack{\lambda\in[0,1],\\ \tau\in\text{aff}(\mathcal{D}),\\ \mathcal{M},\mathcal{G}\in\text{CPTP}}} \left\{\lambda : \lambda\mathcal{R}_\tau + (1-\lambda)\mathcal{M} = \mathcal{N}\,,\, \mathcal{N} = (1+\lambda)\mathcal{G} - \lambda\mathcal{R}_\tau\right\}\,. \tag{C.6}$$

This completes the proof of the lemma. □

Proof of Proposition 28. First, by re-parameterizing Lemma 4, we have

$$1 - \alpha_{\text{Ŧ}}(\mathcal{N}) = \min_{\substack{\lambda\in[0,1],\\ \tau\in\text{aff}(\mathcal{D}),\\ \mathcal{M},\mathcal{G}\in\text{CPTP}}} \left\{\lambda : (1-\lambda)\mathcal{R}_\tau + \lambda\mathcal{M} = \mathcal{N}\,,\, \mathcal{N} = (2-\lambda)\mathcal{G} - (1-\lambda)\mathcal{R}_\tau\right\}\,, \tag{C.7}$$

which is what we use in this proof. Now consider quantum channels $\mathcal{N}_1, \mathcal{N}_2$. Using (C.7), let $(\lambda_1, \tau_1, \mathcal{M}_1, \mathcal{G}_1)$ and $(\lambda_2, \tau_2, \mathcal{M}_2, \mathcal{G}_2)$ be optimizers of $1 - \alpha(\mathcal{N}_1)$ and $1 - \alpha(\mathcal{N}_2)$ respectively. We will use these to construct a feasible point for $(1 - \alpha_{\text{Ŧ}}(\mathcal{N}_2 \circ \mathcal{N}_1))$. Define

$$\lambda' \coloneqq \lambda_2\lambda_1 \quad \tau' \coloneqq \frac{1-\lambda_2}{1-\lambda_1\lambda_2}\tau_2 + \frac{\lambda_2\,(1-\lambda_1)}{1-\lambda_1\lambda_2}\mathcal{M}_2(\tau_1) \quad \mathcal{M}' \coloneqq \mathcal{M}_2 \circ \mathcal{M}_1\,. \tag{C.8}$$

Using that the first constraint in (C.7) is the same as the sole constraint for $1 - \alpha(\mathcal{N})$ given in Eq. (4.30), by the same argument as in the proof of Proposition 19, we conclude that these choices are feasible for the first constraint of $1 - \alpha_{\text{Ŧ}}(\mathcal{N}_2 \circ \mathcal{N}_1)$. Moreover, define

$$\mathcal{G}' \coloneqq \frac{1}{2-\lambda'}[\mathcal{N}_2 \circ \mathcal{N}_1 + (1-\lambda')\mathcal{R}_{\tau'}]\,, \tag{C.9}$$

which is completely positive as it is the positive sum of completely positive maps and is trace preserving as $\mathcal{G}'(\rho) = \frac{1}{2-\lambda'}[1 + (1-\lambda)] = 1$ for every quantum state ρ. Re-ordering Eq. (C.9), one finds $\mathcal{N}_2 \circ \mathcal{N}_1 = (2-\lambda')\mathcal{G}' - (1-\lambda)\mathcal{R}_{\tau'}$; i.e., the choice of $\mathcal{G}'$ in (C.9) satisfies the second constraint in Eq. (C.7). Thus, $(\lambda', \tau', \mathcal{M}', \mathcal{G}')$ is feasible for $1 - \alpha_{\text{Ŧ}}(\mathcal{N}_2 \circ \mathcal{N}_1)$. As it is a minimization, we have $1 - \alpha_{\text{Ŧ}}(\mathcal{N}_2 \circ \mathcal{N}_1) \leq \lambda' = \lambda_2\lambda_1 = (1 - \alpha_{\text{Ŧ}}(\mathcal{N}_2))(1 - \alpha_{\text{Ŧ}}(\mathcal{N}_1))$. This completes the proof of the proposition. □

D Qubit Channel Parameterizations

Here we recall or establish all the properties of qubit channel parameterizations that we will need.

Stokes Parameterization First we recall the Stokes parameterization of a qubit linear map. As $\{I, \sigma_X, \sigma_Y, \sigma_Z\}$ form a basis of $\mathcal{L}(\mathbb{C}^2)$, any $M \in \mathcal{L}(\mathbb{C}^2)$ may be written as $M = w_0 I + \vec{w} \cdot \vec{\sigma}$ where $\vec{w}$ is the Bloch vector. We may identify any qubit-to-qubit linear map, $\Phi \in \mathcal{T}(\mathbb{C}^2, \mathbb{C}^2)$, with a transformation map $\mathbb{T}_\Phi \in \mathbb{R}^{4\times 4}$. Then $\Phi(M)$ can be determined from the matrix multiplication $\begin{bmatrix} w_0' \\ \vec{w}' \end{bmatrix} = \mathbb{T}_\Phi \begin{bmatrix} w_0 \\ \vec{w} \end{bmatrix}$. The most important facts we will need about the transformation map are the following, which are all established or reported in [74]:

1. If Φ is trace-preserving, then $\mathbb{T}_\Phi = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{t} & T \end{bmatrix}$ where $T \in \mathbb{R}^{3\times 3}$ is the 'transfer matrix' and $\vec{t}$ is a three-dimensional column vector.
2. If Φ is trace-preserving and unital, then $\vec{t} = 0$.
3. If Φ is a trace-preserving, so it has the form given in the first item, then its complete positivity conditions are
$$|T_{11} \pm T_{22}| \leq |1 + T_{33}| \ . \tag{D.1}$$
4. If Φ is a trace-preserving and unital, then up to local pre- and post- processing $\mathbb{T} = \mathrm{diag}(1, \lambda_1, \lambda_2, \lambda_3)$ and it is CP if and only if $|\lambda_1 \pm \lambda_2| \leq |1 + \lambda_3|$. In particular, this means $(\lambda_1, \lambda_2, \lambda_3)$ need to be in the tetrahedron defined by the extreme points
$$\{(1,1,1), (1,-1,-1), (-1,1,-1), (-1,-1,1)\} \ . \tag{D.2}$$

Example 2.

1. *The depolarizing channel has $\lambda_1 = \lambda_2 = \lambda_3 = 1 - q$ where $q \in [0, 4/3]$.*
2. *The extreme points are*
$$\begin{aligned} (1,1,1) &\mapsto \mathrm{id}(\cdot) \quad (1,-1,-1) \mapsto \sigma_X \cdot \sigma_X^\dagger \\ (-1,1,-1) &\mapsto \sigma_Y \cdot \sigma_Y^\dagger \quad (-1,-1,1) \mapsto \sigma_Z \cdot \sigma_Z^\dagger \ , \end{aligned} \tag{D.3}$$
which may be verified using $\sigma_X \sigma_Y \sigma_X = -\sigma_Y$ and similar identities.

Choi Operator Pauli Parameterization The other representation that will be useful will be that of the Choi operator of a linear map Eq. (2.2). We recall that to be trace-preserving, we should have that $\mathrm{Tr}_B[\Gamma^\Phi] = I_A$, and similarly, to be unital, we should have that $\mathrm{Tr}_A[\Gamma^\Phi] = I_B$. From these criteria, it follows that for qubit-to-qubit linear maps,
$$\Phi \text{ is TP} \quad \Leftrightarrow \quad \Gamma^\Phi = \frac{1}{2}[I_A \otimes I_B + I_A \otimes \vec{r} \cdot \vec{\sigma} + \sum_{i,j} a_{i,j} \sigma_i \otimes \sigma_j] \tag{D.4}$$

$$\Phi \text{ is Unital} \quad \Leftrightarrow \quad \Gamma^{\Phi} = \frac{1}{2}[I_A \otimes I_B + \vec{s}\cdot\vec{\sigma} \otimes I_B + \sum_{i,j} b_{i,j}\sigma_i \otimes \sigma_j] \ , \tag{D.5}$$

where $\vec{s}, \vec{r} \in \mathbb{R}^3$ and $a_{i,j}, b_{i,j} \in \mathbb{R}$ for all $i, j \in \{1,2,3\}$ and we define the matrix forms $A = [a_{i,j}]$, $B = [b_{i,j}]$ are again the 'transfer matrices.' To see that the above hold, note that as $\{I, \sigma_X, \sigma_Y, \sigma_Z\}^{\times 2}$ form a basis of $\mathcal{L}(\mathbb{C}^4)$ and the only terms we have dropped in each case would have to be zero to satisfy the partial trace conditions.

The above shows a TP map is defined by $(\vec{r}, A)$ and a unital map is defined by $(\vec{s}, B)$. Note that (D.4) and (D.5) define a bijection on these parameterizations. Namely we may define the bijection $(\vec{r}, A) \leftrightarrow (\vec{s} = \vec{r}, B = A^T)$. Note this corresponds to applying the 'swap map' to convert a TP map $\mathcal{N}$ to a unital one $\mathcal{M}$, i.e., $\mathbb{F}\Gamma^{\mathcal{N}}\mathbb{F}^\dagger = \Gamma^{\mathcal{M}}$ where $\mathbb{F} : A \otimes B \to B \otimes A$ is the swap operator that may be defined via $\mathbb{F}(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) = |b_1\rangle\langle b_2| \otimes |a_1\rangle\langle a_2|$ for all $|a_1\rangle, |a_2\rangle \in A$, $|b_1\rangle, |b_2\rangle \in B$. Note that this map is positivity-preserving, i.e., $\Gamma^{\mathcal{N}} \geq 0 \Leftrightarrow \mathbb{F}\Gamma^{\mathcal{N}}\mathbb{F}^\dagger \geq 0$. This means we may immediately lift the positivity conditions of (D.1) to unital maps as the diagonal entries are left invariant under the transpose.

Proposition 54. *Let Φ be a qubit-to-qubit unital map with the parameterization in (D.5). It is unital if and only if $|B_{11} \pm B_{22}| \leq |1 + B_{33}|$.*

We also state the following fact established in [49].

Proposition 55 ([49])**.** *For every trace-preserving qubit-to-qubit map, there exist qubit unitaries U and V such that*

$$\Gamma^{\mathcal{V}\circ\Phi\circ\mathcal{U}} = U \otimes V\Gamma^{\Phi}(U \otimes V)^\dagger = \frac{1}{2}[I_A \otimes I_B + I_A \otimes \vec{r}'\cdot\vec{\sigma} + \sum_k t'_k \sigma_k \otimes \sigma_k] \ , \tag{D.6}$$

where $\mathcal{U}, \mathcal{V}$ denote using U, V as post and pre-processing respectively.

We provide a proof for completeness as it does not seem to follow from the Stokes parameterization in a straightforward manner.

Proof. Let Φ have the form in (D.4). Then there exists a singular value decomposition of the transfer matrix $A = \sum_k t'_k \vec{a}_k \vec{b}_k^T$ where $\{\vec{a}_k\}_k, \{\vec{b}_k\}_k$ are orthonormal basis of $\mathbb{R}^3$. Then there exist rotations $O_1, O_2 \in SO(3)$ such that $O_1 : \{\vec{a}_k\}_k \mapsto \{e_0, e_1, \pm e_2\}_k$, $O_2 : \{\vec{b_k}\} \mapsto \{e_0, e_1, \pm e_2\}$, where the $\pm$ denotes that the rotation must respect the determinant of the initial set of vectors. Given the isomorphism $SO(3) \cong SU(2) \setminus \mathbb{Z}_2$, this means there exist unitaries $U, V \in SU(2)$ such that

$$U(\vec{a}_k\cdot\vec{\sigma})U^\dagger \otimes V(\vec{b}_k\cdot\vec{\sigma})V^\dagger = \begin{cases} \sigma_k \otimes \sigma_k & k \in \{1,2\} \\ \pm\sigma_3 \otimes \sigma_3 & k = 3 \end{cases} \ . \tag{D.7}$$

Thus,

$$U \otimes V\Gamma^{\Phi}(U \otimes V)^\dagger = \frac{1}{2}[I_A \otimes I_B + I_A \otimes V(\vec{r}'\cdot\vec{\sigma}) + \sum_k t'_{k'} \sigma_k \otimes \sigma_k] \ , \tag{D.8}$$

where $t'_{k'}$ captures the signedness and $\vec{r}'$ captures that the values of $\vec{r}$ have been shifted. The pre- and post-processing relation is a well-known relation on Choi matrices and may be verified via direct calculation. □

E Another Alternative Notion of Quantum Doeblin Coefficient

Definition 24. *Define*

$$\alpha_{\curlywedge}(\mathcal{N}) \coloneqq \exp\left(-\inf_{\beta>1}\left\{\inf_{\tau\in\mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}) + \frac{2}{\beta-1}\ln d\right\}\right), \tag{E.1}$$

where the extended sandwiched Rényi relative entropy is defined for $\tau \in \mathrm{aff}(\mathcal{D})$ and $\sigma \geq 0$ as follows [111, Eq. (13)]:

$$\widetilde{D}_\beta(\tau\|\sigma) \coloneqq \frac{\beta}{\beta-1}\ln\left\|\sigma^{\frac{1-\beta}{2\beta}}\tau\sigma^{\frac{1-\beta}{2\beta}}\right\|_\beta \tag{E.2}$$

with $\|A\|_p \coloneqq \left(\mathrm{Tr}\left[\left(\sqrt{AA^\dagger}\right)^p\right]\right)^{1/p}$.

Several properties of the extended sandwiched Rényi relative entropy have been investigated in [111] and [32, Section III].

Proposition 56. *Let $\mathcal{N}$ be a quantum channel. The following inequality holds for all $n \in \mathbb{N}$:*

$$\eta_{\mathrm{Tr}}(\mathcal{N}^{\otimes n}) \leq 1-\alpha(\mathcal{N}^{\otimes n}) \leq 1-\alpha_{\curlywedge}(\mathcal{N})^n \leq n(1-\alpha_{\curlywedge}(\mathcal{N})), \tag{E.3}$$

where $c(\mathcal{N})$ is defined in (E.1).

Proof. Applying Proposition 13 and [32, Lemma 6] for $\beta > 1$, we conclude that

$$-\ln\left(\frac{1}{d^2}\alpha(\mathcal{N})\right) = \inf_{\tau\in\mathrm{aff}(\mathcal{D})} D_H^{1-\frac{1}{d^2}}(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}) \tag{E.4}$$

$$\leq \inf_{\tau\in\mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}) + \frac{\beta}{\beta-1}\ln d^2 \tag{E.5}$$

$$= \inf_{\tau\in\mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}) + \frac{2\beta}{\beta-1}\ln d. \tag{E.6}$$

We can rewrite this as

$$-\ln\alpha(\mathcal{N}) \leq \inf_{\tau\in\mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}) + \frac{2\beta}{\beta-1}\ln d - 2\ln d \tag{E.7}$$

$$= \inf_{\tau\in\mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d\otimes\tau\|\Phi^{\mathcal{N}}) + \frac{2}{\beta-1}\ln d. \tag{E.8}$$

Since this inequality holds for all $\beta > 1$, we conclude that

$$-\ln\alpha(\mathcal{N}) \leq -\ln\alpha_{\curlywedge}(\mathcal{N}). \tag{E.9}$$

Now applying this to a tensor-power channel $\mathcal{N}^{\otimes n}$, we find that

$$-\ln\alpha(\mathcal{N}^{\otimes n}) \leq -\ln\alpha_{\curlywedge}(\mathcal{N}^{\otimes n}) \tag{E.10}$$

$$= \inf_{\beta>1}\left\{\inf_{\tau^{(n)}\in\mathrm{aff}(\mathcal{D}^{(n)})} \widetilde{D}_\beta(\pi_d^{\otimes n}\otimes\tau^{(n)}\|(\Phi^{\mathcal{N}})^{\otimes n}) + \frac{2}{\beta-1}\ln d^n\right\} \tag{E.11}$$

$$\leq \inf_{\beta>1}\left\{\inf_{\tau\in\mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d^{\otimes n}\otimes\tau^{\otimes n}\|(\Phi^{\mathcal{N}})^{\otimes n}) + \frac{2}{\beta-1}\ln d^n\right\} \tag{E.12}$$

$$= n \cdot \inf_{\beta>1} \left\{ \inf_{\tau \in \mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d \otimes \tau \| \Phi^{\mathcal{N}}) + \frac{2}{\beta - 1} \ln d \right\} \tag{E.13}$$

$$= n \left(- \ln \alpha_{\curlywedge}(\mathcal{N}) \right). \tag{E.14}$$

The established inequality is equivalent to the inequality $1 - \alpha(\mathcal{N}^{\otimes n}) \leq 1 - \alpha_{\curlywedge}(\mathcal{N})^n$, thus establishing the second inequality in (E.3).

The first inequality in (E.3) follows because $\eta_{\mathrm{Tr}}(\mathcal{N}) \leq 1 - \alpha(\mathcal{N})$. The last inequality in (E.3) follows because $1 - x^n \leq n(1 - x)$ for all $n \in \mathbb{N}$ and $x \in [0, 1]$. □

Corollary 18. *The quantity $\alpha_{\curlywedge}(\mathcal{N})$ is weakly supermultiplicative; i.e., for all $n \in \mathbb{N}$, the following inequality holds:*

$$\alpha_{\curlywedge}(\mathcal{N})^n \leq \alpha_{\curlywedge}(\mathcal{N}^{\otimes n}). \tag{E.15}$$

Proof. This claim follows by rearranging terms in (E.10)–(E.14). □

Proposition 57. *For a quantum channel $\mathcal{N}$, the following inequalities hold:*

$$\alpha(\mathcal{N}) \geq \alpha_{\curlywedge}(\mathcal{N}) \geq \alpha_{\curlyvee}(\mathcal{N}). \tag{E.16}$$

Proof. The first inequality was established in (E.9). The second inequality follows by taking the $\beta \to \infty$ limit in the objective function of $c(\mathcal{N}$ and noting that the limit gives $\alpha_{\curlyvee}(\mathcal{N})$; i.e.,

$$\lim_{\beta \to \infty} \exp\left(- \inf_{\tau \in \mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d \otimes \tau \| \Phi^{\mathcal{N}}) + \frac{2}{\beta - 1} \ln d \right) = \alpha_{\curlyvee}(\mathcal{N}). \tag{E.17}$$

This concludes the proof. □

It is an open question to determine whether $\alpha_{\curlyvee}(\mathcal{N})$ is equal to $\alpha_{\curlywedge}(\mathcal{N})$.

Proposition 58. *Let $\mathcal{N}$ be a quantum channel. The following inequality holds:*

$$- \ln \alpha_{\curlywedge}(\mathcal{N}) \leq \inf_{\sigma \in \mathcal{D}} D_{\max}(\pi_d \otimes \sigma \| \Phi^{\mathcal{N}}). \tag{E.18}$$

where $\alpha_{\curlywedge}(\mathcal{N})$ is defined in (E.1), π_d is the maximally mixed state of dimension d, and $\Phi^{\mathcal{N}}$ is the Choi state of the channel $\mathcal{N}$.

Proof. Let $\sigma \in \mathcal{D}$. Then

$$- \ln \alpha_{\curlywedge}(\mathcal{N}) = \inf_{\beta>1} \left\{ \inf_{\tau \in \mathrm{aff}(\mathcal{D})} \widetilde{D}_\beta(\pi_d \otimes \tau \| \Phi^{\mathcal{N}}) + \frac{2}{\beta - 1} \ln d \right\} \tag{E.19}$$

$$\leq \inf_{\beta>1} \left\{ \widetilde{D}_\beta(\pi_d \otimes \sigma \| \Phi^{\mathcal{N}}) + \frac{2}{\beta - 1} \ln d \right\} \tag{E.20}$$

$$\leq \lim_{\beta \to \infty} \left\{ \widetilde{D}_\beta(\pi_d \otimes \sigma \| \Phi^{\mathcal{N}}) + \frac{2}{\beta - 1} \ln d \right\} \tag{E.21}$$

$$= D_{\max}(\pi_d \otimes \sigma \| \Phi^{\mathcal{N}}). \tag{E.22}$$

Since the inequality holds for all $\sigma \in \mathcal{D}$, the desired inequality follows. □

Remark 16 (Comparison to Existing Bounds)**.** *By Proposition 58, we conclude that $\alpha_{\curlywedge}(\mathcal{N})$ in (E.3) gives an improved bound compared to [14, Eq. (III.61)].*

F Extra Proofs on Mixing and Decoupling Times of Quantum Processes

F.1 Proof of Proposition 45

(Strongly Mixing $\subsetneq$ Mixing): For both homogeneous and inhomogeneous Markov chains, by definition, if a Markov chain is strongly mixing, it is mixing. On the other hand, any replacement channel which prepares a pure state is mixing but not strongly mixing. This establishes the strict containment of strongly mixing in mixing.

(Mixing $\subseteq$ Weakly Mixing): We prove the inhomogeneous case as it is more general. Let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be mixing with state $\omega \in \mathcal{D}(A)$. Then, for any $\rho, \sigma \in \mathcal{D}(A)$,

$$\|\mathcal{G}_n(\rho) - \mathcal{G}_n(\sigma)\|_1 = \|\mathcal{G}_n(\rho) - \omega + \omega - \mathcal{G}_n(\sigma)\|_1 \tag{F.1}$$

$$\leq \|\mathcal{G}_n(\rho) - \omega\|_1 + \|\mathcal{G}_n(\sigma) - \omega\|_1 \ , \tag{F.2}$$

where we have used the triangle inequality. It follows,

$$\lim_{n\to\infty} \|\mathcal{G}_n(\rho) - \mathcal{G}_n(\sigma)\|_1 \leq \lim_{n\to\infty} [\|\mathcal{G}_n(\rho) - \omega\|_1 + \|\mathcal{G}_n(\sigma) - \omega\|_1] = 0 \ , \tag{F.3}$$

where we used each limit on the RHS exists and is zero by assumption. As ρ, σ were arbitrary and we only assumed $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is mixing, this proves mixing implies weakly mixing, which proves the containment.

(Always Weakly $\subseteq$ Weakly): Weakly mixing is the $k = 1$ case of being always weakly mixing (Definition 18), so being always weakly mixing implies weakly mixing.

(Weakly $\supseteq$ Always Weakly for Homogeneous Markov Chains): For a homogeneous Markov chain $\mathcal{N}_{A\to A}$, for any $k \leq n \in \mathbb{N}$, $\mathcal{G}_{k:n} = \bigcirc_{i=1}^{n-k} \mathcal{N}$, so weakly mixing implies always weakly mixing by definition.

(Mixing $\supseteq$ Weakly Mixing for Homogeneous Markov Chains): Let $\mathcal{N}_{A\to A}$ be a homogeneous Markov chain that is weakly mixing. By definition, for all $\rho, \sigma \in \mathcal{D}(A)$,

$$\lim_{n\to\infty} \|\mathcal{N}^n(\rho) - \mathcal{N}^n(\sigma)\|_1 = 0 \ . \tag{F.4}$$

It is known that any quantum channel has at least one fixed point state (see e.g. [25, Theorem 4.24]). Let $\omega \in \mathcal{D}(A)$ denote this existing fixed point for $\mathcal{N}$. Then by letting σ in the above be ω, we conclude for all $\rho \in \mathcal{D}(A)$

$$0 = \lim_{n\to\infty} \|\mathcal{N}^n(\rho) - \mathcal{N}^n(\omega)\|_1 = \lim_{n\to\infty} \|\mathcal{N}^n(\rho) - \omega\|_1 \ , \tag{F.5}$$

where the first equality is by the definition of weakly mixing and the second is that ω is a fixed point. By Definition 16, this means the homogeneous Markov chain is mixing. Thus, for a homogeneous Markov chain, being weakly mixing implies mixing, which establishes the containment.

((Strongly) Mixing but not Always Weakly Mixing for Inhomogeneous Markov Chains): Consider $(\mathcal{G}_n)_{n\in\mathbb{N}}$ where $\mathcal{N}^1$ is any replacer channel and $\mathcal{N}^i = \mathrm{id}_{A_2\to A_2}$ for all $i > 1$. Then

$(\mathcal{G}_n)_{n\in\mathbb{N}}$ is mixing by direct calculation, but for any $n \geq 2$, $\mathcal{G}_{2:n}$ is not mixing by direct calculation, and thus not always weakly mixing. To separate from strongly mixing, let the replacer channel prepare a full rank state in $\mathcal{D}(A_2)$.

(Always Weakly $\subsetneq$ Weakly for Inhomogeneous Markov Chains): The same example as the previous case is weakly mixing but not always weakly mixing.

(Always Weakly Mixing but not (Strongly) Mixing for Inhomogeneous Markov Chains): Consider $(\mathcal{G}_n)_{n\in\mathbb{N}}$ where

$$\mathcal{N}^i = \begin{cases} \mathcal{R}^{|0\rangle\langle 0|} & i \text{ is odd} \\ \mathcal{R}^{|1\rangle\langle 1|} & i \text{ is even} \end{cases} . \tag{F.6}$$

Recall that $\mathcal{R}^\tau(\rho) = \mathrm{Tr}[\rho]\tau$ for all ρ. Then $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is not mixing as it never converges to a fixed point. As it is not mixing, it is not strongly mixing. However, it is always weakly mixing by direct calculation.

(Always Weakly Mixing $\cap$ Mixing $\neq \emptyset$ and Always Weakly Mixing $\cap$ Strongly Mixing $\neq \emptyset$): These follow from the containments for homogeneous Markov Chains.

(Mixing $\subsetneq$ Weakly Mixing for Inhomogeneous Markov Chains): The construction that separates Always Weakly Mixing and Mixing also works for this case.

F.2 Proof of Proposition 50

In effect, this is the same proof as for Proposition 45 as all the examples considered there use replacer channels, which are decoupling. However, for completeness, we include the proof.

(Strongly Decoupling $\subsetneq$ Decoupling): For both homogeneous and inhomogeneous Markov chains, by definition, if a Markov chain is strongly decoupling, it is decoupling. On the other hand, any replacement channel which prepares a pure state is decoupling but not strongly decoupling. This establishes the strict containment of strongly mixing in mixing.

(Decoupling $\subseteq$ Weakly Decoupling): We prove the inhomogeneous case as it is more general. Let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be decoupling with state $\omega \in \mathcal{D}(A)$. Then, for any $\rho, \sigma \in \mathcal{D}(R \otimes A)$ such that $\rho_R = \sigma_R$,

$$\|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho) - (\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma)\|_1 \tag{F.7}$$
$$= \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho) - \rho_R \otimes \omega + \rho_R \otimes \omega - (\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma)\|_1 \tag{F.8}$$
$$\leq \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho) - \rho_R \otimes \omega\|_1 + \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma) - \sigma_R \otimes \omega\|_1 , \tag{F.9}$$

where we have used the triangle inequality and that $\rho_R = \sigma_R$. It follows,

$$\lim_{n\to\infty} \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho) - (\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma)\|_1 \tag{F.10}$$
$$\leq \lim_{n\to\infty} [\|(\mathrm{id}_R \otimes \mathcal{G}_n)(\rho) - \rho_R \otimes \omega\|_1 + \|(\mathrm{id}_R \otimes \mathcal{G}_n)(\sigma) - \sigma_R \otimes \omega\|_1] = 0 , \tag{F.11}$$

where we used each limit on the RHS exists and is zero by assumption. As ρ, σ were arbitrary other than $\rho_R = \sigma_R$ and we only assumed $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is decoupling, this proves decoupling

implies weakly decoupling, which proves the containment.

(Always Weakly $\subseteq$ Weakly): Weakly decoupling is the $k = 1$ case of being always weakly decoupling (Definition 22), so being always weakly decoupling implies weakly decoupling.

(Weakly $\supseteq$ Always Weakly for Homogeneous Markov Chains): For a homogeneous Markov chain $\mathcal{N}_{A\to A}$, for any $k \leq n \in \mathbb{N}$, $\mathcal{G}_{k:n} = \bigcirc_{i=1}^{n-k}\mathcal{N}$, so weakly decoupling implies always weakly decoupling by definition.

(Decoupling $\supseteq$ Weakly Decoupling for Homogeneous Markov Chains): Let $\mathcal{N}_{A\to A}$ be a homogeneous Markov chain that is weakly decoupling. By definition, for all $\rho, \sigma \in \mathcal{D}(R \otimes A)$ such that $\rho_R = \sigma_R$,

$$\lim_{n\to\infty} \|(\mathrm{id}_R \otimes \mathcal{N}^n)(\rho) - (\mathrm{id}_R \otimes \mathcal{N}^n)(\sigma)\|_1 = 0 \ . \tag{F.12}$$

It is known that any quantum channel has at least one fixed point state (see e.g. [25, Theorem 4.24]). Let $\omega \in \mathcal{D}(A)$ denote this existing fixed point for $\mathcal{N}$. Then by letting $\rho \in \mathcal{D}(R \otimes A)$ be arbitrary and $\sigma_{RA} = \rho_R \otimes \omega_A$ in the above, we conclude

$$0 = \lim_{n\to\infty} \|(\mathrm{id}_R \otimes \mathcal{N}^n)(\rho_{RA}) - (\mathrm{id}_R \otimes \mathcal{N}^n)(\rho_R \otimes \omega_A)\|_1 \tag{F.13}$$

$$= \lim_{n\to\infty} \|(\mathrm{id}_R \otimes \mathcal{N}^n)(\rho) - \rho_R \otimes \omega_A\|_1 \ , \tag{F.14}$$

where the first equality is by the definition of weakly decoupling and the second is that ω is a fixed point. As this held for all $\rho \in \mathcal{D}(R \otimes A)$, by Definition 21, this means the homogeneous Markov chain is decoupling. Thus, for a homogeneous Markov chain, being weakly decoupling implies decoupling, which establishes the containment.

((Strongly) Decoupling but not Always Weakly Decoupling for Inhomogeneous Markov Chains): Consider $(\mathcal{G}_n)_{n\in\mathbb{N}}$ where $\mathcal{N}^1$ is any replacer channel and $\mathcal{N}^i = \mathrm{id}_{A_2\to A_2}$ for all $i > 1$. Then, as a replacer channel is decoupling, $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is decoupling by direct calculation, but for any $n \geq 2$, $\mathcal{G}_{2:n}$ is not decoupling by direct calculation, and thus not always weakly decoupling. To separate from strongly decoupling, let the replacer channel prepare a full rank state in $\mathcal{D}(A_2)$.

(Always Weakly $\subsetneq$ Weakly for Inhomogeneous Markov Chains): The same example as the previous case is weakly decoupling but not always weakly decoupling.

(Always Weakly Decoupling but not (Strongly) Decoupling for Inhomogeneous Markov Chains): Consider $(\mathcal{G}_n)_{n\in\mathbb{N}}$ where

$$\mathcal{N}^i = \begin{cases} \mathcal{R}^{|0\rangle\langle 0|} & i \text{ is odd} \\ \mathcal{R}^{|1\rangle\langle 1|} & i \text{ is even} \end{cases} , \tag{F.15}$$

where we recall that $\mathcal{R}^\tau(\rho) = \mathrm{Tr}[\rho]\tau$ for all ρ. Then $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is not decoupling as it never converges to a fixed point. As it is not decoupling, it is not strongly decoupling. However, it is always weakly decoupling by direct calculation.

(Always Weakly Decoupling $\cap$ (Strong) Decoupling $\neq \emptyset$): These follow from the containments for homogeneous Markov Chains.

(Decoupling $\subsetneq$ Weakly Decoupling for Inhomogeneous Markov Chains): The construction that separates Always Weakly Decoupling and Decoupling also works for this case.