Invalidity of a strong capacity for a quantum channel with memory

Tony Dorlas[*]

School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland

Ciara Morgan[†]

Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543

The strong capacity of a particular channel can be interpreted as a sharp limit on the amount of information which can be transmitted reliably over that channel. To evaluate the strong capacity of a particular channel one must prove both the direct part of the channel coding theorem and the strong converse for the channel. Here we consider the strong converse theorem for the periodic quantum channel and show some rather surprising results. We first show that the strong converse does not hold in general for this channel and therefore the channel does not have a strong capacity. Instead, we find that there is a scale of capacities corresponding to error probabilities between integer multiples of the inverse of the periodicity of the channel. A similar scale also exists for the random channel.

PACS number(s): 03.67.Hk, 89.70.Kn

I. INTRODUCTION

The full channel coding theorem provides a limit on the rate at which a sender can communicate an encoded message to a receiver, such that the probability of a decoding error at the receiver's side decays exponentially in the number of channel uses. The theorem is composed of two parts: the direct part of the theorem, which refers to the construction of the code, and the converse to the theorem. The direct part of the quantum channel coding theorem states that using n copies of the channel, we can code with an exponentially small probability of error at a rate $R = \frac{1}{n}\log_2|\mathcal{M}|$, provided $R < C$ in the asymptotic limit, where $\mathcal{M}$ denotes the set of possible codewords to be transmitted over the channel and C denotes the capacity of the channel. If the rate at which classical information is transmitted over a quantum channel exceeds the capacity of the channel, that is, if $R > C$, then the probability of decoding the information correctly goes to zero in the number of channel uses. The latter is known as the strong converse to the channel coding theorem. The *weak* converse, on the other hand, states that if $R > C$, then the probability of decoding the information correctly is bounded away from 1; that is, the error probability does not tend to zero, whatever encoding-decoding scheme is used.

Shannon [1] first proposed the theorem for classical discrete memoryless channels and the first rigorous proof of the direct part of the theorem was provided by Feinstein [2] and the strong converse by Wolfowitz [3].

However, it was observed that the existence of the strong converse, and therefore strong capacity, for other types of classical channels does not always hold [4]. See Ahlswede [5] for a more complete discussion of converse results for various types of classical channels.

The strong converse to the channel coding theorem for memoryless *classical-quantum* channels with product state inputs was determined independently by Winter [6] and by

Ogawa and Nagaoka [7]. Their result implies that every memoryless discrete classical-quantum channel has a strong capacity which provides a sharp upper bound on the rate at which classical information can be transmitted over this type of channel using product states.

Recent results include a proof by Bjelaković and Boche [8,9] of a full coding theorem for the discrete memoryless compound classical-quantum channel. Wehner and König [10] proved the fully general strong converse theorem for a family of channels; that is, they proved that the strong converse theorem holds for a family of quantum channels even in the case when entangled state inputs are allowed.

In this article we relax the assumption that the communication channel in question is memoryless and we concentrate on a particular quantum channel with memory, that is, a channel with correlations between successive channel uses. In our case the correlations between successive uses of the channel can be described by a Markov chain. Communication channels with memory are widely considered to be more realistic than memoryless channels since real-world channels may not exhibit independence between successive errors and correlations are common. Noise correlations are also necessary for certain models of quantum communication [11]. See, for example, Kretschmann and Werner [12] and Mancini [13] for models of quantum memory channels.

The article is organized as follows. We introduce notation and necessary definitions and define the quantum periodic channel in Sec. II. In Sec. III we prove that the periodic channel does not have a strong capacity. The observation relies on a result which is proved in the Appendix, involving a particular instance of a periodic channel and consequently the strong converse does not hold in general for the periodic channel. In Sec. V we remark on a scale of capacities for the random channel. We then state and prove the main result involving a scale of capacities for the channel.

II. PRELIMINARIES

We begin by introducing some notation. A memoryless channel is given by a completely positive trace-preserving

---

[*]dorlas@stp.dias.ie
[†]cqtciara@nus.edu.sg

(CPT) map $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$, where $\mathcal{B}(\mathcal{H})$ and $\mathcal{B}(\mathcal{K})$ denote the states on the input and output Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, respectively.

Equivalently, we can describe a *classical-quantum* channel, here also denoted Φ, as a mapping from the classical message to the output state of the channel on $\mathcal{B}(\mathcal{K})$ as follows:

$$\Phi : \mathcal{X} \mapsto \mathcal{B}(\mathcal{K}), \qquad (2.1)$$

where the message is first encoded into a sequence belonging the set $\mathcal{X}^n$, where $\mathcal{X}$ represents the input alphabet.

We can combine the two mapping descriptions as follows. We wish to send classical information in the form of quantum states over a quantum channel Φ. A (discrete) memoryless quantum channel, Φ, carrying classical information can be thought of as a map from a (finite) set, or alphabet, $\mathcal{X}$ into $\mathcal{B}(\mathcal{K})$, taking each $x \in \mathcal{X}$ to $\Phi_x = \Phi(\rho_x)$, where the input state to the channel is given by $\{\rho_x\}_{x \in \mathcal{X}}$ and each $\rho_x \in \mathcal{B}(\mathcal{H})$. Let $d = \dim(\mathcal{H})$ and $a = |\mathcal{X}|$.

For a probability distribution P on the input alphabet $\mathcal{X}$, the average output state of a channel Φ is given by

$$P\sigma = \sum_{x \in \mathcal{X}} P(x) \Phi(\rho_x). \qquad (2.2)$$

The conditional von Neumann entropy of Φ given P is defined by

$$S(\Phi|P) = \sum_{x \in \mathcal{X}} P(x) S(\Phi(\rho_x)), \qquad (2.3)$$

and the mutual information between the probability distribution P and the channel Φ is defined as follows:

$$I(P;\Phi) = S(P\sigma) - S(\Phi|P). \qquad (2.4)$$

An *n-block code* for a quantum channel Φ is a pair (C^n, E^n), where C^n is a mapping from a finite set of messages $\mathcal{M}$, of length n, into $\mathcal{X}^n$; that is, a sequence $x^n \in \mathcal{X}$ is assigned to each of the $|\mathcal{M}|$ messages, and E^n is a Positive Operator Valued Measure (POVM) that is, a quantum measurement, on the output space $\mathcal{K}^{\otimes n}$ of the channel $\Phi_{x^n}^n$. The maximum error probability of the code (C^n, E^n) is defined as

$$p_e(C^n, E^n) = \max \left\{ 1 - \text{Tr}\left(\Phi_{C^n(m)}^n E_m^n\right) : m \in \mathcal{M} \right\}. \quad (2.5)$$

The code (C^n, E^n) is called an (n,λ)-code, if $p_e(C^n, E^n) \leqslant \lambda$. The maximum size $|\mathcal{M}|$ of an (n,λ)-code is denoted $N(n,\lambda)$. Define an finite alphabet $\mathcal{X}$ and sequences $x^n = x_1, \ldots, x_n \in \mathcal{X}^n$ and let

$$N(x|x^n) = |\{i \in \{1, \ldots, n\} : x_i = x\}| \qquad (2.6)$$

for $x \in \mathcal{X}$. The *type* of the sequence x^n is given by the empirical distribution P_{x^n} on $\mathcal{X}$ such that

$$P_{x^n}(x) = \frac{N(x|x^n)}{n}. \qquad (2.7)$$

Clearly, the number of types is upper bounded by $(n+1)^a$, where $a = |\mathcal{X}|$.

A. Coding theorem and strong converse

The *strong* capacity of a particular channel provides a *sharp* threshold on the rate at which information may be transmitted over that channel with *exponentially decreasing* probability

of decoding error in the number of channel uses. In order to establish a strong capacity for a particular channel one must prove both existence of a capacity achieving code and the strong converse.

The direct part of the coding theorem for memoryless quantum channels with product-state inputs was determined independently by Holevo [14] and Schumacher and Westmoreland [15]. Winter [6] and Ogawa and Nagaoka [7] independently proved the strong converse for memoryless quantum channels.

In Sec. III we require a version of the strong converse theorem proved by Winter [6] which holds for a single codeword type. We therefore provide this version (Lemma II.1) below, following both the direct part and the strong converse theorems for memoryless classical quantum channels as stated and proved in [6].

Theorem II.1 (direct part). For all $\lambda \in (0,1)$ and $\delta > 0$ there exists $n_0(\lambda,\delta) \in N$ such that for all $n \geqslant n_0$ and every classical quantum channel Φ and probability distribution P on $\mathcal{X}$, there exists an (n,λ)-code such that the number of messages satisfies

$$|\mathcal{M}_n| \geqslant 2^{n(\chi^*(\Phi)-\delta)}, \qquad (2.8)$$

where the Holevo capacity χ^* is given by

$$\chi^*(\Phi) = \sup_P I(P;\Phi), \qquad (2.9)$$

the supremum being over all probability distributions P on $\mathcal{X}$.

Theorem II.2 (strong converse). For all $\lambda \in (0,1)$ and all $\delta > 0$ there exists $n_1(\lambda,\delta)$ such that for all $n \geqslant n_1$ and every memoryless classical quantum channel Φ and the number of messages of an (n,λ)-code is bounded by

$$|\mathcal{M}_n| \leqslant 2^{n(\chi^*(\Phi)+\delta)}. \qquad (2.10)$$

Remark. Winter, in fact, proved a stronger version of these theorems in which δ is replaced by a constant times $1/\sqrt{n}$.

In the following we follow the approach of Winter ([6], Theorem 13), in which the strong converse is derived from a bound on the number of codewords of a given *type P*.

Lemma II.1 (single-type strong converse). For $\lambda \in (0,1)$ and $\delta > 0$ there exists $n_1(\lambda,\delta)$ such that for $n \geqslant n_1$, every (n,λ)-code for which all codewords are of the same type P,

$$|\mathcal{M}_{n,P}| \leqslant 2^{n[I(P;\Phi)+\delta]}. \qquad (2.11)$$

The strong converse follows immediately from this lemma using the fact that the number of types is upper bounded by $(1+n)^a$ (see [16] Lemma 2.2).

Remark. In contrast to the strong converse where the decoding error goes to 1 exponentially in the number of channel applications if $R > C$, the *weak* converse states that if $R > C$, then the probability of decoding the information correctly is bounded away from 1.

B. Quantum channels with classical memory

Next, we provide definitions needed to describe quantum channels with classical memory [17]. Let I denote a countable set and let $\lambda_i = \mathbb{P}(X = i)$, where X is a random variable taking values in the state space I. Let Q denote a transition matrix, with entries labeled $q_{j|i}$. A discrete time random process denoted X_n can be considered to be a Markov chain with

transition matrix Q and initial distribution λ, if and only if the following holds for $i_0, \ldots, i_{n-1} \in I$,

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1})$$
$$= \lambda_{i_0} q_{i_1|i_0} q_{i_2|i_1} \cdots q_{i_{n-1}|i_{n-2}}. \quad (2.12)$$

In Ref. [18] Datta and Dorlas analyze a quantum channel of length n with Markovian noise correlations, first defined by Bowen and Mancini [19], as follows:

$$\Phi^n(\rho^n) = \sum_{i_0, \ldots, i_{n-1}} q_{i_{n-1}|i_{n-2}} \cdots q_{i_1|i_0} \lambda_{i_0} \big(\Phi_{i_0} \otimes \cdots \otimes \Phi_{i_{n-1}} \big)(\rho^n),$$

$$(2.13)$$

where $q_{j|i}$ are the elements of the transition matrix of a discrete-time Markov chain, and $\{\lambda_i\}$ represents an invariant distribution on the Markov chain.

In Sec. III we analyze a particular channel with classical memory, namely, the periodic channel. We describe this channel below.

A periodic channel acting on an n-fold input state can be described as follows:

$$\Phi^n(\rho^n) = \frac{1}{L} \sum_{i=0}^{L-1} (\Phi_i \otimes \Phi_{i+1} \otimes \cdots \otimes \Phi_{i+n-1})(\rho^n), \quad (2.14)$$

where Φ_i are CPT maps acting on the same Hilbert space and the index is cyclic, modulo the period L, that is, $\Phi_{i+L} = \Phi_i$. In this case the elements of the corresponding transition matrix are given by $q_{j|i} = \theta_{i,j}$, where

$$\theta_{i,j} = \begin{cases} 1, & \text{if} \quad j = i+1 \mod L, \\ 0, & \text{otherwise.} \end{cases} \quad (2.15)$$

The product-state capacity of the channel, denoted C_p is given by

$$C_p(\Phi) = \frac{1}{L} \sup_P \sum_{i=0}^{L-1} I(P; \Phi_i). \quad (2.16)$$

The proof of the direct part of the channel coding theorem for the periodic quantum channel is provided in Appendix B of [20]. This is, in fact, a special case of the main result proved by Datta and Dorlas in Ref. [18]. Note that the proof of the direct part of the coding theorem for this channel makes use of a *preamble* to the code which the receiver uses upon receipt to determine which branch of the channel was selected.

Another channel of the general type (2.13) is the random channel. It is given by

$$\Phi^n(\rho^n) = \sum_{i=1}^{M} q_i \, \Phi_i^{\otimes n}(\rho^n), \quad (2.17)$$

where Φ_i $(i = 1, \ldots, M)$ are CPT maps acting on the same Hilbert space and $q_1, \ldots, q_M$ is a probability distribution. In this case the elements of the corresponding transition matrix are given by $q_{j|i} = \delta_{ij}$. It was shown in Ref. [21] that the product state capacity of this channel is given by

$$C_p(\Phi) = \sup_P \min_{i=1}^{M} I(P; \Phi_i). \quad (2.18)$$

We remark on this channel, which like the periodic channel has long-term memory, in Sec. V.

III. CHANNEL WITHOUT A STRONG CONVERSE

The strong converse for the periodic quantum channel does not hold in general because the following inequality holds:

$$C_p < \overline{C_p}, \quad (3.1)$$

where

$$\overline{C_p} = \frac{1}{L} \sum_{i=0}^{L-1} \sup_P I(P, \Phi_i). \quad (3.2)$$

The strict inequality above can be shown explicitly for a periodic channel consisting of two branches of qubit amplitude-damping channels (see the Appendix below for detailed proof). On the other hand, equality for expression (3.1) can be shown to hold for a periodic channel with depolarizing channel branches [22].

Let us now investigate whether we can prove a full coding theorem for rates R such that

$$C_p < R < \overline{C_p}. \quad (3.3)$$

We first define the average probability of error as follows:

$$\overline{p_e} = \frac{1}{L} \sum_{i=0}^{L-1} p_e^i \leqslant \lambda, \quad (3.4)$$

where p_e^i denotes the probability of error for the ith channel branch.

Our coding strategy is as follows. We choose a code, that is, a particular encoding-decoding scheme, suitable for a particular channel branch labeled by the index $i \in \{0, \ldots, L-1\}$. Here a "branch" is defined as one term in the sum (2.14), that is, $\Phi_i^{(n)} = \Phi_i \otimes \Phi_{i+1} \otimes \cdots \otimes \Phi_{i+n-1}$. According to the coding theorem for memoryless channels, there is a code with error probability tending to zero for this branch with rate R. Indeed, for each j there exists a probability distribution P_j of states optimizing $\chi_j^* = \sup_P I(P; \Phi_j)$ and we can choose states from a typical subspace for these distributions, which can be interlaced at the positions $j - i + kL$, where $k \in [\frac{n}{L} - 1]$. The probability of choosing a particular branch correctly is given by $\frac{1}{L}$ and therefore the probability of error approaches

$$p_e = \frac{L-1}{L} < 1. \quad (3.5)$$

We thus have a λ-code for all $\lambda > 1 - \frac{1}{L}$. In particular, the error probability is bounded away from 1, and the *strong* converse does not hold.

On the other hand the strong converse does hold for $R > \overline{C_p}$. Indeed, the codewords can be decomposed into sub-codewords corresponding to the different stages of a period: $x^n = (x_0^n, \ldots, x_{L-1}^n)$, where the components of the x_i^n are understood to be interlaced in x^n. We distinguish types $P_0, \ldots, P_{L-1}$ for the sub-codewords. Then we have an analog of the single-type strong converse given by Lemma II.1.

Lemma III.1. For $\lambda \in (0,1)$ and $\delta > 0$ there exists $n_1(\lambda, \delta)$ such that for $n \geqslant n_1$, every (n, λ)-code for which all sub-codewords are of the same type $P_0, \ldots, P_{L-1}$, given that the ith branch is selected,

$$|\mathcal{M}_{n, P_0, \ldots, P_{L-1}}| \leqslant 2^{\frac{n}{L} \sum_{k=0}^{L-1} [I(P_k; \Phi_{i+k}) + \delta]}. \quad (3.6)$$

Clearly, for the complete channel, it follows that the number of codewords such that the sub-codewords are of types $P_0, \ldots, P_{L-1}$, satisfies

$$|\mathcal{M}_{n,P_0,\ldots,P_{L-1}}| \leqslant 2^{\frac{n}{L}\sum_{k=0}^{L-1}[\sup_P I(P;\Phi_{i+k})+\delta]}. \quad (3.7)$$

Summing over the types, we obtain the strong converse.

We can conclude that the strong converse holds for rates $R > \overline{C_p}$.

IV. A SCALE OF CAPACITIES

The above obviously raises the question of whether smaller error probabilities can be attained for smaller rates, but still above C_p. For this, we define a "pair capacity" $C_p^{(2)}$ as follows:

$$C_p^{(2)} = \frac{1}{2L} \max_{0 \leqslant i_1 < i_2 < L} \sum_{k=0}^{L-1} \sup_P [I(P;\Phi_{i_1+k}) + I(P;\Phi_{i_2+k})]. \quad (4.1)$$

Suppose the maximum is attained at a certain pair (i_1,i_2). With probability $2/L$, one of the two branches i_1 or i_2 is chosen. We attach a preamble to the code as in the proof of the product-state capacity of the periodic channel (2.16). If, for example, the branch i_1 is selected by the channel, the receiver can determine that this is the case by measuring the preamble and can then choose states for each value of k from the typical space corresponding to the maximizing distribution P_k for the CPT map Φ_{i_1+k}. This constitutes an encoding with rate given by the average of the mutual information $I(P;\Phi_{i_1+k})$ for $k = 0, \ldots, L-1$, which is greater than or equal to the pair capacity $C^{(2)}$ given by Eq. (4.1). We have thus constructed a λ-code for $\lambda > 1 - \frac{2}{L}$.

On the other hand, let $R > C_p^{(2)}$ and suppose that (C^n, E^n) is a sequence of (n,λ)-codes with $\lambda < 1 - \frac{1}{L}$, and assume that

$$\frac{1}{n}\log_2|\mathcal{M}_n| \geqslant R > C_p^{(2)}. \quad (4.2)$$

First note that we may assume that the number of codewords with sub-codewords of types $P_0, \ldots, P_{L-1}$ is bounded by

$$\frac{1}{n}\log_2|\mathcal{M}_{n,P_0,\ldots,P_{L-1}}| \leqslant \frac{1}{L}\sum_{k=0}^{L-1}[I(P_k,\Phi_{i+k})+\delta] \quad (4.3)$$

for some *fixed* $i = 0, \ldots, L-1$. Indeed, otherwise, by Lemma III.1, $p_e^i > \lambda$ for all i and hence $p_e > \lambda$.

We now claim that for every other $j \neq i$, and $\epsilon > 0$ small enough,

$$\frac{1}{n}\log_2|\mathcal{M}_{n,P_0,\ldots,P_{L-1}}| > \frac{1}{L}\sum_{k=0}^{L-1}[I(P_k,\Phi_{j+k})+\epsilon]. \quad (4.4)$$

If this were not the case then the pair capacity for a single type P_k can be written as

$$\frac{1}{n}\log_2|\mathcal{M}_{n,P_0,\ldots,P_{L-1}}| \leqslant \frac{1}{2L}\sum_{k=0}^{L-1}[I(P_k,\Phi_{i+k}) + I(P_k,\Phi_{j+k})+\delta] \quad (4.5)$$

and hence

$$\frac{1}{n}\log_2|\mathcal{M}_{n,P_0,\ldots,P_{L-1}}| \leqslant \frac{1}{2L}\sum_{k=0}^{L-1}\big(\sup_P\{I(P,\Phi_{i+k}) + I(P,\Phi_{j+k})\} + \delta\big). \quad (4.6)$$

Summing over the types $P_0, \ldots, P_{L-1}$ leads to a contradiction with (4.2).

Now, expression (4.4) implies with Lemma III.1 that, if the jth branch is selected by the channel, then the error probability $p_e^j > 1 - \eta$ for any $\eta > 0$. Since with probability $1 - \frac{1}{L}$ one of the branches j other than i is selected, we conclude that the error probability $p_e > (1 - \frac{1}{L})(1 - \eta) > \lambda$ if $\eta < 1 - \frac{1}{L} - \lambda$ is small enough.

It is now clear that this argument can be generalized to prove the following theorem.

Theorem IV.1. Define, for $r = 1, \ldots, L$, a *scale* of capacities $C_p^{(r)}$ by

$$C_p^{(r)} = \frac{1}{rL} \max_{0 \leqslant i_1 < \cdots < i_r < L} \sum_{k=0}^{L-1} \sup_P \sum_{m=1}^r I(P,\Phi_{i_m+k}). \quad (4.7)$$

(Note that $C_p^{(1)} = C_p$ and $C_p^{(L)} = \overline{C_p}$.) Then, if $\lambda > 1 - \frac{r}{L}$ and $R < C_p^{(r)}$, there exists a sequence of (n,λ)-codes with rate R. Conversely, if $\lambda < 1 - \frac{r-1}{L}$, there exists no sequence of (n,λ)-codes with rate $R > C_p^{(r)}$.

V. THE RANDOM CHANNEL

The situation for the random channel is similar, but more complicated due to the fact that different branches can have different probabilities q_i. We can, in general, distinguish break points at values of the error probability given by

$$q(\Delta) = \sum_{i \in \Delta} q_i, \quad \Delta \subset \{1, \ldots, M\}. \quad (5.1)$$

We have an analog of the detailed theorem for periodic channels above.

Theorem V.1. Define, for $\Delta \subset \{1, \ldots, M\}$, a *scale* of capacities C_p^Δ by

$$C_p^\Delta = \sup_P \min_{i \in \Delta} I(P;\Phi_i). \quad (5.2)$$

Then, if $\lambda > 1 - q(\Delta)$ and $R < C_p^\Delta$, there exists a sequence of (n,λ)-codes with rate R.

For the converse to the theorem, we introduce another scale as follows:

$$\overline{C_p^\Delta} = \sup_P \max_{i \in \Delta} I(P;\Phi_i). \quad (5.3)$$

Then, if $\lambda < 1 - q(\Delta)$ there exist no (n,λ)-codes with rate $R > \overline{C_p^\Delta}$.

The situation is less clear cut than it seems, however. In fact, not every $q(\Delta)$ is necessarily a point of discontinuity for the capacity, because C_p^Δ is, in general, not monotonic in the probabilities $q(\Delta)$!

VI. DISCUSSION

One of the most surprising and interesting results which has emerged from Shannon theory is the observation that the *strong* information-carrying capacity of a memoryless channel is independent of the upper bound on the maximum error probability of that channel, usually denoted λ. The independence of the parameter λ is crucial to the existence of a so-called *strong* capacity for the channel [3].

The dependency of some channel capacities on this parameter λ, including nonstationary discrete memoryless classical channels, led to the definition of a capacity function [5]. Note that recently Ahlswede [23] proved that the capacity functions can now be thought of as so-called capacity-sequences.

For the case of the quantum periodic channel, and also the random channel, we have shown that an analogous parameter-dependent capacity can be defined, which takes the form of a scale of capacities applicable for various ranges of the error parameter.

Note added. It appears that similar results to ours were obtained by Datta, Hsieh, and Brandão [24], using different methods.

APPENDIX: THE PERIODIC CHANNEL WITH AMPLITUDE-DAMPING CHANNEL BRANCHES

The qubit amplitude-damping channel acting on the state $\rho = \left(\begin{smallmatrix} a & b \\ \overline{b} & 1-a \end{smallmatrix}\right)$ is given by

$$\Phi_{\text{amp}}(\rho) = \begin{pmatrix} a + (1-a)\gamma & b\sqrt{1-\gamma} \\ \overline{b}\sqrt{1-\gamma} & (1-a)(1-\gamma) \end{pmatrix}. \quad (A1)$$

The expression for the product-state capacity of the qubit amplitude-damping channel is given as follows,

$$\begin{aligned}
\chi(&\Phi_{\text{amp}}(\{p_j, \rho_j\})) \\
&= S\left[\sum_j \begin{pmatrix} p_j[a_j + (1-a_j)\gamma] & p_j b_j \sqrt{(1-\gamma)} \\ p_j \overline{b}_j \sqrt{(1-\gamma)} & p_j(1-a_j)(1-\gamma) \end{pmatrix}\right] \\
&\quad - \sum_j p_j\, S \begin{pmatrix} a_j + (1-a_j)\gamma & b_j\sqrt{1-\gamma} \\ \overline{b}_j\sqrt{1-\gamma} & (1-a_j)(1-\gamma) \end{pmatrix}.
\end{aligned} \quad (A2)$$

We now investigate whether the following equation holds for a periodic channel with two amplitude-damping channel branches:

$$\frac{1}{2}\sup_{\{p_j,\rho_j\}}\sum_{i=0}^{1}\chi_i(\{p_j,\rho_j\}) = \frac{1}{2}\sum_{i=0}^{1}\sup_{\{p_j,\rho_j\}}\chi_i(\{p_j,\rho_j\}). \quad (A3)$$

Let γ_0 and γ_1 represent the error parameters for two amplitude-damping channels Φ_0 and Φ_1, respectively. We have argued [25] that the Holevo quantity for the qubit amplitude-damping channel can be increased using an ensemble containing two mirror-image pure states, each with probability $\frac{1}{2}$. Using this minimal ensemble we investigate both sides of Eq. (A3), for a periodic channel with two qubit amplitude-damping channel branches.

Clearly the *left-hand side* of Eq. (A3) will be attained for a single parameter which we denote by a_{max}. However,

the *right-hand side* of Eq. (A3) cannot be obtained by a single a_{max}. Instead, the supremum for each channel will be attained at a different value of the input state parameter a. We denote by a_{max_0} and a_{max_1} the state parameter that achieves the product-state capacity for the channels Φ_0 and Φ_1, respectively. Let $\chi_0(a)$ and $\chi_1(a)$ denote the Holevo quantities of the channels Φ_0 and Φ_1, respectively. Denoting $x_i = \sqrt{1 - 4\gamma_i(1-\gamma_i)(1-a^2)}$ the eigenvalues for each of the amplitude-damping channels can be written as

$$\lambda_{\text{amp}_i\pm} = \tfrac{1}{2}(1 \pm \sqrt{1 - 4\gamma_i(1-\gamma_i)(1-a)^2}). \quad (A4)$$

The values for a_{max_0} and a_{max_1} can be determined by separately solving the following equation for each channel:

$$\begin{aligned}
\frac{d\chi_i(a)}{da} &= (1-\gamma_i)\ln\left(\frac{(1-a)(1-\gamma_i)}{a + (1-a)\gamma_i}\right) \\
&\quad + \frac{2\gamma_i(1-\gamma_i)(1-a)}{x_i}\ln\left(\frac{1+x_i}{1-x_i}\right) \\
&= 0. \quad\quad (A5)
\end{aligned}$$

Let $\chi_{\text{avg}}^*(\gamma_0, \gamma_1, a_{\text{max}_0}, a_{\text{max}_1})$ denote the average of the supremum of the Holevo capacities of the channels Φ_0 and Φ_1, that is,

$$\chi_{\text{avg}}^*(\gamma_0, \gamma_1, a_{\text{max}_0}, a_{\text{max}_1}) = \tfrac{1}{2}\left[\chi_0^*(a_{\text{max}_0}) + \chi_1^*(a_{\text{max}_1})\right]. \quad (A6)$$

It is not difficult to show that

$$\chi^*(\gamma_0 = 1, \gamma_1, a_{\text{max}}) = \chi_{\text{avg}}^*(\gamma_0 = 1, \gamma_1, a_{\text{max}_0}, a_{\text{max}_1}). \quad (A7)$$

Similarly, we can show that

$$\chi^*(\gamma_0, \gamma_1 = 1, a_{\text{max}}) = \chi_{\text{avg}}^*(\gamma_0, \gamma_1 = 1, a_{\text{max}_0}, a_{\text{max}_1}).$$

Next, we show separately for (a) $\gamma_i = 0$ and for (b) $0 < \gamma_i < 1$ that the following inequality holds:

$$\chi^*(\gamma_0, \gamma_1, a_{\text{max}}) < \chi_{\text{avg}}^*(\gamma_0, \gamma_1, a_{\text{max}_0}, a_{\text{max}_1}). \quad (A8)$$

(a) Taking $\gamma_0 = 0$, the expression $\chi^*(\gamma_0, \gamma_1, a_{\text{max}})$ becomes

$$\begin{aligned}
\chi^*(\gamma_0 = 0, \gamma_1, a_{\text{max}}) &= H_{\text{bin}}(a_{\text{max}_1}) + H_{\text{bin}}\left[(1-a_{\text{max}_1})(1-\gamma_1)\right] \\
&\quad - S[\Phi_1(\rho_{a\text{max}})]. \quad (A9)
\end{aligned}$$

Denoting $\chi_{\text{avg}}^*(\gamma_0, \gamma_1, a_{\text{max}_0}, a_{\text{max}_1})$ by $\chi_{\text{avg}}^*(\gamma_1)$ the right-hand side becomes

$$\begin{aligned}
\chi_{\text{avg}}^*(\gamma_1, a_{\text{max}_1}) &= H_{\text{bin}}(a_{\text{max}_1}) + H_{\text{bin}}\left[(1-a_{\text{max}_1})(1-\gamma_1)\right] \\
&\quad - S[\Phi_1(\rho_{a\text{max}_1})]. \quad (A10)
\end{aligned}$$

Clearly, $a_{\text{max}_0} = \frac{1}{2}$. To show that $a_{\text{max}} < a_{\text{max}_1}$, we must show that $\frac{d}{da}\sum_i \chi_i(a) < 0$ at $a = a_{\text{max}_1} = \frac{1}{2}$.

For $\gamma_0 = 0$ the Holevo quantity of the channel Φ_0 becomes

$$\chi_0(a) = S\begin{pmatrix} a & 0 \\ 0 & (1-a) \end{pmatrix} - S(\rho). \quad (A11)$$

However, ρ is a pure state and therefore $S(\rho) = 0$. Therefore, from Eq. (A5),

$$\frac{d\chi_0(a)}{da} = \ln\left(\frac{(1-a)}{a}\right). \quad (A12)$$

We have previously shown that the maximizing state parameter for the amplitude-damping channel is achieved at $a \geqslant \frac{1}{2}$ [25].

We are considering the case where $\gamma_0 \neq \gamma_1$, that is, $\gamma_1 \neq 0$; therefore, $a_{\max_1} > \frac{1}{2}$. The expression $\chi_0(a)$ now represents the binary entropy, $H(a)$, and is therefore maximized at $a = \frac{1}{2}$. It was shown above that the entropy $S(a)$ is a strictly concave function for $\gamma_0 = 0$ and $\chi_0(a)$ is therefore decreasing at $a = a_{\max_1}$.

The capacity $\chi_1^*(a)$ is achieved at $a = a_{\max_1}$. Therefore, $\frac{d\chi_1(a)}{da}$ is equal to zero at this point.

We can now conclude that $\frac{d}{da} \sum_i \chi_i(a) < 0$ when $a = a_{\max_1}$ and therefore

$$\chi^*(\gamma_0 = 0, \gamma_1, a_{\max}) < \chi_{\text{avg}}^*\left(\gamma_0 = 0, \gamma_1, a_{\max_0}, a_{\max_1}\right). \quad \text{(A13)}$$

(b) We now show that an inequality exists between the expressions $\chi^*(\gamma_0, \gamma_1, a_{\max})$ and $\chi_{\text{avg}}^*(\gamma_0, \gamma_1, a_{\max_0}, a_{\max_1})$ for fixed γ_0, such that $0 < \gamma_0 < 1$.

In Ref. [25] we proved that if $\gamma_0 < \gamma_1$, then $\chi(\gamma_0) > \chi(\gamma_1)$, and therefore $a_{\max_0} < a_{\max_1}$. Therefore, $\frac{d\chi_0(a)}{da} < 0$ at $a = a_{\max_1}$ and $a_{\max} < a_{\max_1}$. Similarly, if $\gamma_0 > \gamma_1$, then $a_{\max_0} > a_{\max_1}$ and $\frac{d\chi_0(a)}{da} > 0$ at $a = a_{\max_1}$ and $a_{\max} > a_{\max_1}$.

As a result, $a_{\max}$ will always lie in between $a_{\max_0}$ and $a_{\max_1}$. We have previously shown [25] that the Holevo quantity for the qubit amplitude-damping channel is concave in its single state parameter. Therefore $a_{\max} > \tilde{a}$, where $\tilde{a}$ is the parameter value associated with $\chi_{\text{avg}}^*(\gamma, \gamma_1, a_{\max_0}, a_{\max_1})$, that is, $\sum_i \sup_a \chi_i(a) = \chi_{\gamma_0, \gamma_1}^*(\tilde{a})$. This proves that $\chi^*(\gamma_0, \gamma_1, a_{\max}) < \chi_{\text{avg}}^*(\gamma, \gamma_1, a_{\max_0}, a_{\max_1})$.

In conclusion, if $\gamma_0 = 1$ or $\gamma_1 = 1$, then $a_{\max} = a_{\max_0}$ or $a_{\max} = a_{\max_1}$, respectively, and $\chi^*(\gamma_0, \gamma_1, a_{\max}) = \chi_{\text{avg}}^*(\gamma, \gamma_1, a_{\max_1}, a_{\max_1})$. However, if $\gamma_0, \gamma_1 \neq 1$, then $\chi^*(\gamma_0, \gamma_1, a_{\max}) < \chi_{\text{avg}}^*(\gamma, \gamma_1, a_{\max_1}, a_{\max_1})$.

---

[1] C. E. Shannon, Bell Syst. Tech. J. **27**, 379, 623 (1948).

[2] A. Feinstein, IRE Trans. Inf. Theory PGIT **4**, 2 (1954).

[3] J. Wolfowitz, *Coding Theorems of Information Theory* (Springer, Berlin, 1964).

[4] J. Wolfowitz, Inf. Control **6**, 49 (1963).

[5] R. Ahlswede, in *General Theory of Information Transfer and Combinatorics* (Springer, Berlin, Heidelberg, 2006), Vol. 4123 of Lecture Notes in Computer Science, pp. 639–663.

[6] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).

[7] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **45**, 2486 (1999).

[8] I. Bjelaković and H. Boche, in *Proceedings of the IEEE Information Theory Workshop* (IEEE, porto, Portugal, 2008).

[9] I. Bjelaković and H. Boche, IEEE Trans. Inf. Theory **55**, 3360 (2009).

[10] R. König and S. Wehner, Phys. Rev. Lett. **103**, 070504 (2009).

[11] S. Bose, Phys. Rev. Lett. **91**, 207901 (2003).

[12] D. Kretschmann and R. Werner, Phys. Rev. A **72**, 062323 (2005).

[13] S. Mancini, J. Phys.: Conf. Ser. **36**, 121 (2006).

[14] A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).

[15] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997).

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. (Cambridge University Press, Cambridge, 2011).

[17] J. R. Norris, *Markov Chains* (Cambridge University Press, Cambridge, 1997).

[18] N. Datta and T. C. Dorlas, J. Stat. Phys. **134**, 1173 (2009).

[19] G. Bowen and S. Mancini, Phys. Rev. A **69**, 012306 (2004).

[20] C. Morgan, Ph.D. thesis, University College Dublin, 2010.

[21] N. Datta and T. Dorlas, J. Phys. A **40**, 8147 (2007).

[22] T. Dorlas and C. Morgan, Phys. Rev. A **79**, 032320 (2009).

[23] R. Ahlswede, in *Proceedings of the IEEE Information Theory Workshop* (IEEE, Dublin, Ireland, 2010).

[24] N. Datta, M. Hsieh, and F. Brandão, e-print arXiv:1106.3089.

[25] T. Dorlas and C. Morgan, Int. J. Quantum Inf. **6**, 745 (2008).