Forgetting boosts the classical environment-assisted and private capacity

David Elkouss\(^1\) and Sergii Strelchuk\(^2\)

\(^1\)QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
\(^2\)Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, U.K.

A channel capacity is non-convex if the capacity of a mixture of different quantum channels exceeds the mixture of the individual capacities. This implies that there is a concrete communication scenario in which a sender can increase the transmission rate by forgetting which channel acts on the channel input. Previously, this surprising property had only been shown for the quantum capacity. Here we prove that both the private and classical environment-assisted capacities are non-convex.

### Communication tasks

The action of a quantum channel can always be defined by an isometry \( V \) that takes the input system \( A' \) to the output \( B \) together with an auxiliary system called the environment \( E: \mathcal{N}^{A'\rightarrow B}(\rho^{A'}) = \text{tr}_E V^{A'\rightarrow BE} \rho^{A'} (V^{A'\rightarrow BE})\dagger. \) This isometry allows to define the action of the complementary channel:

\[
\begin{array}{c|c|c|c}
\text{Computability} & \text{Additivity} & \text{Convexity} \\
\hline
\mathcal{Q} &?&\text{No} &\text{[17]} & \text{No} &\text{[17]} \\
\mathcal{P} &?&\text{No} &\text{[18,19]}& \text{No} \\
\mathcal{C} &?&?&? \\
\mathcal{C}_c &\text{Yes} &\text{[3]} & \text{Yes} &\text{[3,4]} \\
\mathcal{C}_H &?&\text{No} &\text{[14]} & \text{No} \\
\end{array}
\]

**TABLE I.** Main properties of quantum channel capacities: convexity, additivity and computability. We consider quantum capacity \( \mathcal{Q} \), private capacity \( \mathcal{P} \) and unassisted and environment-assisted classical capacities, \( \mathcal{C} \) and \( \mathcal{C}_H \) respectively.

\( N_1 \) and \( N_2 \) and \( p \in (0,1) \) such that:

\[
pT(N_1) + (1-p)T(N_2) < T(pN_1 + (1-p)N_2).
\]

Non-convexity is a particularly surprising property especially in connection to the following two scenarios depicted on Fig. [1] In the first case, corresponding to the left hand side of [1], Alice has access to two channels separately: she chooses \( N_1 \) with probability \( p \) and \( N_2 \) with probability \( 1-p \). The encoding over the two channels is independent. In the second case, corresponding to the right hand side, Alice has no control over which of the channels is applied; instead a black box applies them at random with the same probabilities \( p \) and \( 1-p \). Another way of looking at the second scenario is that Alice can choose between both channels but then she forgets which one she applied. The difference between the two scenarios lies in the fact that Alice loses all the control over the applied channel which intuitively should severely handicap her ability to transmit information.

Here, we report that private capacity and classical environment-assisted classical capacities are non-convex and thus Alice could send more information in the latter setting where the adverse effects of noise are amplified.
The system evolves by a superset of input states purified with a reference system. Let $\rho^A$ be a quantum state, we denote by $H(A) = -tr\rho \log \rho$ the von Neumann entropy. Let $\rho^{AB}$ be a bipartite quantum state, we denote by $I(A;B) = H(A) + H(B) - H(AB)$ the mutual information between the systems $A$ and $B$.

We are interested in the following communication tasks and the associated channel capacities.

The first task is the transmission of quantum information. The quantum capacity characterizes the ability of a quantum channel for this task in the absence of additional resources. [3][8][10]

$$Q(N) = \lim_{n \to \infty} \frac{1}{n} Q^{(1)}(N^{\otimes n}),$$

where $Q^{(1)}(N) = \max_{\phi AA'} Q^{(1)}(N, \phi AA')$ is the coherent information of a quantum channel. The maximum is taken over all input states purified with a reference system $A$. The quantity $Q^{(1)}(N, \phi AA') = H(B) - H(AB)$, where $H(AB), H(AB)$ are the von Neumann entropies of $\rho^{A'} = N(tr_A \phi AA'), \rho^{AB} = \text{id}^A \otimes N^{A' \to B}(\phi AA')$ and $\text{id}$ denotes the identity channel.

For some channels the coherent information is additive and thus exactly characterizes their capacity. In these cases it is possible to compute the capacity exactly [20]. However there are examples when this is not the case [21][22]: coherent information is superadditive, but also the quantum capacity itself is superadditive [17][23] – there exist pairs of channels such that their joint capacity is strictly larger than the sum of their capacities.

The second task is the transmission of private classical information. The capacity of a channel for this task without additional resources is called the private capacity [9][10]. We define the private information to be

$$P^{(1)}(N) = \max_{\nu \in P_d} \max_{\rho^{X \otimes N^{A'-B}}} I(X;B) - I(X;E),$$

where $I(X;B)$ and $I(X;E)$ are evaluated on the states $\text{id}^X \otimes N^{A' \to B}(\sum x p_x |x\rangle \langle x| \otimes \rho^{A'})$ and $\text{id} \otimes N^{A' \to E}(\sum x p_x |x\rangle \langle x| \otimes \rho^{A'})$. The private capacity is given by the regularization of the private information

$$P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N^{\otimes n}).$$

Both private information [24][20] and the private capacity [18][19][27] were found to be superadditive.

The third task is the transmission of classical information. The classical capacity [6][7] characterizes the capacity of a quantum channel for transmitting classical information without additional resources. To characterize the classical capacity we first define the Holevo information

$$C^{(1)}(N) = \max \sum_{p_x|x} I(x;B).$$

The classical capacity is given by the regularization of the Holevo information

$$C(N) = \lim_{n \to \infty} \frac{1}{n} C^{(1)}(N^{\otimes n}).$$

Holevo information is superadditive [28] but it is a challenging open question whether or not the classical capacity verifies any of the three properties of convexity, additivity and computability.

In some scenarios sender and receiver may share additional resources which they can leverage to increase their communication rates. The capacities of a channel for a communication task assisted by additional resources turn out to have completely different properties than their unassisted counterparts. One such example is shared entanglement. The entanglement-assisted classical capacity of a quantum channel $C_e(N)$ is both convex and additive and can be computed efficiently [4].

Alternatively, one may consider the environment of the channel as a friendly helper that ‘assists’ the sender during information transmission [29]. This third party can input states independently of the sender or even interact with the sender by exchanging messages. This gives rise to a host of environment-assisted classical capacities depending on whether we have active or passive environment assistance [13] or whether the sender and environment are allowed to share entanglement or interact by means of local operations and classical communication. In our work we focus on the weakest variant of assistance for classical communication when the helper is in the product state with the sender [14]. The corresponding capacity is given by

$$C_{\eta}(N) = \lim_{n \to \infty} \frac{1}{n} \max_{\eta \in P_{d}} C^{(1)}(N^{\otimes n}),$$

where $N^{\otimes n}(\rho) = \text{tr}_E W^{\otimes n}(\rho \otimes \eta)(W^{\otimes n})$. $W^{A' \to E}$ is an isometric extension of the channel such that $N^{A' \to B}(\rho^{A'}) = \text{tr}_F W^{\otimes n}(\rho^{A'} \otimes |0\rangle \langle 0|^E W^{\otimes n})$ and $\eta$ is a state of the system $E$ over $n$ uses of the channel.

**Private capacity.** We first show that private capacity is non-convex. Let us first define two families of channels. The first is the $d$-dimensional erasure channel $E_{d,p}$. Its action is defined as follows:

$$E_{d,p}(\rho) = (1-p)\rho + p|e\rangle \langle e|.$$ 

That is, $E_{d,p}$ takes the input to the output with probability $1-p$ and with probability $p$ it outputs an erasure flag. The private capacity of the erasure channel is known to be [20]:

$$P(E_{d,p}) = \max \{0, (1 - 2p) \log d\}.$$ 

The second is the ‘rocket channel’ $R_d$. It was introduced by Smith and Smolin in [19]. It takes two $d$-dimensional inputs that we label $C$ and $D$. The channel chooses two unitaries $U$ and $V$ at random [30] and applies
The intuition behind this fact is that if Alice knew $U_D$ to compensate for the dephasing of erasure channel and a flagged rocket channel:

$$\text{The average} N_{q,d,p} \text{ transmit classical information at a rate } \log d \text{ strictly higher rate in the scenario below.}$$

Below, a black box chooses the channel for Alice (with the same probabilities). Non-convexity implies that Alice might communicate at a strictly higher rate in the scenario. Below, a black box chooses the channel for Alice (with the same probabilities). Non-convexity implies that Alice might communicate at a strictly higher rate in the scenario below.

Then:

$$\mathcal{P}(N_{q,d,p}) \geq \mathcal{Q}(N_{q,d,p}) \geq \frac{1}{2} \mathcal{Q}^{(1)} \left( N_{q,d,p}^{2}, \rho \right).$$

Now, let the input be:

$$\rho^{A^1A^2C_1D_1C_2D_2} = \Phi^{A^1D_1} \otimes \Phi^{C_1C_2} \otimes \Phi^{A^2D_2},$$

where $\Phi^{AB}$ represents a maximally entangled state between systems $A$ and $B$. We use a subscript if the register corresponds to a concrete channel use and a superscript to number the subsystem: $C_1^A$ stands for the first subsystem of the register $C$ in the second use of the channel and $A^2$ the second subsystem of an auxiliary register $A$.

Now we analyze the coherent information achieved by the input in (15). After sending $\rho$ through the channel, the resulting state is:

$$\mathcal{N}_{q,d,p}^{2} = q^2 \left( \mathcal{E}_{d^2,p} \otimes \mathcal{E}_{d^2,p} \right)(\rho) \otimes |0\rangle \langle 0| \otimes |0\rangle \langle 0|$$

$$+ q(1-q) \left( \mathcal{E}_{d^2,p} \otimes \mathcal{R}_d \right)(\rho) \otimes |0\rangle \langle 0| \otimes |1\rangle \langle 1|$$

$$+ (1-q)q \left( \mathcal{R}_d \otimes \mathcal{E}_{d^2,p} \right)(\rho) \otimes |1\rangle \langle 1| \otimes |0\rangle \langle 0|$$

$$+ (1-q)^2 \left( \mathcal{R}_d \otimes \mathcal{R}_d \right)(\rho) \otimes |1\rangle \langle 1| \otimes |1\rangle \langle 1|. \quad (19)$$

Since the channel is a flagged combination of $\mathcal{E}$ and $R$, the coherent information is just the weighed sum of four terms

$$\mathcal{Q}^{(1)}(N_{q,d,p}^{2}, \rho) = q^2 \mathcal{Q}^{(1)}(\mathcal{E}_{d^2,p} \otimes \mathcal{E}_{d^2,p}, \rho)$$

$$+ q(1-q) \mathcal{Q}^{(1)}(\mathcal{R}_d \otimes \mathcal{E}_{d^2,p}, \rho)$$

$$+ (1-q)q \mathcal{Q}^{(1)}(\mathcal{E}_{d^2,p} \otimes \mathcal{R}_d, \rho)$$

$$+ (1-q)^2 \mathcal{Q}^{(1)}(\mathcal{R}_d \otimes \mathcal{R}_d, \rho). \quad (20)$$

By symmetry of the input state one has

$$\mathcal{Q}^{(1)}(\mathcal{E}_{d^2,p} \otimes \mathcal{R}_d, \rho) = \mathcal{Q}^{(1)}(\mathcal{R}_d \otimes \mathcal{E}_{d^2,p}, \rho). \quad (21)$$

Let us compute each of the three terms. First we consider two erasure channels. The resulting state is

$$\left( \Phi^{A^1A^2C_1D_1B_1^2} \otimes \Phi^{C_2D_2B_2^2} \right) (\rho)$$

$$= (1-p)^2 \Phi^{A^1B_1^2} \otimes \Phi^{B_2^2} \otimes \Phi^{A^2B_2^2}$$

$$+ p(1-p) \Phi^{A^1B_1^2} \otimes \pi B_2^2 \otimes \pi A^2 \otimes |e\rangle \langle e| B_2^2$$

$$+ p(1-p) \Phi^{A^2B_2^2} \otimes \pi B_1^2 \otimes \pi A^1 \otimes |e\rangle \langle e| B_1^2$$

$$+ p^2 \Phi^{A^1A^2} \otimes |e\rangle \langle e| B_1^2 \otimes |e\rangle \langle e| B_2^2. \quad (22)$$
where $\pi$ stands for the maximally mixed state. The four states of this mixture can be differentiated by checking the erasure flag. This implies that the coherent information can also be divided into the sum of the coherent information of each term.

$$Q^{(1)}(\mathcal{E}_{d^2,p} \otimes \mathcal{E}_{d^2,p}, \rho) = (1-p)^2 2 \log d$$
$$+ 2p(1-p)0 + p^2 (-2\log d)$$
$$= (1-2p)2\log d.$$  

(23)

The resulting state in the case of one erasure channel and one rocket channel is

$$(R_d \otimes \mathcal{E}_{d^2,p})(\rho) = (1-p)(R_d \otimes I)(\rho) + p(R_d \otimes \mathcal{E}_{d^2,1})(\rho)$$

(25)

which yields

$$Q^{(1)}(R_d \otimes R_d, \rho) = (2 - 3p) \log d.$$  

(26)

Finally, the use of two rocket channels yields

$$Q^{(1)}(R_d \otimes R_d, \rho) \geq 0.$$  

(27)

For justification of (26) and (27), see the Supplemental material.

We plug (23), (26), and (27) back into (20)

$$Q^{(1)}(\mathcal{N}^{\otimes 2}_{q,d,p}, \rho) = 2q((1-q)(2-3p)+q(1-2p)) \log d$$

(28)

and we obtain that

$$\mathcal{P}(\mathcal{N}_{q,d,p}) \geq \frac{1}{2}Q^{(1)}(\mathcal{N}^{\otimes 2}_{q,d,p}, \rho)$$

$$\geq q ((1-q)(2-3p)+q(1-2p)) \log d.$$  

(30)

It remains to compare the achievable bound in (30) with the converse bound in (15). For any triple $(q,d,p)$ such that (30) is strictly greater than (15) the private capacity is non-convex. Figure 2 depicts the achievable region for which we exhibit non-convexity.

**Classical environment-assisted capacity.** We turn to proving the second claim – the non-convexity of classical capacity with the weakest environment assistance. We start with providing two channels and a special entangled input state which is used to demonstrate this effect. Consider a flagged combination of the two channels used in [14] to show superadditivity of $\mathcal{C}_H$.

The first channel is defined by a controlled unitary

$$V^{A \rightarrow E \rightarrow FB} = \sum_{x,z} |xz\rangle^F |xz\rangle^A \otimes (W(x,z))^{E \rightarrow B}$$

where $W(x,z) = X(x)Z(z)$, $X(x)j \mod d$, $Z(z)j \mod d$ and $\omega$ is again the primitive $d$-th root of unity.

The second channel is a SWAP channel: \text{SWAP}(|\phi \rangle^A \otimes |\psi \rangle^E) = |\psi \rangle^B \otimes |\phi \rangle^F.

Thus, our channels will have the form $\mathcal{N}_1 = |0\rangle\langle 0| \otimes V^{A \rightarrow E \rightarrow BF}$ and $\mathcal{N}_2 = |1\rangle\langle 1| \otimes \text{SWAP}^{A \rightarrow E \rightarrow BF}$. Fix $|A| =$

FIG. 2. The figure shows the difference between [30] and [15] normalized by $\log d$ when $d$ goes to infinity. A value larger than zero implies non-convexity of $\mathcal{P}$.

$|F| = d^2, |E| = d, |B| = d$. In the following we prove that for some range of $p$:

$$\mathcal{C}_H(p\mathcal{N}_1 + (1-p)\mathcal{N}_2) > p\mathcal{C}_H(\mathcal{N}_1) + (1-p)\mathcal{C}_H(\mathcal{N}_2).$$  

(31)

It follows from [14] that $\mathcal{C}_H(\mathcal{N}_1) = \log d$ and $\mathcal{C}_H(\mathcal{N}_2) = 0$. Hence, the right hand side of (31) is bounded from above by

$$p\mathcal{C}_H(\mathcal{N}_1) + (1-p)\mathcal{C}_H(\mathcal{N}_2) \leq p \log d.$$  

(32)

In order to bound from below the left hand side of (31), consider two uses of the channel $\mathcal{M} = p\mathcal{N}_1 + (1-p)\mathcal{N}_2$. Let the state of the environment be the maximally entangled state between $E_1$ and $E_2$: $\Phi^{E_1E_2}$ and the input state to the channel:

$$\rho^{X_{A_1A_2}} = \frac{1}{d^2} \sum_{i,j=0}^{d-1} |ij\rangle\langle ij|^{X} \otimes |ij\rangle\langle ij|^{A_1} \otimes |ij\rangle\langle ij|^{A_2}.$$  

(33)

Then,

$$\mathcal{C}_H(\mathcal{M}^{\otimes 2}) \geq I(X : B_1B_2)_{\mathcal{M}^{\otimes 2}(\rho)},$$  

(34)

and since $\mathcal{M}$ is flagged, we can also divide the mutual information into the sum of the mutual information associated with each channel action. Let us compute the
corresponding output states:

\[ N_1^{\otimes 2}(\rho) = \frac{1}{d^2} \sum_{i,j=0}^{d-1} |ij\rangle \langle ij|^X \otimes Z(j) \otimes Z(j)(\Phi_{B_1B_2}). \]  

(35)

\[ N_1 \otimes N_2(\rho) = \frac{1}{d^2} \sum_{i,j=0}^{d-1} |ij\rangle \langle ij|^X \otimes \text{id} \otimes W(i,j) (\Phi_{B_1B_2}). \]  

(36)

\[ N_2^{\otimes 2}(\rho) = \frac{1}{d^2} \sum_{i,j=0}^{d-1} |ij\rangle \langle ij|^X \otimes \Phi_{B_1B_2}. \]  

(37)

Note that \( N_2 \otimes N_1(\rho) \) is just \( N_1 \otimes N_2(\rho) \) with \( B_1 \) and \( B_2 \) swapped. The state obtained from the action of \( N_1^{\otimes 2}(\rho) \) follows from the observation that \( W(x,z) \otimes W(x,z)\Phi = \text{id} \otimes W(x,z)^TW(x,z)\Phi = Z(j) \otimes Z(j)\Phi \).

It is easy to verify that \( I(X;B_1B_2) \) vanishes when \( N_2 \otimes N_2 \) is applied and takes the value \( 2 \log d \) when either \( N_2 \otimes N_1 \) or \( N_1 \otimes N_2 \) is applied. In the case of \( N_1 \otimes N_1 \) (see Supplemental material for details) we can bound the mutual information by:

\[ I(X;B_1B_2)_\rho = \begin{cases} 
\log d & \text{if } d \text{ is odd} \\
2 \log d & \text{if } d \text{ is even} 
\end{cases} \]  

(38)

Adding all the contributions we obtain for odd \( d \):

\[ C_H(\mathcal{M}) \geq \frac{1}{2} \left( 2p(1-p)2 \log d + p^2 \log d \right) = \left( 2p - \frac{3}{2}b^2 \right) \log d. \]  

(39)

Comparing the achievable bound in (39) with the converse bound in (32) one observes that for odd \( d > 1 \) and \( 0 < p < \frac{2}{3} \) the classical capacity with passive environment-assisted capacity is non-convex.

**Discussion**

The classical capacity of classical channels is efficiently computable, convex and additive. Quantum channels can be used for several different communication tasks. In particular, for the transmission of quantum information, exchanging private classical bits and classical communication. One may endow a quantum channel with auxiliary resources but with the exception of the entanglement-assisted classical capacity the corresponding channel capacities do not share the same properties which make their classical counterparts efficiently computable.

We have focused our attention in non-convexity. This is a very curious property since, operationally, it can be interpreted as a sender Alice being able to transmit at higher rates by losing control over or forgetting which channel is applied to the input state. Prior to this work, non-convexity had only been proven for the quantum capacity. Here, we have shown that also the private capacity and the classical environment-assisted classical capacity are non-convex.

**Acknowledgements:** We thank Kenneth Goodeough, Jonas Helsen and Stephanie Wehner for useful discussions and feedback. SS acknowledges the support of Sidney Sussex College and European Union under project QALGO (Grant Agreement No. 600700). DE has been partially supported by STW, the NWO Vidi grant “Large quantum networks from small quantum devices” and by the project HyQuNet (Grant No. TEC2012-35673), funded by Ministerio de Economía y Competitividad (MINECO), Spain.
Supplemental Material

Forgetting boosts the classical environment-assisted and private capacity

Appendix A: Justification of (26) and (27).

The arguments follow from [19]. Let us analyze the action of one rocket channel and one erasure. This action can be decomposed into the action of the identity channel with probability \((1 - p)\) and an erasure with probability \(p\) as stated in (25). Let us compute the resulting state in both situations. For \(A^1A^2C_1D_1C_2D_2 = \Phi A^1D_1 \otimes \Phi C_1C_2 \otimes \Phi A^2D_2\) we get:

\[
(R_d \otimes \text{id})(\rho) = R_d^{C_1D_1 \rightarrow B_1} \left( \Phi A^1D_1 \otimes \Phi C_1B_1^2 \right) \otimes \Phi A^2B_2^2
\]

(A1)

The key idea here is that the register \(C_1\) is maximally entangled with the register \(B_2\), which is available to the receiver. Hence, the receiver can undo each unitary applied to \(C_1\) by applying the inverse of the transpose of the corresponding unitary. More precisely, for each choice of \(U\) and \(V\) from the channel:

\[
\left( (V^T)B_1 \circ pB_1B_1^2 \circ ((U^T)^T)^2 \circ R_d^{UV} \right) \left( \Phi A^1D_1 \otimes \Phi C_1B_1^2 \right) = \Phi A^1D_1 \otimes \pi B_1^2
\]

(A2)

In the case of rocket channel and erasure we obtain:

\[
(R_d \otimes E_d \otimes id)(\rho) = R_d^{C_1D_1 \rightarrow B_1} \left( \Phi A^1D_1 \otimes \pi C_1 \right) \otimes \pi A^2 \otimes |e\rangle\langle e|B_1^2B_2^2
\]

(A3)

Let us denote by \(\Phi AB = (\text{id} \otimes U) \Phi AB (\text{id} \otimes U^\dagger)\), then: \(\Phi AB = (U \otimes \text{id}) \Phi AB (U^\dagger \otimes \text{id})\). If we focus our attention in the action of the rocket channel for some concrete \(U\) and \(V\):

\[
R_d^{UV} \left( \Phi A^1D_1 \otimes \pi C_1 \right) = tr_{C_1} \left( \sum_{ijkl} \omega^{ijkl} |ij\rangle\langle ij|D_1C_1 (\Phi V \otimes \pi) |kl\rangle\langle kl|D_1C_1 \right)
\]

(A4)

\[
= \sum_{ijkl} \omega^{ijkl} |ij\rangle\langle ij|D_1C_1 (\Phi V \otimes \pi) |kl\rangle\langle kl|D_1C_1
\]

(A5)

\[
= \sum_{ij} |ij\rangle\langle ij|D_1 (\Phi V \otimes \pi) |ij\rangle\langle ij|D_1
\]

(A6)

\[
= \sum_{j} (V^T \otimes |j\rangle\langle j|D_1) \Phi \left( (V^T)^\dagger \otimes |j\rangle\langle j|D_1 \right)
\]

(A7)

\[
= (V^T \otimes D) (\Phi)
\]

(A8)

where \(\mathcal{D}\) denotes the completely dephasing channel in the computational basis. We can conclude that \(Q^{(1)}(R_d \otimes \text{id}, \rho) = 2 \log d, \ Q^{(1)}(R_d \otimes E_d \otimes \rho) = - \log d\) and \(Q^{(1)}(R_d \otimes E_d \otimes \rho) = (2 - 3p) \log d\).

Now, let us analyze the action of two rocket channels. From the data processing inequality for coherent information we have that \(Q^{(1)}(R_d \otimes R_d, \rho) \geq Q^{(1)}(\mathcal{D} \circ R_d \otimes \mathcal{D} \circ R_d, \rho)\). Now we will show that,

\[
id A^1 \otimes \mathcal{D} \circ R_d^{UV} \otimes \mathcal{D} \circ R_d^{WX} \left( \Phi A^1D_1 \otimes \Phi C_1C_2 \otimes \Phi A^2D_2 \right) = (V^T A^1 \otimes \mathcal{D} \otimes (X^T A^2 \otimes \mathcal{D} \left( \Phi A^1D_1 \otimes \Phi A^2D_2 \right)
\]

(A9)
Then, since coherent information is invariant under the application of local unitaries:

$$Q^{(1)}(R_d^{UV} \otimes R_d^{WX}, \rho) \geq Q^{(1)}(\tilde{D} \otimes \tilde{D}, \Phi^{A^1D_1} \otimes \Phi^{A^2D_2}) = 0 .$$

(A10)

We can write explicitly the form of the ouput after acting on the input state with $R_d^{UV} \otimes R_d^{WX}$:

$$\sigma^{A^1A^2D_1D_2} =$$

(A11)

$$= \sum_{ijkl} \omega^{i-j-k+l+a-b-c-d} \left( |i\rangle \langle j| \Phi_V^{A^1D_1} |id \otimes |l\rangle \langle l| \right) \otimes \left( |i\rangle \langle j| \Phi_X^{A^2D_2} |id \otimes |l\rangle \langle l| \right) \text{tr} \left[ (|i\rangle \langle j| \otimes |a\rangle \langle a| \Phi^{U\tau W} |k\rangle \langle k| \otimes |c\rangle \langle c| \right]$$

(A12)

$$= \sum_{ijkl} \omega^{i-j-k+l+a-b-c-d} \left( |i\rangle \langle j| \Phi_V^{A^1D_1} |id \otimes |l\rangle \langle l| \right) \otimes \left( |i\rangle \langle j| \Phi_X^{A^2D_2} |id \otimes |l\rangle \langle l| \right) \langle ia| \Phi^{U\tau W} |ia\rangle .$$

(A13)

If we apply a dephasing channel at the output we obtain the following:

$$\tilde{D} \otimes \tilde{D} \left( \sigma^{A^1A^2D_1D_2} \right) =$$

(A14)

$$= \sum_{xy} \text{id}^{A^1A^2} \otimes |xy\rangle \langle xy| |D_1D_2 \sigma^{A^1A^2D_1D_2} \text{id}^{A^1A^2} \otimes |x\rangle \langle y|.$$}

We apply the dephasing channel at the output we obtain the following:

$$\tilde{D} \otimes \tilde{D} \left( \sigma^{A^1A^2D_1D_2} \right) =$$

(A15)

$$= \sum_{ijkl} \left( |i\rangle \langle j| \Phi_V^{A^1D_1} |id \otimes |j\rangle \langle j| \right) \otimes \left( |i\rangle \langle j| \Phi_X^{A^2D_2} |id \otimes |j\rangle \langle j| \right) \langle ia| \Phi^{U\tau W} |ia\rangle$$

(A16)

$$= \sum_{j} \left( |i\rangle \langle j| \Phi_V^{A^1D_1} |id \otimes |j\rangle \langle j| \right) \otimes \left( |i\rangle \langle j| \Phi_X^{A^2D_2} |id \otimes |j\rangle \langle j| \right)$$

(A17)

$$= (V^T)^{A_1} \otimes \tilde{D} \otimes (X^T)^{A_2} \otimes \tilde{D} \left( \Phi^{A^1D_1} \otimes \Phi^{A^2D_2} \right).$$

Appendix B: Justification of (38).

The input state is a classical-quantum state of the form: $\sum_{ij} |ij\rangle \langle ij|^X \otimes \rho^{A^1A^2}_{ij}$. We can write explicitly the input states as:

$$\rho^{A^1A^2}_{ij} = |ij\rangle \langle ij|^{A^1} \otimes |ij\rangle \langle ij|^{A^2} .$$

(B1)

If we apply the channel to an input state, we can combine from (35) that the output does only depend on $j$ and it simplifies to:

$$\Phi_{B_1B_2} := \sum_{i} W(i,j) \otimes W(i,j) \Phi$$

(B2)

$$= Z(j) \otimes Z(j) \sum_{i=0}^{d-1} |ii\rangle \langle ii|$$

(B3)

$$= \sum_{i=0}^{d-1} \omega^{2ij} |ii\rangle .$$

(B4)

Let $0 \leq a, b \leq d-1$ and $a \neq b$, we can check the orthogonality between two output states:

$$\langle \Phi_a|\Phi_b\rangle = \frac{1}{d} \sum_{i,j=0}^{d-1} \omega^{-2ai} \omega^{2bj} |ii\rangle \langle jj|$$

(B5)

$$= \frac{1}{d} \sum_{j=0}^{d-1} (\omega^{2(b-a)})^j .$$

(B6)

$$|\Phi_a|\Phi_b\rangle$$

is a geometric series. Then, if $\omega^{2(b-a)} - 1 \neq 0$:

$$\langle \Phi_a|\Phi_b\rangle = \frac{(\omega^{2(b-a)})^d - 1}{(\omega^{2(b-a)}) - 1} = 0 .$$

(B7)

That is, $\Phi_a$ and $\Phi_b$ are orthogonal except if $\omega^{2(b-a)} = 1$ and then $\Phi_a = \Phi_b$. This is the case if $d$ divides $2(b-a)$ which can only occur for $2(b-a) = d$. Hence if $d$ is even there are $d/2$ orthogonal states and if $d$ is odd there are $d$ orthogonal states. We conclude that $I(X; B_1B_2)$ equals $\log d$ if $d$ is odd and $\log d/2$ if $d$ is even as claimed.