

© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# A decoupling approach to classical data transmission over quantum channels

Frédéric Dupuis, Oleg Szehr, and Marco Tomamichel

**Abstract**—Most coding theorems in quantum Shannon theory can be proven using the decoupling technique: to send data through a channel, one guarantees that the environment gets no information about it; Uhlmann’s theorem then ensures that the receiver must be able to decode. While a wide range of problems can be solved this way, one of the most basic coding problems remains impervious to a direct application of this method: sending classical information through a quantum channel. We will show that this problem can, in fact, be solved using decoupling ideas, specifically by proving a “dequantizing” theorem, which ensures that the environment is only classically correlated with the sent data.

Our techniques naturally yield a generalization of the Holevo-Schumacher-Westmoreland Theorem to the one-shot scenario, where a quantum channel can be applied only once.

**Index Terms**—Coding, Decoupling, HSW Theorem, Smooth entropies.

## I. INTRODUCTION

One of the most fruitful ideas that arose in quantum Shannon theory in the past few years is that of *decoupling*: the fact that, in quantum mechanics, the absence of correlations between two systems implies perfect correlations of those two systems with a third one. More precisely, the core idea is as follows: suppose that we have a tripartite pure state  $|\rho\rangle_{ABC}$ , and that we know that the reduced state on  $AB$  is a product state, i.e.  $\text{tr}_C[|\rho\rangle\langle\rho|] = \rho_A \otimes \rho_B$ . Then, we know from the unitary equivalence of purifications that there exists a partial isometry  $V_{C \rightarrow CA} C_B$  with the property that  $V|\rho\rangle = |\psi\rangle_{ACA} \otimes |\varphi\rangle_{BCB}$ . In other words, if  $A$  and  $B$  are completely uncorrelated, then  $C$  contains perfect correlations with both  $A$  and  $B$ . Furthermore, this observation remains true if the state on  $A$  and  $B$  is only close to a product state, as can be shown via Uhlmann’s theorem [23].

This observation can be used to prove coding theorems for quantum Shannon theory problems. To see this, suppose that we have a channel  $\mathcal{T}_{A \rightarrow B}$ , with a Stinespring dilation  $(U_{\mathcal{T}})_{A \rightarrow BE}$ , and that we want to use this channel to send quantum information from Alice (who has access to the input  $A$ ) to Bob (who receives the output system  $B$ ). Let  $\psi_M$  (with purification  $|\psi\rangle_{MR}$ ) be the state of the message Alice wants to send, and let  $W_{M \rightarrow A}$  be the encoding isometry she uses to map her state to the channel input. After encoding the state and sending it through the channel, we have  $\varphi_{RBE} := U_{\mathcal{T}} W \psi W^\dagger U_{\mathcal{T}}^\dagger$ . Now, suppose that the encoding operation is such that  $\varphi_{RE} = \psi_R \otimes \varphi_E$ . Then, the argument in the previous paragraph tells us that there exists an isometry  $V_{B \rightarrow ME'}$  such

that  $V \varphi_{RBE} V^\dagger = \psi_{MR} \otimes \xi_{EE'}$  for some state  $\xi$ . If we then trace out  $EE'$ , we see that  $V$  acted as a decoder to recover the initial state  $\psi_{MR}$ . One can also show that the condition that  $R$  and  $E$  be decoupled is not only sufficient but necessary in order to be able to transmit arbitrary quantum information. This simplifies our task as information theorists: as long as we can design an encoder  $W$  that ensures that this decoupling condition is fulfilled, we know that a decoder must exist, and do not need to explicitly construct it. Furthermore, our aim becomes to *destroy* correlations rather than to ensure their presence, which seems to be a rather less delicate task at first glance.

To enforce the decoupling condition, a number of *decoupling theorems* have arisen [12], [1], [7], [8]. The version from [7], [8], whose approach we will broadly follow here, goes as follows. Let  $\tilde{\mathcal{T}}_{A \rightarrow E}$  be a complementary channel for  $\mathcal{T}_{A \rightarrow B}$  and let  $\rho_{AR}$  be a quantum state. We consider the state  $(\tilde{\mathcal{T}} \otimes \mathcal{I}_R)((U_A \otimes \mathbb{1}_R)\rho_{AR}(U_A^\dagger \otimes \mathbb{1}_R))$  on  $ER$ , where  $U_A$  is chosen randomly according to the Haar measure on  $\mathbb{U}(A)$ . It turns out that this state is decoupled (i.e. that it is close to  $\tilde{\mathcal{T}}(\mathbb{1}/d_A) \otimes \rho_R$  in trace distance) if the state and the channel fulfill a certain entropic criterion, namely that  $H_{\min}^\varepsilon(A|R)_\rho + H_{\min}^\varepsilon(A'|E)_\tau \gtrsim 0$  (these *smooth min-entropies* will be defined in the next section). Roughly speaking, the first term measures how hard the state  $\rho_{AR}$  is to decouple, and the second term measures the “decoupling power” of the channel  $\tilde{\mathcal{T}}$ ; if the decoupling power of the channel exceeds the difficulty of decoupling the state, decoupling does indeed happen.

By appropriately applying the outlined procedure, one can get a variety of coding theorems. This general approach has now become a staple of quantum Shannon theory, and has been used in quantum state merging [12], state transfer (also known as “Fully Quantum Slepian-Wolf”) [1], for sending quantum information through quantum channels [10], for quantum broadcast channels [9], quantum channels with side-information [6], among other examples.

The common point in all of the previous papers is that they use this argument to send *quantum* information. For sending classical information, on the other hand, the argument does not work directly. The reason for this is that if one sends classical information, the channel environment (the system  $E$  above) can also receive a copy of the message without impairing the protocol. However, it turns out that for the protocol to work,  $E$  can *only* share classical correlations with the message; in particular,  $E$  cannot contain any phase information about the message, otherwise Bob cannot decode. Hence, while the vast majority of quantum Shannon theory

F. Dupuis is with ETH Zürich, O. Szehr is with TU München and M. Tomamichel is with CQT, National University of Singapore.

can now be done using decoupling methods, classical coding over quantum channels, the so-called Holevo-Schumacher-Westmoreland theorem (HSW theorem) [11], [18], remains a notable outlier. The purpose of this paper is to close this gap and provide a decoupling proof, based on the above argument, of the HSW theorem.

The results presented here have a somewhat similar flavor to those presented in [16], but a rather different emphasis. In both papers, the idea that the environment cannot have information about the phase of the classical message arises as a central theme. In [16], this occurs in the context of using complementary bases to get coding theorems from privacy amplification and information reconciliation, whereas here it arises as a natural analog of the concept of decoupling.

The paper will be structured as follows. Section II will explain the notation and basic concepts needed for this paper, Section III will give a *dequantizing theorem*, which will be the analog of the decoupling theorem that we will need for the classical case, and Section IV will show how to use it to derive coding theorems for sending classical information over quantum channels. Finally, we discuss the results in Section V.

## II. PRELIMINARIES AND NOTATION

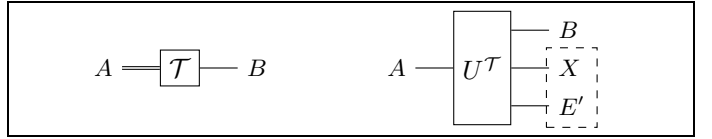
### A. Quantum States and Maps

Let  $\mathcal{H}$  be a finite dimensional, complex Hilbert space. The set of linear operators on  $\mathcal{H}$  will be denoted by  $\mathcal{L}(\mathcal{H})$ , the set of Hermitian operators by  $\mathcal{L}^\dagger(\mathcal{H})$  and the set of positive-semidefinite operators is given by  $\mathcal{P}(\mathcal{H})$ . The set of quantum states is given by  $\mathcal{S}_=(\mathcal{H}) := \{\rho \in \mathcal{P}(\mathcal{H}) \mid \text{tr} \rho = 1\}$  and the set of subnormalized quantum states is  $\mathcal{S}_\leq(\mathcal{H}) := \{\rho \in \mathcal{P}(\mathcal{H}) \mid \text{tr} \rho \leq 1\}$ . A subscript letter following some mathematical object denotes the physical system to which it belongs. However, when it is clear which systems are described we might drop the subscripts to shorten the notation. Given two physical systems  $A$  and  $B$ , the joint bipartite system  $AB$  is represented by a tensor product space  $\mathcal{H}_A \otimes \mathcal{H}_B =: \mathcal{H}_{AB}$ .

We will denote by  $\mathbb{1}_A$  the identity operator on  $\mathcal{H}_A$  and by  $\pi_A := \mathbb{1}_A/d_A$  the completely mixed state on  $A$ , where  $d_A = \dim \mathcal{H}_A$ . For  $d_A \geq d_B$  the states  $T_{AB} := \frac{1}{d_B} \sum_i |i\rangle\langle i|_A \otimes |i\rangle\langle i|_B$  and  $\Phi_{AB} := \frac{1}{d_B} \sum_{i,j} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B$  in  $\mathcal{S}_=(\mathcal{H}_{AB})$  represent maximal classical and, respectively, quantum correlations between the systems  $A$  and  $B$ .

Suppose  $|\psi\rangle_{AB}$  is a pure state of the bipartite system  $AB$  (i.e. the system is in the state  $\psi_{AB} = |\psi\rangle\langle\psi|_{AB}$ ) and  $d_A \geq d_B$ . Then there exist lists of orthonormal vectors  $\{|i\rangle_A\}_{i=1,\dots,d_B} \in \mathcal{H}_A$  and  $\{|i\rangle_B\}_{i=1,\dots,d_B} \in \mathcal{H}_B$  such that  $|\psi\rangle_{AB} = \sum_i \lambda_i |i\rangle_A |i\rangle_B$ , where  $\lambda_i \geq 0$  and  $\sum_i \lambda_i^2 = 1$  [15]. The corresponding basis  $\{|i\rangle\}_{i=1,\dots,d_B}$  is called *Schmidt basis* and the numbers  $\lambda_i$  are *Schmidt coefficients*.

A quantum state  $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_{AB})$  is said to be classical with respect to a fixed basis  $\{|i\rangle\}_{i=1,\dots,d_A}$  of  $\mathcal{H}_A$  if  $\rho_{AB} \in \text{span}_{\mathbb{R}}\{|i\rangle\langle i|_{i=1,\dots,d_A}\} \otimes \mathcal{L}^\dagger(\mathcal{H}_B)$ . If in addition  $\rho_{AB}$  is not classical on  $\mathcal{H}_B$  we call it a hybrid classical-quantum or shortly *CQ-state*. Moreover, we call a state  $\rho_{XX'B} \in \mathcal{S}_\leq(\mathcal{H}_{XX'B})$  *coherent classical* on  $X$  and  $X'$  if it commutes with the projector  $P_{XX'} = \sum_x |x\rangle\langle x|_X \otimes |x\rangle\langle x|_{X'}$ .



**Fig. 1:** Diagram illustrating the purification of a CQ-channel. The environment (depicted with a dashed box) of a CQ-channel can be split into two parts, a register  $X$  that contains a copy of the input and a system  $E'$ .

Linear maps from  $\mathcal{L}(\mathcal{H}_A)$  to  $\mathcal{L}(\mathcal{H}_B)$  will be denoted by calligraphic letters, e.g.  $\mathcal{T}_{A \rightarrow B} \in \text{Hom}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$ . Quantum operations are in one-to-one correspondence with trace preserving completely positive maps (TPCPMs). The TPCPM we will encounter most often is the partial trace (over the system  $B$ ), denoted  $\text{tr}_B(\cdot)$ , which is defined to be the adjoint mapping of  $\mathcal{T}_{A \rightarrow AB}(\xi_A) = \xi_A \otimes \mathbb{1}_B$  for  $\xi_A \in \mathcal{L}^\dagger(\mathcal{H}_A)$  with respect to the Schmidt scalar product  $\langle A, B \rangle := \text{tr}(A^\dagger B)$ . This means  $\text{tr}((\xi_A \otimes \mathbb{1}_B)\zeta_{AB}) = \text{tr}(\xi_A \text{tr}_B(\zeta_{AB}))$  for any  $\zeta_{AB} \in \mathcal{L}^\dagger(\mathcal{H}_{AB})$ . Given a bipartite state  $\xi_{AB}$ , we write  $\xi_A := \text{tr}_B \xi_{AB}$  for the reduced density operator on  $A$  and  $\xi_B := \text{tr}_A \xi_{AB}$ , respectively, on  $B$ . If  $\xi_{AB}$  is pure, we call  $|\xi\rangle_{AB}$  a purification for  $\xi_A$  and  $\xi_B$ .

The map  $\mathcal{C}(\cdot)_A = \sum_i |i\rangle\langle i|_A (\cdot) |i\rangle\langle i|_A$  classicalizes an arbitrary density operator on  $A$  by removing all off-diagonal elements. When  $\mathcal{C}$  is applied to part of a bipartite state  $\rho_{AB}$ , we get the CQ state  $\rho_{AB}^{\text{cl}} := (\mathcal{C}_A \otimes \mathcal{I}_B)(\rho_{AB})$ . Here,  $\mathcal{I}_B$  denotes the operator identity on  $B$ , which we will only write explicitly if it is not clear from the context.

The Choi-Jamiołkowski representation [3], [13] of  $\mathcal{T}_{A \rightarrow B} \in \text{Hom}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$  is given by the operator  $\omega_{A'B} := (\mathcal{T}_{A \rightarrow B} \otimes \mathcal{I}_{A'}) (\Phi_{AA'})$ , where  $\mathcal{H}_{A'}$  is a copy of  $\mathcal{H}_A$ . We say that  $\mathcal{T}_{A \rightarrow B} \in \text{Hom}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$  has classical-quantum (CQ) structure if its Choi-Jamiołkowski representation is a CQ-state. For a map  $\mathcal{T}$  with Choi-Jamiołkowski  $\omega_{A'B}$  we define the map  $\mathcal{T}^{\text{cl}}$  to be the unique map whose Choi-Jamiołkowski representation is  $\omega_{A'B}^{\text{cl}} = \mathcal{C}_{A'}(\omega_{A'B})$ .

In our context it will also be important to purify quantum channels. Given a TPCPM  $\mathcal{T}_{A \rightarrow B} \in \text{Hom}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$ , we define the unitary  $(U_{\mathcal{T}})_{A \rightarrow BE}$  to be any particular Stinespring dilation of  $\mathcal{T}$ . The purifying system  $E$  will be called the *environment* of the channel. For a channel  $\mathcal{T}_{A \rightarrow B}$ , we define the complementary channel  $\tilde{\mathcal{T}}_{A \rightarrow E} : X \mapsto \text{tr}_B[U_{\mathcal{T}} X (U_{\mathcal{T}})^\dagger]$  to be the channel to the environment.

The purification of a CQ-channel  $\mathcal{T}$  with  $\mathcal{T}(\xi_A) = \sum_i \text{tr}(|i\rangle\langle i|_A \xi_A) \rho_B^{[i]}$  is given by  $U_{\mathcal{T}} |i\rangle_A = |i\rangle_X \otimes |\rho^{[i]}\rangle_{BE}$ . Thus, the environment of such a channel can conceptually be split into two parts: a register  $X$ , which contains a copy of the input to the channel and a system  $E'$  which stems from the purification of the operators  $\rho_B^{[i]}$ . See Figure 1 for an illustration of this. The Choi-Jamiołkowski representation of a complementary channel of a CQ-channel can be written as  $\omega_{A'E'X}$ , where the systems  $X$  and  $A'$  are classically coherent. Furthermore, we will frequently be considering channels that are complementary to CQ channels; we will call such channels “complementary CQ channels”.

The swap operator  $F_{AA'}$  acting on the bipartite space  $\mathcal{H}_{AA'}$

is given by  $F_{AA'} := \sum_{i,j} |i\rangle\langle j|_A \otimes |j\rangle\langle i|_{A'}$ . It is easy to verify that for any  $M_A, N_{A'} \in \mathcal{L}(\mathcal{H}_A)$  the swap operator satisfies  $\text{tr}(M_A N_{A'}) = \text{tr}((M_A \otimes N_{A'}) F_{AA'})$ .

For any operator in  $\xi_A \in \mathcal{L}(\mathcal{H}_A)$  we denote by  $\|\xi_A\|_1$  and  $\|\xi_A\|_2$  the Schatten 1 and 2-norms of  $\xi_A$ , respectively. These norms are unitarily invariant and satisfy  $\|\xi_A\|_2 \leq \|\xi_A\|_1 \leq \sqrt{d_A} \|\xi_A\|_2$ . The metric induced on  $\mathcal{L}(\mathcal{H})$  via the Schatten 1-norm is  $D(\rho, \sigma) := \|\rho - \sigma\|_1$ . Another measure of closeness between states on  $\mathcal{P}(\mathcal{H})$  is the fidelity,  $F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1$ .

### B. Permutation operators

The symmetric group  $S_d$  is the set of all bijective maps of  $\{1, \dots, d\}$  to itself together with the concatenation of maps as the group multiplication. Elements  $\pi$  of  $S_d$  are called permutations. Let  $\mathcal{H}$  be a Hilbert space together with a fixed basis  $\{|i\rangle\}_{i=1, \dots, d}$ . For  $\pi \in S_d$ , we define the permutation operator  $P(\pi)$  on  $\mathcal{H}$  such that  $P(\pi)|i\rangle = |\pi(i)\rangle$ . The group of all such matrices will be denoted by  $\mathbb{P}$ . Typically in this paper  $\{|i\rangle\}_{i=1, \dots, d}$  will be the Schmidt basis of a given density matrix. The above permutation matrices then act by reordering the elements of this basis.

Given a random variable  $X : \mathbb{P} \rightarrow \Omega$  ( $\Omega$  some measurable space), we shall write  $\mathbb{E}_{\mathbb{P}}[X] := \frac{1}{d!} \sum_{P \in \mathbb{P}} X(P)$  for the expectation value of  $X$  with respect to the uniform probability distribution on  $\mathbb{P}$ .

### C. Smooth entropies

Entropies are used to quantify the uncertainty an observer has about a quantum state. Moreover, conditional entropies quantify the uncertainty of an observer about one subsystem of a bipartite state when he has access to another subsystem. The most commonly used quantity is the von Neumann entropy. Given a state  $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ , we denote by  $H(A|B)_\rho := H(\rho_{AB}) - H(\rho_B)$  the von Neumann entropy of  $A$  conditioned on  $B$ , where  $H(\rho) := -\text{tr}(\rho \log \rho)$ .

While the von Neumann entropy is appropriate for analyzing processes involving a large number of copies of an identical system, the min-entropy is relevant when a single system is considered [17].

**Min-Entropy [17]** Let  $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_{AB})$ , then the min-entropy of  $A$  conditioned on  $B$  of  $\rho_{AB}$  is defined as

$$H_{\min}(A|B)_\rho := \max_{\sigma_B \in \mathcal{S}_=(\mathcal{H}_B)} \sup\{\lambda \in \mathbb{R} \mid \rho_{AB} \leq 2^{-\lambda} \mathbb{1}_A \otimes \sigma_B\}.$$

More generally, the smooth conditional min-entropy is defined as the largest conditional min-entropy one can get within a distance of at most  $\varepsilon$  from  $\rho$ . Here closeness is measured with respect to the *purified distance*,  $P(\rho, \sigma)$ , which is defined as [22]

$$P(\rho, \sigma) := \sqrt{1 - \bar{F}(\rho, \sigma)^2},$$

where  $\bar{F}(\rho, \sigma)$  is the *generalized fidelity*;  $\bar{F}(\rho, \sigma) := F(\rho, \sigma) + \sqrt{(1 - \text{tr} \rho)(1 - \text{tr} \sigma)}$  for  $\rho, \sigma \in \mathcal{S}_\leq(\mathcal{H})$ . The purified distance constitutes a metric [22] on  $\mathcal{S}_\leq(\mathcal{H})$  and satisfies the Fuchs-van

de Graaf inequalities

$$\begin{aligned} \frac{1}{2} \|\rho - \sigma\|_1 + \frac{1}{2} |\text{tr} \rho - \text{tr} \sigma| &\leq P(\rho, \sigma) \\ &\leq \sqrt{\|\rho - \sigma\|_1 + |\text{tr} \rho - \text{tr} \sigma|}. \end{aligned} \quad (1)$$

We say that  $\rho$  is  $\varepsilon$ -close to  $\tilde{\rho}$ , denoted  $\tilde{\rho} \approx_\varepsilon \rho$ , if  $P(\rho, \tilde{\rho}) \leq \varepsilon$ .

**Smooth Min-Entropy [17], [22]** Let  $\varepsilon \geq 0$  and let  $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_{AB})$  with  $\sqrt{\text{tr} \rho} > \varepsilon$ , then the  $\varepsilon$ -smooth min-entropy of  $A$  conditioned on  $B$  of  $\rho_{AB}$  is defined as

$$H_{\min}^\varepsilon(A|B)_\rho = \max_{\tilde{\rho}} H_{\min}(A|B)_{\tilde{\rho}},$$

where we maximize over all  $\tilde{\rho} \approx_\varepsilon \rho$ .

Next, we define the smooth max-entropy.

**Smooth Max-Entropy [17], [22]** Let  $\varepsilon \geq 0$ , let  $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_{AB})$  and let  $\rho_{ABC} \in \mathcal{S}_\leq(\mathcal{H}_{ABC})$  be an arbitrary purification of  $\rho_{AB}$ . The  $\varepsilon$ -smooth max-entropy of  $A$  conditioned on  $B$  of  $\rho_{AB}$  is defined as

$$H_{\max}^\varepsilon(A|B)_\rho = -H_{\min}^\varepsilon(A|C)_{\rho_{ABC}}.$$

The fully quantum asymptotic equipartition property (QAEF) states that in the limit of an infinite number of identical states the smooth min- and max-entropies converge to the von Neumann entropy [21], [20]: Let  $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ , then

$$\begin{aligned} H(A|B)_\rho &= \lim_{n \rightarrow \infty} \frac{1}{n} H_{\min}^\varepsilon(A^n|B^n)_{\rho^{\otimes n}} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} H_{\max}^\varepsilon(A^n|B^n)_{\rho^{\otimes n}}. \end{aligned} \quad (2)$$

In that sense, the smooth conditional min- and max-entropies can be seen as one-shot generalizations of the von Neumann entropy.

### D. Uhlmann's theorem and existence of a decoding operation

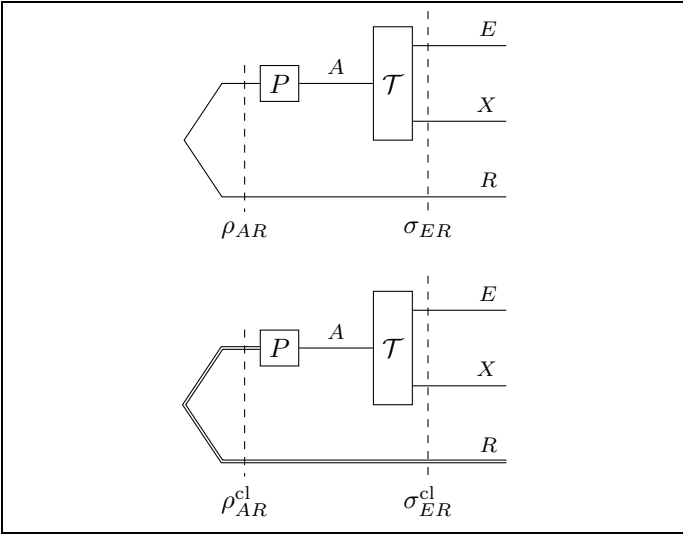
To prove a coding theorem it is necessary to establish the existence of a decoding operation. That is, given a quantum state that results from the execution of some quantum channel, we would like to recover the message originally encoded into the input of the channel. It turns out that this can be achieved if and only if the environment of the channel and some reference system purifying the original message are left uncorrelated after the execution of the channel. In this situation the existence of a decoding operation follows from Uhlmann's Theorem [23], which we shall state here for completeness.

**Theorem 2.1 (Uhlmann's Theorem):** Let  $\rho_A, \sigma_A \in \mathcal{S}(\mathcal{H}_A)$  be two quantum states with respective purifications  $|\phi\rangle_{AB}$  and  $|\psi\rangle_{AC}$ . Then,

$$F(\rho_A, \sigma_A) = \max_{V_{B \rightarrow C}} |\langle \psi | V | \phi \rangle|,$$

where the maximization goes over all partial isometries from  $B$  to  $C$ .

Since our decoupling results involve the Schatten 1-norm rather than the Fidelity it will be useful to transform the above theorem into a statement formulated in terms of Schatten 1-norms. The following Corollary [5] follows from Uhlmann's



**Fig. 2:** Illustration of the dequantizing theorem. The top diagram illustrates the situation in which we apply the dequantizing theorem: we apply a random permutation  $P$  to  $\rho_{AR}$  following by the channel. The bottom diagram illustrates the “ideal” state we would like to get at the end: a state containing only classical correlations between  $R$  and  $E$ .

Theorem with an application of the Fuchs van de Graaf Inequalities (cf. Equation (1)).

*Corollary 2.2:* Let  $\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$  be pure quantum states and assume that  $\|\rho_A - \sigma_A\|_1 \leq \varepsilon$ . Then there exists some isometry  $U_{B \rightarrow C}$  such that  $\|U\rho_{AB}U^\dagger - \sigma_{AB}\|_1 \leq 2\sqrt{\varepsilon}$ .

### III. DEQUANTIZING THEOREM

In this section, we will derive the dequantizing theorem which will be the core technical ingredient for our coding theorems. Our aim will be to derive conditions under which the output of a channel contains only classical correlations with a reference system. More precisely, we will prove the following:

*Theorem 3.1 (Dequantizing Theorem):* Let  $\mathcal{T}_{A \rightarrow EX}$  be a complementary CQ channel, and let  $\omega_{A'EX} \in \mathcal{S}_{\leq}(\mathcal{H}_{A'EX})$  be its Choi-Jamiołkowski representation. Let  $\rho_{AR}$  be a pure state on  $\mathcal{H}_{AR}$ , and let  $\rho_{AR}^{cl} := \mathcal{C}_A(\rho_{AR})$ . Then

$$\mathbb{E}_{P_A \in \mathbb{P}} \left\| \mathcal{T} \left( P_A (\rho_{AR} - \rho_{AR}^{cl}) P_A^\dagger \right) \right\|_1 \leq \sqrt{\frac{1}{d_A - 1} 2^{-H_{\min}(A|EX)_\omega - H_{\min}(A|R)_\rho}}, \quad (3)$$

where the permutation operators act by permuting the Schmidt-basis vectors of  $\rho_{AR}$ .

In other words, Theorem 3.1 gives a bound on how close the state  $\mathcal{T}(P_A \rho_{AR} P_A^\dagger)$  is from a state containing only classical correlations between  $R$  and  $E$  (namely,  $\mathcal{T}(P_A \rho_{AR}^{cl} P_A^\dagger)$ ). See Figure 2 for an illustration.

The rest of this section is devoted to the proof of Theorem 3.1 and is organized in three subsections. In the first one we calculate the above expectation value with the Schatten 1-norm replaced by the Schatten 2-norm. We conclude the proof of the above theorem in the second subsection showing

how the statement found about the Schatten 2-norm can be transformed into the one above. In the third subsection we reformulate the upper bound of Theorem 3.1 using the smooth conditional min-entropy. This enables us to make statements about independent, identically distributed channels via the QAEF, Equation (2).

#### A. Dequantizing with Schatten 2-Norms

We first prove a statement that holds for general hermiticity preserving, linear maps  $\mathcal{N}_{A \rightarrow B} \in \text{Hom}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$  with Choi-Jamiołkowski representation  $\omega_{A'B} \in \mathcal{L}^\dagger(\mathcal{H}_{A'B})$ . In Proposition 3.3 below, we will compute the expectation value

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left\| \mathcal{N} \left( P_A (\rho_{AR} - \rho_{AR}^{cl}) P_A^\dagger \right) \right\|_2^2 \\ = \frac{d_A}{d_A - 1} \left\| \rho_{AR} - \rho_{AR}^{cl} \right\|_2^2 \left\| \omega_{A'B} - \omega_{A'B}^{cl} \right\|_2^2, \end{aligned}$$

where the permutation operators act by permuting the basis vectors of the Schmidt-basis<sup>1</sup>. To prove this, we first need the following lemma:

*Lemma 3.2:* Let  $\mathcal{H}_A$  be a Hilbert space with orthonormal basis  $\{|i\rangle\}_{i=1, \dots, d_A}$  and let  $\mathbb{P}$  be the corresponding set of permutation operators. Then for any  $i \neq j$  one has that

$$\begin{aligned} \mathbb{E}_{P \in \mathbb{P}} \left( P^{\otimes 2} (|i\rangle\langle j|_A \otimes |j\rangle\langle i|_{A'}) (P^\dagger)^{\otimes 2} \right) \\ = \frac{1}{d_A(d_A - 1)} (F_{AA'} - d_A T_{AA'}). \end{aligned}$$

*Proof:* There are  $d_A!$  permutation operators in  $\mathbb{P}$ . For arbitrary but fixed  $i \neq j$  and  $k \neq l$  there are  $(d_A - 2)!$  permutation operators with  $(P_A)^{\otimes 2} |i\rangle_A \otimes |j\rangle_{A'} = |k\rangle_A \otimes |l\rangle_{A'}$ . On the other hand there is no permutation such that for  $i \neq j$  the operator  $(P_A)^{\otimes 2}$  maps  $|i\rangle_A \otimes |j\rangle_{A'}$  to  $|k\rangle_A \otimes |l\rangle_{A'}$  with  $k = l$ . We conclude that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left( (P_A)^{\otimes 2} (|i\rangle\langle j|_A \otimes |j\rangle\langle i|_{A'}) (P_A^\dagger)^{\otimes 2} \right) \\ = \frac{(d_A - 2)!}{d_A!} \sum_{k \neq l} |k\rangle\langle l|_A \otimes |l\rangle\langle k|_{A'}. \end{aligned}$$

*Proposition 3.3 (Distance from classicality):* Let  $\mathcal{N}_{A \rightarrow B} \in \text{Hom}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$  be a linear map with Choi-Jamiołkowski representation  $\omega_{A'B} \in \mathcal{L}^\dagger(\mathcal{H}_{B A'})$ , and let  $|\rho\rangle_{AR} = \sum_i \sqrt{\lambda_i} |ii\rangle_{AR}$ . Then

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left\| \mathcal{N} \left( P_A (\rho_{AR} - \rho_{AR}^{cl}) P_A^\dagger \right) \right\|_2^2 \\ = \frac{d_A}{d_A - 1} \left\| \rho_{AR} - \rho_{AR}^{cl} \right\|_2^2 \left\| \omega_{A'B} - \omega_{A'B}^{cl} \right\|_2^2, \end{aligned}$$

where the operators  $P$  permute the Schmidt-basis vectors of  $\rho_A$ .

*Proof:* Rewriting the Schatten 2-norm in terms of the trace, we get

<sup>1</sup>An extension of this result to an arbitrary basis is known [19].

$$\begin{aligned}
& \mathbb{E}_{\mathbb{P}} \left\| \mathcal{N}(P_A \otimes \mathbb{1}_R (\rho_{AR} - \rho_{AR}^{\text{cl}}) P_A^\dagger \otimes \mathbb{1}_R) \right\|_2^2 \\
&= \mathbb{E}_{\mathbb{P}} \text{tr} \left( \mathcal{N}(P_A \otimes \mathbb{1}_R (\rho_{AR} - \rho_{AR}^{\text{cl}}) P_A^\dagger \otimes \mathbb{1}_R)^2 \right) \\
&= \mathbb{E}_{\mathbb{P}} \text{tr} \left( \mathcal{N} \left( \sum_{i \neq j}^{d_R} \sqrt{\lambda_i \lambda_j} P_A |i\rangle\langle j|_A P_A^\dagger \otimes |i\rangle\langle j|_R \right)^2 \right) \\
&= \sum_{i \neq j}^{d_R} \lambda_i \lambda_j \left[ \mathbb{E}_{\mathbb{P}} \text{tr} \left( \mathcal{N} \left( P_A |i\rangle\langle j|_A P_A^\dagger \right) \mathcal{N} \left( P_A |j\rangle\langle i|_A P_A^\dagger \right) \right) \right] \\
&= \sum_{i \neq j}^{d_R} \left\{ \lambda_i \lambda_j \cdot \right. \\
&\quad \left. \text{tr} \left( \mathcal{N}^{\otimes 2} \left( \mathbb{E}_{\mathbb{P}} \left( P_A^{\otimes 2} (|i\rangle\langle j|_A \otimes |j\rangle\langle i|_{A'}) (P_A^\dagger)^{\otimes 2} \right) F_{BB'} \right) \right) \right\} \\
&\quad (4) \\
&= \sum_{i \neq j}^{d_R} \frac{\lambda_i \lambda_j}{d_A (d_A - 1)} \text{tr} \left( \mathcal{N}^{\otimes 2} \left( (F_{AA'} - d_A T_{AA'}) F_{BB'} \right) \right) \quad (5) \\
&= \frac{\|\rho_{AR} - \rho_{AR}^{\text{cl}}\|_2^2}{d_A (d_A - 1)} \\
&\quad \left( \text{tr} \left( \mathcal{N}^{\otimes 2} (F_{AA'}) F_{BB'} \right) - \text{tr} \left( (\mathcal{N}^{\text{cl}})^{\otimes 2} (F_{AA'}) F_{BB'} \right) \right). \quad (6)
\end{aligned}$$

Equation (4) is by an application of the swap trick and in equation (5) we applied Lemma 3.2. To simplify (6) we evaluate the term  $\text{tr} \left( \mathcal{N}^{\otimes 2} (F_{AA'}) F_{BB'} \right)$  using the inverse Choi-Jamiołkowski isomorphism:

$$\begin{aligned}
& \text{tr} \left( \mathcal{N}^{\otimes 2} (F_{AA'}) F_{BB'} \right) \\
&= d_A^2 \text{tr} \left( \text{tr}_{AA'} \left( \omega_{AB}^{\otimes 2} (F_{AA'} \otimes \mathbb{1}_{BB'}) \right) F_{BB'} \right) \\
&= d_A^2 \text{tr} \left( \omega_{AB}^{\otimes 2} (F_{AA'} \otimes \mathbb{1}_{BB'}) (\mathbb{1}_{AA'} \otimes F_{BB'}) \right) \quad (7) \\
&= d_A^2 \text{tr} \left( \omega_{AB}^2 \right).
\end{aligned}$$

In Equation (7) we used the fact that the adjoint mapping of the partial trace is tensoring with the identity. Analogously, one has

$$\text{tr} \left( (\mathcal{N}^{\text{cl}})^{\otimes 2} (F_{AA'}) F_{BB'} \right) = d_A^2 \text{tr} \left( (\omega_{AB}^{\text{cl}})^2 \right)$$

and the second factor of (6) becomes

$$\begin{aligned}
& \text{tr} \left( \mathcal{N}^{\otimes 2} (F_{AA'}) F_{BB'} \right) - \text{tr} \left( (\mathcal{N}^{\text{cl}})^{\otimes 2} (F_{AA'}) F_{BB'} \right) \\
&= d_A^2 \left( \text{tr} \left( \omega_{AB}^2 \right) - \text{tr} \left( (\omega_{AB}^{\text{cl}})^2 \right) \right) \\
&= d_A^2 \|\omega_{AB} - \omega_{AB}^{\text{cl}}\|_2^2.
\end{aligned}$$

### B. Dequantizing with the Schatten 1-norm

In this subsection we derive Theorem 3.1 with an application of Proposition 3.3. We use the following lemma.

*Lemma 3.4:* For any  $\xi_{AR} \in \mathcal{S}_{\leq}(\mathcal{H}_{AR})$ , there exists an operator  $\zeta_R \in \mathcal{S}_{=}(\mathcal{H}_R)$  with

$$\frac{1}{\text{tr}[\xi_{AR}]} \text{tr} \left( \left( (\mathbb{1}_A \otimes \zeta_R^{-1/2}) \xi_{AR} \right)^2 \right) \leq 2^{-H_{\min}(A|R)\xi}.$$

*Proof:* Choose  $\zeta_R$  such that it maximizes the min-entropy, i.e. it satisfies  $\xi_{AR} \leq 2^{-H_{\min}(A|R)\xi} \mathbb{1}_A \otimes \zeta_R$ . Hence,

$$\sqrt{\xi_{AR}} (\mathbb{1}_A \otimes \zeta_R^{-\frac{1}{2}}) \xi_{AR} (\mathbb{1}_A \otimes \zeta_R^{-\frac{1}{2}}) \sqrt{\xi_{AR}} \leq 2^{-H_{\min}(A|R)\xi} \xi_{AR}$$

Taking the trace on both sides concludes the proof.  $\blacksquare$

*Proof of Theorem 3.1:* We first introduce some notation. We abbreviate the difference between  $\rho_{AR}$  and its classicalized version by writing  $\bar{\rho}_{AR} := \rho_{AR} - \rho_{AR}^{\text{cl}}$ . By Lemma 3.4 there are operators  $\sigma_{EX}$  and  $\tau_R$  such that

$$\frac{1}{\text{tr}[\omega_{A'EX}]} \text{tr} \left( \left( (\mathbb{1}_{A'} \otimes \sigma_{EX}^{-1/2}) \omega_{A'EX} \right)^2 \right) \leq 2^{-H_{\min}(A'|EX)\omega}$$

and

$$\frac{1}{\text{tr}[\rho_{AR}]} \text{tr} \left( \left( (\mathbb{1}_A \otimes \tau_R^{-1/2}) \rho_{AR} \right)^2 \right) \leq 2^{-H_{\min}(A|R)\rho}.$$

Since  $\bar{\mathcal{T}}_{A' \rightarrow EX}$  is a complementary channel of a CQ-channel we can assume that  $\sigma_{EX}$  has CQ-structure, i.e.  $\sigma_{EX} = \sum_x \sigma_E^x \otimes |x\rangle\langle x|_X$ . Furthermore the operator  $\rho_{AR}$  is given in its Schmidt-basis, such that  $\tau_R$  can be written as  $\tau_R = \sum_x r_x |x\rangle\langle x|_R$ . (Both facts follow from Lemma A.1 in Appendix A with  $\varepsilon = 0$ .)

We introduce a system  $P$  with  $(d_A!)$ -dimensional Hilbert space,  $\mathcal{H}_P$ , and canonical basis vectors  $|P\rangle_P$  that correspond to the permutation operators of  $P \in \mathbb{P}$ . We define the operator

$$\begin{aligned}
& \zeta_{PREX} := \\
& \mathbb{E}_{\mathbb{P}} \left( \sum_{x=1}^{d_R} r_x |P\rangle\langle P|_P \otimes |x\rangle\langle x|_R \otimes \sigma_E^{P(x)} \otimes |P(x)\rangle\langle P(x)|_X \right),
\end{aligned}$$

which can be inverted on its support to yield

$$\begin{aligned}
& \zeta_{PREX}^{-1} := d_A! \sum_{P \in \mathbb{P}} \left( \sum_{x=1}^{d_R} r_x^{-1} |P\rangle\langle P|_P \otimes |x\rangle\langle x|_R \otimes \right. \\
& \quad \left. \left( \sigma_E^{P(x)} \right)^{-1} \otimes |P(x)\rangle\langle P(x)|_X \right).
\end{aligned}$$

We note that the operator  $\zeta^{\frac{1}{4}} \zeta^{-\frac{1}{4}}$  is a projector and

$$\begin{aligned}
& \mathbb{E}_{\mathbb{P}} |P\rangle\langle P|_P \otimes \bar{\mathcal{T}}(P_A \otimes \mathbb{1}_R \bar{\rho}_{AR} P_A^\dagger \otimes \mathbb{1}_R) \\
&= (\zeta^{\frac{1}{4}} \zeta^{-\frac{1}{4}}) \left[ \mathbb{E}_{\mathbb{P}} |P\rangle\langle P|_P \otimes \bar{\mathcal{T}}(P_A \otimes \mathbb{1}_R \bar{\rho}_{AR} P_A^\dagger \otimes \mathbb{1}_R) \right] (\zeta^{-\frac{1}{4}} \zeta^{\frac{1}{4}}) \quad (8)
\end{aligned}$$

Using these operators, we write

$$\begin{aligned}
& \mathbb{E}_{\mathbb{P}} \left\| \bar{\mathcal{T}}(P_A \bar{\rho}_{AR} P_A^\dagger) \right\|_1 \\
&= \mathbb{E}_{\mathbb{P}} \left\| |P\rangle\langle P|_P \otimes \left( \bar{\mathcal{T}}(P_A \bar{\rho}_{AR} P_A^\dagger) \right) \right\|_1 \\
&= \left\| \mathbb{E}_{\mathbb{P}} |P\rangle\langle P|_P \otimes \left( \bar{\mathcal{T}}(P_A \bar{\rho}_{AR} P_A^\dagger) \right) \right\|_1 \\
&\leq \sqrt{\text{tr}(\zeta_{PREX})} \left\| \zeta^{-\frac{1}{4}} \left( \mathbb{E}_{\mathbb{P}} |P\rangle\langle P|_P \otimes \bar{\mathcal{T}}(P_A \bar{\rho}_{AR} P_A^\dagger) \right) \zeta^{-\frac{1}{4}} \right\|_2. \quad (9)
\end{aligned}$$

Inequality (9) follows from Equation (8) together with an application of the Hölder-type inequality  $\|ABC\|_1 \leq$

$\| |A|^4 \|_1^{\frac{1}{4}} \| |B|^2 \|_1^{\frac{1}{2}} \| |C|^4 \|_1^{\frac{1}{4}}$ , [2]. The trace term on the right hand side of Inequality (9) can be evaluated directly to be

$$\text{tr}(\zeta_{PREX}) = \sum_{x=1}^{d_R} r_x \mathbb{E}_{\mathbb{P}} \text{tr} \left( \sigma_E^{P(x)} \right) = \frac{1}{d_A}.$$

Thus, it is sufficient to evaluate the term with the Schatten 2-norm. For notational convenience we introduce the map  $\tilde{\mathcal{T}}(\cdot) := (\sigma_{EX})^{-\frac{1}{4}} \tilde{\mathcal{T}}(\cdot) (\sigma_{EX})^{-\frac{1}{4}}$  with Choi-Jamiołkowski representation  $\tilde{\omega}_{A'EX}$  and the operator  $\tilde{\rho}_{AR} := (\mathbb{1}_A \otimes \tau_R)^{-\frac{1}{4}} \tilde{\rho}_{AR} (\mathbb{1}_A \otimes \tau_R)^{-\frac{1}{4}}$ . Using the fact that  $\tilde{\mathcal{T}}$  is the complementary channel of a CQ-channel one can verify that

$$\begin{aligned} & \zeta^{-\frac{1}{4}} \left( \mathbb{E}_{\mathbb{P}} |P\rangle\langle P|_{\mathbb{P}} \otimes \tilde{\mathcal{T}}(P_A \tilde{\rho}_{AR} P_A^\dagger) \right) \zeta^{-\frac{1}{4}} \\ &= \frac{1}{\sqrt{d_A!}} \sum_{P \in \mathbb{P}} |P\rangle\langle P|_{\mathbb{P}} \otimes \\ & \quad \left( (\sigma_{EX} \otimes \tau_R)^{-\frac{1}{4}} \tilde{\mathcal{T}}(P_A \tilde{\rho}_{AR} P_A^\dagger) (\sigma_{EX} \otimes \tau_R)^{-\frac{1}{4}} \right) \\ &= \frac{1}{\sqrt{d_A!}} \sum_{P \in \mathbb{P}} |P\rangle\langle P|_{\mathbb{P}} \otimes \tilde{\mathcal{T}}(P_A \tilde{\rho}_{AR} P_A^\dagger). \end{aligned}$$

Using this, we find

$$\begin{aligned} & \left\| \zeta^{-\frac{1}{4}} \left( \mathbb{E}_{\mathbb{P}} |P\rangle\langle P|_{\mathbb{P}} \otimes \tilde{\mathcal{T}}(P_A \tilde{\rho}_{AR} P_A^\dagger) \right) \zeta^{-\frac{1}{4}} \right\|_2^2 \\ &= \mathbb{E}_{\mathbb{P}} \left\| \tilde{\mathcal{T}}(P_A \tilde{\rho}_{AR} P_A^\dagger) \right\|_2^2 \\ &= \frac{d_A}{d_A - 1} \left\| \tilde{\omega}_{A'EX} - \tilde{\omega}_{A'EX}^{\text{cl}} \right\|_2^2 \left\| \tilde{\rho}_{AR} - \tilde{\rho}_{AR}^{\text{cl}} \right\|_2^2 \\ &\leq \frac{d_A}{d_A - 1} \text{tr}(\tilde{\omega}_{A'EX}^2) \text{tr}(\tilde{\rho}_{AR}^2) \\ &\leq \frac{d_A}{d_A - 1} 2^{-H_{\min}(A'|EX)_\omega - H_{\min}(A|R)_\rho}, \end{aligned}$$

where we apply Proposition 3.3 to obtain the second equality and use the special choice of  $\sigma_{EX}$  and  $\tau_R$  (cf. Lemma 3.4) for the last inequality. Plugging this into Equation (9) concludes the proof.  $\blacksquare$

### C. A smoothed version of the dequantizing theorem

The reason for introducing smooth versions of the min- and max-entropy is that these quantities can vary a lot with small variations in the underlying states while the quantities that we are bounding with them do not. This is such a case; it is therefore desirable to have a version of the dequantizing theorem which involves the smooth entropies. The smooth quantities have the additional advantage that they converge to the corresponding von Neumann quantities in the i.i.d. case (cf. Equation 2). We therefore prove the following:

*Theorem 3.5:* Let  $\tilde{\mathcal{T}}_{A \rightarrow EX}$  be a complementary channel of a CQ-channel, let  $\omega_{A'EX} \in \mathcal{S}_{\leq}(\mathcal{H}_{EXA'})$  be the Choi-Jamiołkowski representation of  $\tilde{\mathcal{T}}$ , and let  $|\rho\rangle_{AR} = \sum_i \sqrt{\lambda_i} |ii\rangle_{AR}$ . Let  $\varepsilon, \varepsilon'$  be such that  $\sqrt{\text{tr}(\rho)} > \varepsilon \geq 0$  and  $\sqrt{\text{tr}(\omega)} > \varepsilon' \geq 0$ . Then,

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}} \left\| \tilde{\mathcal{T}} \left( P_A (\rho_{AR} - \rho_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1 \\ & \leq \sqrt{\frac{1}{d_A - 1} 2^{-H_{\min}^{\varepsilon'}(A'|EX)_\omega - H_{\min}^\varepsilon(A|R)_\rho}} + 8\varepsilon + 8\varepsilon', \end{aligned}$$

where the permutation operators act by permuting the Schmidt-basis vectors of  $\rho_A$ .

*Proof:* Let  $\hat{\omega}_{A'EX} \in \mathcal{S}_{\leq}(\mathcal{H}_{A'EX})$  be a state that saturates the bound in the definition of the smooth min-entropy, i.e.  $P(\omega_{A'EX}, \hat{\omega}_{A'EX}) \leq \varepsilon'$  and  $H_{\min}^{\varepsilon'}(A'|EX)_\omega = H_{\min}^{\varepsilon'}(A'|EX)_{\hat{\omega}}$ . Analogously,  $\hat{\rho}_{AR}$  satisfies  $P(\rho_{AR}, \hat{\rho}_{AR}) \leq \varepsilon$  and  $H_{\min}^\varepsilon(A|R)_\rho = H_{\min}^\varepsilon(A|R)_{\hat{\rho}}$ .

Using inequality (1), we find that

$$\|\omega_{A'EX} - \hat{\omega}_{A'EX}\|_1 \leq 2\varepsilon' \quad \|\rho_{AR} - \hat{\rho}_{AR}\|_1 \leq 2\varepsilon. \quad (10)$$

We decompose  $\hat{\omega} - \omega$  into positive operators with orthogonal support writing  $\hat{\omega} - \omega = \Delta_+ - \Delta_-$  and conclude from (10) that  $\|\Delta_+\|_1 \leq 2\varepsilon'$  and  $\|\Delta_-\|_1 \leq 2\varepsilon'$ .

Similarly, we decompose  $(\rho_{AR} - \rho_{AR}^{\text{cl}}) - (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}}) = \Gamma_+ - \Gamma_-$  with the operators  $\Gamma_+$  and  $\Gamma_-$  again chosen to be positive and with orthogonal support. From the second inequality in (10), we conclude that  $\|\Gamma_+\|_1 \leq 4\varepsilon$  and  $\|\Gamma_-\|_1 \leq 4\varepsilon$ .

Let  $\hat{\mathcal{T}}$ ,  $\mathcal{D}_+$  and  $\mathcal{D}_-$  be the unique Choi-Jamiołkowski preimages of  $\hat{\omega}_{A'EX}$ ,  $\Delta_+$  and  $\Delta_-$  respectively. We note that from the fact that  $\omega_{A'EX}$  is classically coherent between  $A'$  and  $X$  it follows that  $\hat{\omega}_{A'EX}$  also shares this property (See Appendix A, Lemma A.1). Furthermore the state  $\rho_{AR}$  is classically coherent between  $A$  and  $R$ , such that the state  $\hat{\rho}_{AR}$  has the same Schmidt-basis as  $\rho_{AR}$  (Lemma A.1). We therefore can apply Theorem 3.1 on the states  $\hat{\omega}$  and  $\hat{\rho}$  to find

$$\begin{aligned} & \sqrt{\frac{1}{d_A - 1} 2^{-H_{\min}^{\varepsilon'}(A'|EX)_\omega - H_{\min}^\varepsilon(A|R)_\rho}} \\ &= \sqrt{\frac{1}{d_A - 1} 2^{-H_{\min}(A'|EX)_{\hat{\omega}} - H_{\min}(A|R)_{\hat{\rho}}}} \\ &\geq \mathbb{E}_{\mathbb{P}} \left\| \hat{\mathcal{T}} \left( P_A (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1. \end{aligned}$$

Applying the triangle inequality twice shows that for any permutation operator we have

$$\begin{aligned} & \left\| \hat{\mathcal{T}} \left( P_A (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1 \\ & \geq \left\| \tilde{\mathcal{T}} \left( P_A (\rho_{AR} - \rho_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1 \\ & \quad - \left\| (\tilde{\mathcal{T}} - \hat{\mathcal{T}}) \left( P_A (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1 \\ & \quad - \left\| \tilde{\mathcal{T}} \left( P_A ((\rho_{AR} - \rho_{AR}^{\text{cl}}) - (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}})) P_A^\dagger \right) \right\|_1 \quad (11) \end{aligned}$$

The first term on the right hand-side of Inequality (11) corresponds to the unsmoothed dequantizing theorem. For the remaining two terms we find upper bounds:

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}} \left\| (\tilde{\mathcal{T}} - \hat{\mathcal{T}}) \left( P_A (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1 \\ & \leq \sum_{a \in \{+, -\}} \mathbb{E}_{\mathbb{P}} \text{tr} \left( D_a \left( P_A \hat{\rho}_{AR} P_A^\dagger \right) \right) \\ & \quad + \sum_{a \in \{+, -\}} \mathbb{E}_{\mathbb{P}} \text{tr} \left( D_a \left( P_A \hat{\rho}_{AR}^{\text{cl}} P_A^\dagger \right) \right) \quad (12) \\ & = 2 \sum_{a \in \{+, -\}} \mathbb{E}_{\mathbb{P}} \text{tr} \left( D_a \left( P_A \hat{\rho}_A P_A^\dagger \right) \right) \\ & \leq 2 (\text{tr}(\Delta_+) + \text{tr}(\Delta_-)) \leq 8\varepsilon' \end{aligned}$$

We bound the third term in a similar way. We have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left\| \bar{\mathcal{T}} \left( P_A \left( (\rho_{AR} - \rho_{AR}^{\text{cl}}) - (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}}) \right) P_A^\dagger \right) \right\|_1 \\ \leq \sum_{a \in \{+, -\}} \mathbb{E}_{\mathbb{P}} \text{tr} \left( \bar{\mathcal{T}} \left( P_A \Gamma_a P_A^\dagger \right) \right) \\ \leq \text{tr}(\Gamma_+) + \text{tr}(\Gamma_-) \leq 8\varepsilon. \end{aligned} \quad (13)$$

Substituting the expressions (12) and (13) into Inequality (11) shows that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left\| \hat{\mathcal{T}} \left( P_A (\hat{\rho}_{AR} - \hat{\rho}_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1 \\ \geq \mathbb{E}_{\mathbb{P}} \left\| \bar{\mathcal{T}} \left( P_A (\rho_{AR} - \rho_{AR}^{\text{cl}}) P_A^\dagger \right) \right\|_1 - 8\varepsilon - 8\varepsilon' \end{aligned}$$

and an application of Theorem 3.1 concludes the proof. ■

#### IV. FROM DEQUANTIZING TO CODING

In the following subsections we show how the dequantizing theorem (Theorem 3.5) yields a one-shot coding theorem for classical information. Then, in the last subsection, we apply this coding theorem to the iid scenario and obtain the HSW theorem as a corollary.

##### A. Sending classical information through a quantum channel

Consider the following scenario: Alice wants to send a classical message  $M$  to Bob using a quantum channel  $\mathcal{T}_{A \rightarrow B}$ . For this purpose she encodes her message using an encoding TPCPM  $\mathcal{E}_{M \rightarrow A}$  into the input  $A$  of the channel. Having received the output of the channel  $B$ , Bob will apply a decoding TPCPM  $\mathcal{D}_{B \rightarrow \hat{M}}$  aiming to recover the original message. Since we are interested in the transmission of classical data through a quantum channel, the operation  $\mathcal{E}$  can be assumed to be classical, while  $\mathcal{T}$  is a CQ-channel.

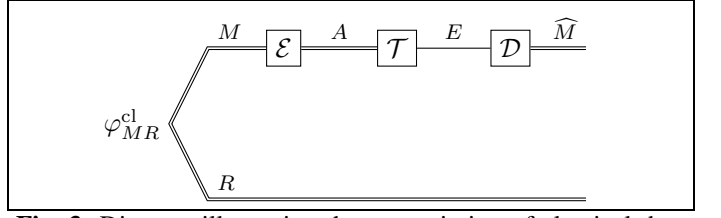
Alice's message  $M$  is assumed to exhibit perfect correlations with some classical reference system  $R$ . This means that the joint state of the message and the reference can be represented by the operator  $\varphi_{MR}^{\text{cl}} = \sum_{i=1}^{d_M} \lambda_i |ii\rangle\langle ii|_{MR}$  for some probability distribution  $\lambda$ . The aim is that after decoding, Bob holds a system  $\hat{M}$  which contains Bob's decoded message. Naturally, we want  $\hat{M}$  to contain the same message as  $M$  with high probability, which is equivalent to saying that  $\hat{M}$  is almost perfectly correlated to the reference system  $R$ . (See Figure 3 for an illustration of this scenario.) Mathematically, this means that we want the probability of error  $p_e$  to be bounded as

$$2p_e = \left\| (\mathcal{D}_{B \rightarrow \hat{M}} \circ \mathcal{T}_{A \rightarrow B} \circ \mathcal{E}_{M \rightarrow A})(\varphi_{MR}^{\text{cl}}) - \varphi_{MR}^{\text{cl}} \right\|_1 \leq \varepsilon.$$

In other words, the state after encoding, the channel, and decoding is within  $\varepsilon$  in trace distance to a state that is perfectly correlated between  $\hat{M}$  and the copy of the message in  $R$ .

##### B. The purified picture

To apply the dequantizing theorem, it is necessary to work with pure states and operations. Hence, for our derivation we will consider the setup depicted in Figure 3, but where all states and operations are replaced with the corresponding purifications.



**Fig. 3:** Diagram illustrating the transmission of classical data through a quantum channel. The state  $\varphi_{MR}^{\text{cl}}$  represents perfect classical correlations of a message  $M$  and a reference  $R$ . The aim is to obtain after decoding a system  $\hat{M}$  with nearly perfect classical correlations to  $R$ .

The present subsection shows how the dequantizing theorem (Theorem 3.1) can be used to show the existence of a decoding operation; in the following subsection, we will apply the theorem to get the encoder.

First, we purify the state  $\varphi_{MR}^{\text{cl}}$  from above to  $|\varphi\rangle_{MR} = \sum_i \sqrt{\lambda_i} |ii\rangle_{MR}$ . Next, we will assume that the encoder  $\mathcal{E}$  is actually a partial isometry  $V_{M \rightarrow A}$ ; this is slightly less general, but it will turn out to be enough for our purposes. Then, we replace  $\mathcal{T}$  by its Stinespring dilation  $U_{A \rightarrow BXE}^{\mathcal{T}}$ , where  $X$  contains a copy of the classical input, as explained in the preliminaries (Section II). Likewise, the decoder  $\mathcal{D}$  becomes the partial isometry  $U_{B \rightarrow \hat{M}E_{\mathcal{D}}}^{\mathcal{D}}$ , with an “environment” system  $E_{\mathcal{D}}$ . See Figure 4 for an illustration of the purified picture.

Now, we will show that if dequantizing holds, then a suitable decoder must exist. Consider the two states  $\rho$  and  $\hat{\rho}$  in Figure 4, which are the states immediately before and immediately after the decoder, and look at the reduced states on  $R$ ,  $E$  and  $X$ . Since these subsystems are untouched by the decoder, we have that  $\rho_{RXE} = \hat{\rho}_{RXE}$ , and  $\rho_{RXEB}$  and  $\hat{\rho}_{RXE\hat{M}E_{\mathcal{D}}}$  are both purifications of this state. The decoder is then simply the partial isometry that relates them and which is guaranteed to exist by the unitary equivalence of purifications. Hence, as long as  $\rho_{RXE}$  is of the right form, we know that a suitable decoder must exist.

We must now find out what this “right form” is. Note that since the encoder is classical, and the channel is CQ, one can show that  $\rho_{RXE\hat{M}E_{\mathcal{D}}}$  must have the form

$$|\rho\rangle_{RXE\hat{M}E_{\mathcal{D}}} = \sum_i \sqrt{\lambda_i} |i\rangle_R \otimes |\pi(i)\rangle_X \otimes |\psi^i\rangle_{E\hat{M}E_{\mathcal{D}}},$$

for some set of states  $\{|\psi^i\rangle\}$  and some permutation  $\pi$ . Furthermore, we know that the error probability must be low; this means that  $|\rho\rangle$  must be close to a state of the form

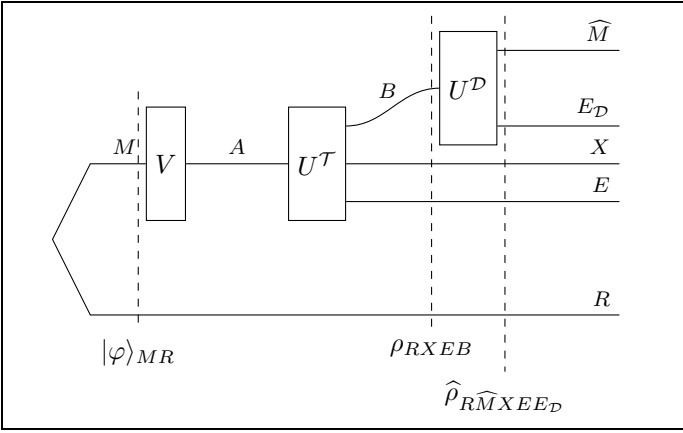
$$|\xi\rangle_{RXE\hat{M}E_{\mathcal{D}}} = \sum_i \sqrt{\lambda_i} |i\rangle_R \otimes |\pi(i)\rangle_X \otimes |i\rangle_{\hat{M}} \otimes |\theta^i\rangle_{EE_{\mathcal{D}}},$$

where here the decoder output  $\hat{M}$  is perfectly correlated with the message in  $R$  (and  $\{|\theta^i\rangle\}$  is some set of states). Tracing out  $\hat{M}E_{\mathcal{D}}$  in  $\xi$ , we get

$$\xi_{RXE} = \sum_i \lambda_i |i\pi(i)\rangle\langle i\pi(i)|_{RX} \otimes \theta_{E}^i.$$

Note that this  $\xi$  has only classical correlations between the three systems — this is the “right form” that we need for the channel output.





**Fig. 4:** Diagram illustrating the completely purified scenario. Since  $\mathcal{T}$  is a CQ-channel the corresponding environment can be split up in two parts  $X$  and  $E$ .

Hence, we have reduced the problem to finding an encoder that ensures that the output of the channel has almost only classical correlations, and this is precisely what the dequantizing theorem does.

### C. A one-shot classical coding theorem

We now put the pieces together and derive a coding theorem based on the argument in the previous subsection. We need to obtain an encoding operation with the property that  $\text{tr}_B(\bar{\mathcal{T}} \circ \mathcal{E}(\varphi_{MR})) \approx \sum_i \lambda_i |i\rangle\langle i|_{RX} \otimes \theta^i_E$ . On the other hand, note that by sending the *classically* correlated state  $\varphi_{MR}^{\text{cl}}$  through the channel (as opposed to  $|\varphi\rangle_{MR}$ ), we automatically get a state of this form, regardless of the encoder. Our strategy will therefore be to show that

$$\bar{\mathcal{T}}(\mathcal{E}(\varphi_{MR})) \approx \bar{\mathcal{T}}(\mathcal{E}(\varphi_{MR}^{\text{cl}})). \quad (14)$$

We are now in a position to use the dequantizing theorem. The encoder is constructed as follows: we first embed the message  $M$  into the input space of the channel  $A$ . (We could denote this using a partial isometry, but to avoid cluttering the notation we will simply consider  $\mathcal{H}_M$  to be a subspace of  $\mathcal{H}_A$  from now on.) We then apply a permutation on the basis elements of  $A$ , as in Theorem 3.5. We will then show that, if we average over the choice of permutations, this strategy works. It then follows that a suitable permutation exists.

Applying Theorem 3.5 to the scenario at hand, we get that there is a permutation operator such that

$$\begin{aligned} & \left\| \bar{\mathcal{T}}(P_A(\varphi_{MR} - \varphi_{MR}^{\text{cl}})P_A^\dagger) \right\|_1 \\ & \leq \sqrt{\frac{1}{d_A - 1}} 2^{-H_{\min}^\varepsilon(A|EX)_\omega - H_{\min}(M|R)_\varphi} + 8\varepsilon. \end{aligned}$$

This gives precise bounds for (14) above. We now get the decoder using Corollary 2.2: there exists a TPCPM  $\mathcal{D}_{B \rightarrow \widehat{M}}$  such that

$$\begin{aligned} & \left\| \mathcal{D}(U_T P_A(\varphi_{MR} - \varphi_{MR}^{\text{cl}})P_A^\dagger U_T^\dagger) \right\|_1 \\ & \leq 2 \sqrt{\sqrt{\frac{1}{d_A - 1}} 2^{-H_{\min}^\varepsilon(A|EX)_\omega - H_{\min}(M|R)_\varphi} + 8\varepsilon}. \end{aligned}$$

Tracing out the systems  $X$  and  $E$ , we get

$$\begin{aligned} & \left\| \mathcal{D} \circ \mathcal{T}(P_A \varphi_{MR} P_A^\dagger) - \varphi_{\widehat{M}}^{\text{cl}} \right\|_1 \\ & \leq 2 \sqrt{\sqrt{\frac{1}{d_A - 1}} 2^{-H_{\min}^\varepsilon(A|EX)_\omega - H_{\min}(M|R)_\varphi} + 8\varepsilon}. \end{aligned}$$

By the duality of the smooth min- and max-entropies [22], we have  $-H_{\min}(M|R)_\varphi = H_{\max}(M)_\varphi$  and  $-H_{\min}^\varepsilon(A|EX)_\omega = H_{\max}^\varepsilon(A|B)_\omega$ , yielding the following theorem:

**Theorem 4.1:** Let  $\mathcal{T}_{A \rightarrow B}$  be a quantum channel with Choi-Jamiołkowski representation  $\omega_{AB}$ , and let  $|\varphi\rangle_{MR} = \sum_i \sqrt{\lambda_i} |ii\rangle_{MR}$ , where  $\lambda_i$  is a probability distribution. Then, we have that there exists a permutation  $P_A$  on the basis elements of  $A$  such that

$$\begin{aligned} & \left\| \mathcal{D} \circ \mathcal{T}(P_A \varphi_{MR} P_A^\dagger) - \varphi_{\widehat{M}}^{\text{cl}} \right\|_1 \\ & \leq 2 \sqrt{\sqrt{\frac{1}{d_A - 1}} 2^{H_{\max}^\varepsilon(A|B)_\omega + H_{\max}(M)_\varphi} + 8\varepsilon}. \quad (15) \end{aligned}$$

Hence, if the right-hand side of the above inequality is small enough, then the scheme succeeds. We formulate this fact as the following corollary of the preceding theorem:

**Corollary 4.2:** Let  $\mathcal{T}_{A \rightarrow B}$  be a quantum channel with Choi-Jamiołkowski representation  $\omega_{AB}$ , and let  $\mathcal{M}$  be a set of messages with  $|\mathcal{M}| \leq d_A$  and  $p$  a probability distribution on  $\mathcal{M}$ . Then, there exists an encoder and a decoder for  $\mathcal{T}_{A \rightarrow B}$  with error probability  $p_e \geq 0$  if

$$H_{\max}(M)_p \leq \log d_A - H_{\max}^\varepsilon(A|B)_\omega - 1 + 2 \log(p_e^2 - 8\varepsilon).$$

for some constant  $\varepsilon$ ,  $0 \leq \varepsilon \leq \frac{p_e^2}{8}$ .

*Proof:* The above inequality ensures that the right-hand side of (15) is at most  $2p_e$ . ■

### D. The i.i.d scenario and the HSW theorem

Applying the above to a channel of the form  $\mathcal{T}_{A \rightarrow B}^{\otimes n}$  allows us to easily recover the HSW theorem. Recall that the HSW theorem states that there exists a family of codes for  $\mathcal{T}$  with increasing block length  $n$  with a vanishing error probability as  $n \rightarrow \infty$  as long as the rate  $Q$  is less than  $I(X, B)_\tau$ , where  $\tau$  is a state of the form  $\tau_{XB} = \sum_x p_x |x\rangle\langle x|_X \otimes \mathcal{T}(\sigma_A^x)$ , where  $p_x$  forms a probability distribution. (The rate of a code for  $n$  uses of a channel is  $\frac{1}{n} \log K$ , where  $K$  is the number of possible messages sent.) The only challenge facing us when attempting to prove this is to relate the quantity  $\log d_A$  to  $nH(X)_\tau$ . We do this using the idea of *types*, explained very briefly in Appendix B. The result is the following theorem:

**Theorem 4.3 (Holevo [11], Schumacher-Westmoreland [18]):** Let  $\mathcal{T}_{A \rightarrow B}$  be a CQ channel, let  $q$  be a probability distribution over the set  $\mathfrak{X}$ , and let  $\tau_{XB} := \sum_{x \in \mathfrak{X}} q(x) |x\rangle\langle x|_X \otimes \mathcal{T}(\sigma_A^x)$ , where  $\{\sigma_A^x : x \in \mathfrak{X}\}$  is a set of states on  $A$ . Then, there exists a family of codes for  $\mathcal{T}^{\otimes n}$  whose rate approaches  $I(X; B)_\tau$ .

*Proof:* Let  $p \in \mathcal{P}_n(\mathfrak{X})$  be the most likely type under the distribution  $q^n$ , and let  $A'$  be a system with  $\mathcal{H}_{A'} \subseteq \mathcal{H}_A^{\otimes n}$  of dimension  $|t(p)|$  defined as

$$\mathcal{H}_{A'} := \text{span}\{|\vec{x}\rangle_{A^n} : \vec{x} \in t(p)\}.$$

Now, consider the channel  $\mathcal{T}'_{A' \rightarrow B^{\otimes n}}$  defined as follows:

$$\mathcal{T}'(\xi_{A'}) = \sum_{\vec{x} \in t(p)} \langle \vec{x} | \xi | \vec{x} \rangle \mathcal{T}^{\otimes n}(\sigma_{\vec{x}}),$$

where  $\sigma_{\vec{x}} = \sigma_{x_1} \otimes \cdots \otimes \sigma_{x_n}$ . Furthermore, let  $\omega'_{A'B^n}$  be the Choi-Jamiołkowski state of  $\mathcal{T}'$ , and let us apply Corollary 4.2 to the uniform distribution over some message set  $\mathcal{M}$  and channel  $\mathcal{T}'$ . We get that there exists an encoder and a decoder such that

$$\begin{aligned} \log |\mathcal{M}| &\leq \log d_{A'} - H_{\max}^{\varepsilon'}(A'|B^n)_{\omega} - o(n) \\ &= \log |t(p)| - H_{\max}^{\varepsilon'}(A'|B^n)_{\omega} - o(n) \end{aligned}$$

From the properties of types, we have that  $|t(p)| \geq |\mathcal{P}_n(\mathfrak{X})|^{-1} 2^{nH(X)_\tau}$ . Also, note that

$$\omega_{A'B^n} = \Pi_{t(p)} \tau_{XB}^{\otimes n} \Pi_{t(p)} / \text{tr}[\Pi_{t(p)} \tau_{XB}^{\otimes n}].$$

Using Lemma A.2, we get that

$$H_{\max}^{\varepsilon}(A'|B^n)_{\omega'} \leq H_{\max}^{\varepsilon}(X^n|B^n)_{\tau^{\otimes n}} + \log \text{tr}[\Pi_{t(p)} \rho_{XB}^{\otimes n}].$$

Hence, we now have that

$$\begin{aligned} \log |\mathcal{M}| &\leq \log |t(p)| \\ &\quad - H_{\max}^{\varepsilon}(X^n|B^n)_{\tau^{\otimes n}} - \log \text{tr}[\Pi_{t(p)} \tau_{XB}^{\otimes n}] - o(n). \end{aligned}$$

Choosing  $\log |\mathcal{M}| = nQ$  for a transmission rate of  $Q$  and using the above bound on  $|t(p)|$  and the fact that the most likely type  $p$  satisfies  $\text{tr}[\Pi_{t(p)} \tau_{XB}^{\otimes n}] \geq |\mathcal{P}_n(\mathfrak{X})|^{-1}$ , this bound becomes:

$$Q \leq H(X)_\tau - \frac{1}{n} H_{\max}^{\varepsilon}(X^n|B^n)_{\tau^{\otimes n}} - o(1).$$

Taking the limit as  $n \rightarrow \infty$ , we get that this bound is satisfied whenever

$$Q < H(X)_\tau - H(X|B)_\tau = I(X; B)_\tau,$$

where we have used the fully quantum asymptotic equipartition property of [21] to bound the  $H_{\max}$  term above. Since this is true for any  $\varepsilon > 0$ , the theorem holds. ■

## V. CONCLUSION AND FURTHER WORK

In this article, we show that it is possible to derive direct bounds for the capacity of classical-quantum channels using decoupling-like techniques, therefore adding the transmission of classical data to the list of problems that are amenable to the decoupling approach to coding. Our derivation also naturally leads to bounds in the one-shot setting, where the channel is only used once and we allow a finite error probability.

We want to emphasize, however, that the bounds resulting from our calculation are somewhat weaker than the best known one-shot direct bounds, found for example in Mosonyi and Datta [14] and Wang and Renner [24]. Furthermore, our one-shot result only applies to uniform inputs of the channel

and the method of types is needed to shape the input into this form in order to achieve the HSW capacity. The latter weakness could potentially be overcome inside the decoupling framework, using an analogue of Theorem 3.14 in [7].

## APPENDIX A

### TECHNICAL FACTS ABOUT THE SMOOTH ENTROPIES

Here we establish some useful properties of the (smooth) min-entropy of classically coherent states. In particular for a state  $\rho_{XX'AB}$  that is classically coherent between  $X$  and  $X'$  we show that the state  $\sigma_{X'B}$  that optimizes

$$\begin{aligned} H_{\min}(XA|X'B)_\rho \\ = \max_{\sigma_{X'B}} \sup \{ \lambda \in \mathbb{R} : \tilde{\rho}_{XX'AB} \leq 2^{-\lambda} \mathbb{1}_{XA} \otimes \sigma_{X'B} \} \end{aligned}$$

can be chosen to have CQ-structure. Furthermore we show how the min-entropy of a classically coherent state behaves under smoothing. The following lemma is a direct consequence of the results obtained in [22].

*Lemma A.1:* Let  $\rho_{XX'AB}$  be coherent classical on  $X$  and  $X'$ . Then, there exists a state  $\tilde{\rho}_{XX'AB} \in \mathcal{B}^\varepsilon(\rho_{XX'AB})$  that is coherent classical on  $X$  and  $X'$  and a state  $\sigma_{X'B} \in \mathcal{S}_{\leq}(\mathcal{H}_{X'B})$  that is classical on  $X'$  such that

$$\begin{aligned} H_{\min}^\varepsilon(XA|X'B)_\rho \\ = \sup \{ \lambda \in \mathbb{R} : \tilde{\rho}_{XX'AB} \leq 2^{-\lambda} \mathbb{1}_{XA} \otimes \sigma_{X'B} \}. \end{aligned}$$

*Proof:* Let  $\hat{\rho}_{XX'AB} \in \mathcal{B}^\varepsilon(\rho_{XX'AB})$  be a state that maximizes the smooth min-entropy, namely it satisfies  $H_{\min}^\varepsilon(XA|X'B)_\rho = H_{\min}(XA|X'B)_{\hat{\rho}}$ . Then, the state  $\tilde{\rho}_{XX'AB} = P_{XX'} \hat{\rho}_{XX'AB} P_{XX'}$  satisfies the criteria.

First, note that we have  $P(\tilde{\rho}_{XX'AB}, \rho_{XX'AB}) \leq P(\hat{\rho}_{XX'AB}, \rho_{XX'AB}) \leq \varepsilon$  due to the monotonicity of the purified distance under projections [22]. Second, by definition of the smooth min-entropy, there exists an operator  $\hat{\sigma}_{X'B}$  such that, for  $\lambda = H_{\min}^\varepsilon(XA|X'B)$ , we have

$$\hat{\rho}_{XX'AB} \leq 2^{-\lambda} \mathbb{1}_{XA} \otimes \hat{\sigma}_{X'B}.$$

Thus,

$$\begin{aligned} \tilde{\rho}_{XX'AB} &\leq 2^{-\lambda} P_{XX'} (\mathbb{1}_{XA} \otimes \hat{\sigma}_{X'B}) P_{XX'} \\ &= 2^{-\lambda} \sum_x |x\rangle\langle x|_X \otimes \mathbb{1}_A \otimes |x\rangle\langle x|_{X'} \otimes \langle x | \hat{\sigma}_{X'B} | x \rangle \\ &\leq 2^{-\lambda} \mathbb{1}_{XA} \otimes \underbrace{\sum_x |x\rangle\langle x|_{X'} \otimes \langle x | \hat{\sigma}_{X'B} | x \rangle}_{=:\sigma_{X'B}}. \end{aligned} \quad (16)$$

Finally, we note that  $\text{tr}(\sigma_{X'B}) \leq 1$  and, thus, Eq. (16) implies that  $H_{\min}^\varepsilon(XA|X'B)_\rho \geq \lambda$ , which concludes the proof. ■

*Lemma A.2:* Let  $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ , let  $\Pi_A$  be an operator such that  $0 \leq \Pi_A \leq \mathbb{1}_A$ , and let  $\varepsilon \geq 0$ . Furthermore, let  $\rho'_{AB} := \Pi_A \rho_{AB} \Pi_A$ . Then,  $H_{\max}^\varepsilon(A|B)_{\rho'} \leq H_{\max}^\varepsilon(A|B)_\rho$ .

*Proof:* Let  $\tilde{\rho}_{AB} \in \mathcal{B}_\varepsilon(\rho)$  and  $\sigma_B$  be such that

$$2^{H_{\max}^\varepsilon(A|B)_\rho} = F(\tilde{\rho}_{AB}, \mathbb{1}_A \otimes \sigma_B)^2.$$

Let  $\tilde{\rho}' := \Pi \tilde{\rho} \Pi \in \mathcal{B}_\varepsilon(\rho')$ , and let  $\omega_B$  be such that

$$2^{H_{\max}^\varepsilon(A|B)_{\tilde{\rho}'}} = F(\tilde{\rho}'_{AB}, \mathbb{1}_A \otimes \omega_B)^2.$$

Then, we immediately have that

$$\begin{aligned}
2^{H_{\max}^{\varepsilon}(A|B)_{\rho}} &= F(\tilde{\rho}_{AB}, \mathbb{1}_A \otimes \sigma_B)^2 \\
&\geq F(\tilde{\rho}_{AB}, \mathbb{1}_A \otimes \omega_B)^2 \\
&\geq F(\tilde{\rho}_{AB}, \Pi_A^2 \otimes \omega_B)^2 \\
&= \text{tr} \left[ \sqrt{(\Pi_A \otimes \sqrt{\omega_B}) \tilde{\rho}_{AB} (\Pi_A \otimes \sqrt{\omega_B})} \right]^2 \\
&= \text{tr} \left[ \sqrt{(\mathbb{1}_A \otimes \sqrt{\omega_B}) \Pi_A \tilde{\rho}_{AB} \Pi_A (\mathbb{1}_A \otimes \sqrt{\omega_B})} \right]^2 \\
&= F(\Pi_A \tilde{\rho}_{AB} \Pi_A, \mathbb{1}_A \otimes \omega_B)^2 \\
&= F(\tilde{\rho}'_{AB}, \mathbb{1}_A \otimes \omega_B)^2 \\
&= 2^{H_{\max}(A|B)_{\rho'}} \\
&\geq 2^{H_{\max}^{\varepsilon}(A|B)_{\rho'}}.
\end{aligned}$$

Taking logarithms then yields the lemma.  $\blacksquare$

#### APPENDIX B THE METHOD OF TYPES

The ‘‘method of types’’ is a technique that is used extensively in classical information theory and that we need here to make statements about discrete memoryless channels. For a complete introduction to this method, we refer the reader to [4]; we will only give here the facts needed for our paper. The basic idea goes as follows. Let  $\mathfrak{X}$  be a finite set, and let  $\vec{x} = x_1 \dots x_n \in \mathfrak{X}^n$  be a sequence of  $n$  symbols from  $\mathfrak{X}$ . For any  $x \in \mathfrak{X}$ , let  $p_{\vec{x}}(x)$  be the relative frequency of the symbol  $x$  in  $\vec{x}$  (i.e. the number of occurrences of  $x$  in  $\vec{x}$  divided by  $n$ ). We call the distribution  $p_{\vec{x}}$  the *type* of  $\vec{x}$ , and, given a type  $p$ , we define  $t(p)$  to be the set of all sequences of type  $p$ . Furthermore, we define  $\mathcal{P}_n(\mathfrak{X})$  to be the set of all types over  $\mathfrak{X}^n$ .

We now list some basic properties of types:

- $|\mathcal{P}_n(\mathfrak{X})| = \binom{n+|\mathfrak{X}|-1}{|\mathfrak{X}|-1}$ .
- For any type  $p \in \mathcal{P}_n(\mathfrak{X})$ , we have that
$$|\mathcal{P}_n(\mathfrak{X})|^{-1} 2^{nH(p)} \leq |t(p)| \leq 2^{nH(p)}.$$
- For any type  $p \in \mathcal{P}_n(\mathfrak{X})$  and any probability distribution  $q$  over  $\mathfrak{X}$ , we have that

$$|\mathcal{P}_n(\mathfrak{X})|^{-1} 2^{-nD(p||q)} \leq \sum_{\vec{x} \in t(p)} q^n(\vec{x}) \leq 2^{-nD(p||q)}.$$

- For any probability distribution  $q$  and any  $n$ , the most likely type  $p$  has total probability

$$\sum_{\vec{x} \in t(p)} q^n(\vec{x}) \geq |\mathcal{P}_n(\mathfrak{X})|^{-1}.$$

Note that  $|\mathcal{P}_n(\mathfrak{X})|$  is polynomial in  $n$  and becomes negligible in most expressions involving exponentials of entropies.

To use these concepts in quantum information, we will define *type projectors*. Let  $X$  be a  $|\mathfrak{X}|$ -dimensional quantum system, with a basis vector  $|x\rangle$  for each  $x \in \mathfrak{X}$ . Let  $p \in \mathcal{P}_n(\mathfrak{X})$ ; we define the type projector  $\Pi_{t(p)}$  as

$$\Pi_{t(p)} = \sum_{\vec{x} \in t(p)} |\vec{x}\rangle\langle\vec{x}|,$$

where  $|\vec{x}\rangle = |x_1\rangle \otimes \dots \otimes |x_n\rangle$  for  $\vec{x} = x_1 \dots x_n$ .

#### ACKNOWLEDGMENTS

We thank Renato Renner for discussions. FD acknowledges support from the Swiss National Science Foundation (grants PP00P2-128455), the National Centre of Competence in Research ‘‘Quantum Science and Technology’’ and the German Science Foundation (grants CH 843/1-1 and CH 843/2-1). OS acknowledges funding by the Elite Network of Bavaria (ENB) project QCCC. MT acknowledges support from the National Research Foundation (Singapore), and the Ministry of Education (Singapore).

#### REFERENCES

- [1] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructuring quantum information’s family tree. *Proceedings of the Royal Society A*, 465(2108):2537–2563, 2009. quant-ph/0606225.
- [2] R. Bhatia. *Matrix Analysis*. Springer, 1996.
- [3] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear algebra and its applications*, 10(3):285–290, 1975.
- [4] I. Csiszár. The method of types. *IEEE Transactions on Information Theory*, 44(6):2505–2523, 1998.
- [5] I. Devetak, A. Harrow, and A. Winter. A resource framework for quantum Shannon theory. *IEEE Transactions on Information Theory*, 54(10):4587–4618, oct. 2008. quant-ph/0512015.
- [6] F. Dupuis. The capacity of quantum channels with side information at the transmitter. *Proceedings of the 2009 IEEE International Symposium on Information Theory*, 2009. arXiv:0805.3352.
- [7] F. Dupuis. *The decoupling approach to quantum information theory*. PhD thesis, Université de Montréal, 2009. arXiv:1004.1641.
- [8] F. Dupuis, M. Berta, J. Wullschlegler, and R. Renner. The decoupling theorem. 2010. arXiv:1012.6044.
- [9] F. Dupuis, P. Hayden, and K. Li. A father protocol for quantum broadcast channels. *IEEE Transactions on Information Theory*, 56(6):2946–2956, june 2010. quant-ph/0612155.
- [10] P. Hayden, M. Horodecki, J. Yard, and A. Winter. A decoupling approach to the quantum capacity. *Open Systems and Information Dynamics*, 15:7–19, 2008. quant-ph/0702005.
- [11] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44:269–273, 1998.
- [12] M. Horodecki, J. Oppenheim, and A. Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269(1):107–136, 2007. quant-ph/0512247.
- [13] A. Jamiolkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.
- [14] M. Mosonyi and N. Datta. Generalized relative entropies and the capacity of classical-quantum channels. *J. Math. Phys.*, 50(072104), 2009. arXiv:0810.3478.
- [15] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, New York, NY, USA, 2000.
- [16] J. Renes and R. Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Transactions on Information Theory*, 57(11):7377–7385, nov. 2011. arXiv:1012.4814.
- [17] R. Renner. *Security of quantum key distribution*. PhD thesis, ETH Zürich, 2005. arXiv:quant-ph/0512258v2.
- [18] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131–138, 1997.
- [19] O. Szehr. *Decoupling Theorems*. Master thesis, ETH Zürich, 2011.
- [20] M. Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, ETH Zurich, 2012.
- [21] M. Tomamichel, R. Colbeck, and R. Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, dec. 2009. arXiv:0811.1221.
- [22] M. Tomamichel, R. Colbeck, and R. Renner. Duality between smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 56(9):4674–4681, sept. 2010. arXiv:0907.5238.
- [23] A. Uhlmann. The ‘transition probability’ in the state space of a \*-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.
- [24] L. Wang and R. Renner. One-shot classical-quantum capacity and hypothesis testing. *Phys. Rev. Lett.*, 108:200501, May 2012. arXiv:1007.5456.