Commun. Math. Phys. 328, 251–284 (2014)
Digital Object Identifier (DOI) 10.1007/s00220-014-1990-4

Communications in
Mathematical
Physics

One-Shot Decoupling

Frédéric Dupuis[1], **Mario Berta**[1], **Jürg Wullschleger**[2,3], **Renato Renner**[1]

[1] Institute for Theoretical Physics, ETH Zurich, Zurich, Switzerland. E-mail: berta@phys.ethz.ch
[2] Department of Computer Science and Operations Research, Université de Montréal, Montreal, QC, Canada
[3] McGill University, Montreal, QC, Canada

Received: 5 November 2012 / Accepted: 22 December 2013
Published online: 21 March 2014 – © The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract: If a quantum system A, which is initially correlated to another system, E, undergoes an evolution separated from E, then the correlation to E generally decreases. Here, we study the conditions under which the correlation disappears (almost) completely, resulting in a decoupling of A from E. We give a criterion for decoupling in terms of two smooth entropies, one quantifying the amount of initial correlation between A and E, and the other characterizing the mapping that describes the evolution of A. The criterion applies to arbitrary such mappings in the general one-shot setting. Furthermore, the criterion is tight for mappings that satisfy certain natural conditions. One-shot decoupling has a number of applications both in physics and information theory, e.g., as a building block for quantum information processing protocols. As an example, we give a one-shot state merging protocol and show that it is essentially optimal in terms of its entanglement consumption/production.

1. Introduction

Correlations in quantum systems, and in particular entanglement, have been in the focus of (both theoretical and experimental) research in quantum information science over the past decades. As a result, one has nowadays a pretty good (although still not complete) understanding of quantum correlations and, in particular, the processes that create them. In this work, we take—so to speak—an opposite approach and study conditions under which two systems can be decoupled, i.e., brought to a state where they are uncorrelated.

We call a system, B, decoupled from another system, E, if the joint state of the two systems, ρ_{BE}, has product form $\rho_B \otimes \rho_E$. Operationally, this means that the outcome of any measurement on B is statistically independent of the outcome of any measurement on E. Or, in information-theoretic terms, the system E does not give any information on B (and can therefore safely be ignored when studying B).

Decoupling theorem. Our goal is to characterize the conditions under which the evolution of a system results in decoupling. For this, we consider a system, A, that may

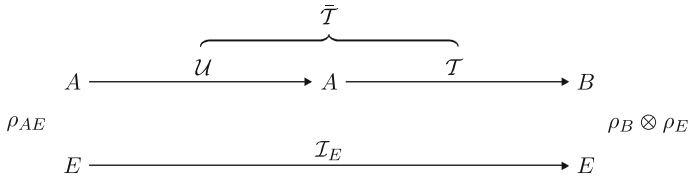


Fig. 1. Decoupling. The initial system, A, may be correlated to a reference system E. The evolution is modeled as a mapping $\bar{\mathcal{T}}$ from A to B. The final state of B is supposed to be independent of E. The subdivision of $\bar{\mathcal{T}}$ into a unitary $\mathcal{U}$ and a mapping $\mathcal{T}$ is required for the formulation of our decoupling criterion

initially be correlated to E. Furthermore, we assume that the system A undergoes an evolution, described by a TPCPM[1] $\bar{\mathcal{T}}$ from A to B, during which no interaction with E takes place (see Fig. 1). The main result of this work is a decoupling theorem, i.e., a criterion that provides necessary and sufficient conditions for decoupling (of B from E). The criterion depends on two entropic quantities, characterizing the initial state, ρ_{AE}, and the mapping $\bar{\mathcal{T}}$, respectively.

The decoupling criterion can be conceptually split into two parts, called achievability and converse part, which we now describe informally. The full technical statements are provided as Theorems 3.1 and 4.1 in Sects. 3 and 4, respectively. For their formulation, it is convenient to view $\bar{\mathcal{T}}$ as a sequence, $\bar{\mathcal{T}} = \mathcal{T} \circ \mathcal{U}$, where $\mathcal{U}$ is an arbitrary unitary on A, and $\mathcal{T}$ a fixed TPCPM from A to B.

Achievability: decoupling up to an error ε is achieved for most choices of $\mathcal{U}$ if

$$H_{\min}^{\varepsilon}(A|E)_{\rho} + H_{\min}^{\varepsilon}(A|B)_{\tau} \gtrapprox 0. \tag{1}$$

Converse: decoupling up to an error ε is not achieved for any choice of $\mathcal{U}$ if

$$H_{\min}^{\varepsilon}(A|E)_{\rho} + H_{\max}^{\varepsilon}(A|B)_{\tau} \lessapprox 0. \tag{2}$$

The criteria refer to the ε-smooth conditional min- and max-entropy introduced in [RW04, Ren05], which can be seen as generalizations of the von Neumann entropy (cf. Sect. 2 for definitions and properties). The ε-smooth conditional min-entropy $H_{\min}^{\varepsilon}(A|E)_{\rho}$ is a measure for the correlation present in the initial state ρ_{AE}—the larger this measure, the less dependent is A on E (see Table 1 for some typical examples). The quantities $H_{\min}^{\varepsilon}(A|B)_{\tau}$ (for the achievability) and $H_{\max}^{\varepsilon}(A|B)_{\tau}$ (for the converse) measure how well the mapping $\mathcal{T}$ conserves correlations. Roughly, they quantify the uncertainty one has about a "copy" of the input, A, given access to the output, B, of $\mathcal{T}$ (cf. Table 2). We note that the expressions for the achievability and for the converse essentially coincide in many cases of interest (see the discussion in Sect. 4).

As a typical example for decoupling, consider m qubits, A, that are classically maximally correlated to E (so that $H_{\min}^{\varepsilon}(A|E)_{\rho} = 0$, cf. second row of Table 1). Furthermore, assume that A undergoes a reversible evolution, $\mathcal{U}$, after which we discard $m - m'$ qubits, corresponding to a partial trace, $\mathcal{T} = \mathrm{Tr}_{m-m'}$ (see last example of Table 2). Our criterion then says that the remaining m' qubits will, for most evolutions $\mathcal{U}$, be decoupled from E whenever $m' < m/2$. Conversely, if this condition is not satisfied, some correlation will necessarily be retained.

We mention that it is possible to phrase our achievability criterion for decoupling (1) in another (but equivalent) way. For TPCPMs $\mathcal{T}$ from A to B such that for every unitary

[1] A trace-preserving completely-positive map (TPCPM) is a linear function that maps density operators to density operators.

Table 1. Dependence on the initial state

Description of initial state	$\rho = \rho_{AE}$	$H_{\min}^{\varepsilon}(A\|E)_{\rho}$
k random bits A independent of E	$2^{-k} \cdot \mathbb{1}_A \otimes \rho_E$	k
k bits A correlated classically to E	$2^{-k} \cdot \sum_{i=1}^{2^k} \|i\rangle\langle i\|_A \otimes \|i\rangle\langle i\|_E$	0
k qubits A fully entangled with E	$\|\Psi\rangle\langle\Psi\|$, where $\Psi = 2^{-k/2} \cdot \sum_{i=1}^{2^k} \|i\rangle_A \otimes \|i\rangle_E$	$-k$

The table illustrates how the term $H_{\min}^{\varepsilon}(A|E)_{\rho}$ (for $\varepsilon \to 0$) in the decoupling criterion depends on the initial state ρ_{AE}. In all three examples, A is assumed to be a k-qubit system with orthonormal basis $\{|i\rangle_A\}_{i=1}^{2^k}$. Similarly, $\{|i\rangle_E\}_{i=1}^{2^k}$ is an orthonormal family of states on E

Table 2. Dependence on the mapping

Description of mapping	$\mathcal{T}$	$H_{\min}^{\varepsilon}(A\|B)_{\tau}$
Identity on m qubits	$\sigma \mapsto \sigma$	$-m$
Orthogonal measurement on m qubits	$\sigma \mapsto \sum_{i=1}^{2^m} \|i\rangle\langle i\|\sigma\|i\rangle\langle i\|$	0
Erasure of m qubits	$\sigma \mapsto \mathrm{Tr}(\sigma)\|0\rangle\langle 0\|$	m
Identity on m', orthogonal measurement on $m - m'$ qubits	$\sigma \mapsto \sum_{i=1}^{2^{m-m'}} (\mathbb{1}_{m'} \otimes \|i\rangle\langle i\|)\sigma(\mathbb{1}_{m'} \otimes \|i\rangle\langle i\|)$	$-m'$
Identity on m', erasure on $m - m'$ qubits	$\sigma \mapsto \mathrm{Tr}_{m-m'}(\sigma)$	$m - 2m'$

The table illustrates how the term $H_{\min}^{\varepsilon}(A|B)_{\tau}$ in the decoupling criterion depends on the mapping $\mathcal{T}$. In all five examples, the input space, A, is assumed to consist of m qubits with orthonormal basis $\{|i\rangle_A\}_{i=1}^{2^m}$. The last two examples have a smaller output space consisting of only m' qubits. The penultimate one can be seen as a combination of the first and the second, and the last one can be seen as a combination of the first and the third. (The smooth conditional min-entropies are evaluated for $\varepsilon \to 0$)

$\mathcal{U}$ on A there exists a unitary $\mathcal{V}$ on B with $\mathcal{V} \circ \mathcal{T} = \mathcal{T} \circ \mathcal{U}$, decoupling up to an error ε is achieved if

$$H_{\min}^{\varepsilon}(A|E)_{\rho} + H_{\min}^{\varepsilon}(A|B)_{\tau} \gtrapprox 0. \tag{3}$$

For more details about this formulation, see the discussion in Sect. 3.1.

Applications. The notion of decoupling has various applications in information theory and in physics. Many of these applications have in common that decoupling of a system B from a system E is used to show that B is maximally entangled with a complementary system, R. Indeed, under the assumption that R is chosen such that the joint state, ρ_{BER}, is pure, $\rho_{BE} = \rho_B \otimes \rho_E$ immediately implies that there exists a subsystem R' of R such that the state on $\rho_{BR'}$ is pure. If, in addition, ρ_B is fully mixed, $\rho_{BR'}$ is necessarily maximally entangled.

In the context of information theory, this type of argument is, for example, used to analyze state merging [HOW05,HOW07], i.e., the task of conveying a subsystem from a sender to a receiver | who already holds a possibly correlated subsystem | using classical communication and entanglement. Another example, where decoupling is used in a similar fashion, is the quantum reverse Shannon theorem [BSST02,BDH+09,BCR11]. In fact, the proof of this theorem given in [BCR11] refers to a coherent form of state merging (also known as the fully quantum Slepian Wolf or mother protocol [ADHW09]) where the classical communication is replaced by quantum communication. Decoupling can also be used for the characterization of correlation and entanglement between systems, erasure processes, as well as channel capacities (see, e.g., [GPW05,Bus09,HHWY08]). In addition, its classical analogue, privacy amplification [BBCM95,RK05], is widely used in classical and quantum cryptography.

Decoupling processes are also crucial in physics. For example, the evolution of a thermodynamical system towards thermal equilibrium can be understood as a decoupling process, where the system under consideration decouples from the observer (somewhat analogous to the considerations in [LPSW09,Par89a,Par89b]). Recent work indeed shows that there is a close relation between smooth entropies and quantities that are relevant in thermodynamics [DRRV09,dRAR+11,Hut11,FDOR12,Abe13,HO13]. Similarly, black hole radiation may be analyzed from such a point of view [HP07,BP07, PZ13]. Finally, one-shot decoupling techniques were also applied in solid state physics in order to show that 1D quantum states with exponential decay of correlations have an efficient classical approximate description as a matrix product state [BH13].

History and related work. While various standard results in quantum information theory have been proved using ideas related to decoupling, the concept came into its own with the discovery of state merging protocols [HOW05,HOW07] and, later, the fully quantum Slepian Wolf protocol [ADHW09]. These are based on specific decoupling processes where the mapping $\mathcal{T}$ is either a projective measurement or a partial trace. In this early work, the decoupling was analyzed in terms of the dimensions of certain subsystems (rather than smooth conditional entropies).

Based on the diploma thesis of one of us [Ber08], we have generalized these decoupling results to include mappings $\mathcal{T}$ that consist of combinations of projective measurements and partial trace-preserving. Furthermore, we expressed the decoupling criterion in terms of smooth conditional entropies. Subsequently, one of the authors derived in his doctoral thesis [Dup09] a general decoupling theorem that can be applied to any type of mapping. This result is essentially (up to the use of different entropy measures) equivalent to Theorem 3.1 presented here. We also note that the aforementioned characterizations of decoupling can be seen as special cases of this general result.

The above work was mostly concerned with achievability. Converse results were so far only known in special cases. In particular, we derived in [BRW07] and [Ber08] (see also [Ren09]) converse theorems for the case where the mapping $\mathcal{T}$ is a projective measurement. The converse theorem presented here, Theorem 4.1, generalizes these results.

We emphasize that the use of smooth conditional entropies is essential for applications of the decoupling technique in physics (see the discussion in Sect. 6).

Structure of the paper. In Sect. 2 we introduce the notation and review the definitions and main properties of the entropy measures used in this work. Our main achievability result for decoupling is given in Sect. 3, whereas Sect. 4 contains a converse that is tight in many cases of interest. The use of the decoupling technique is illustrated in Sect. 5,

where we show how to obtain optimal one-shot quantum state merging. We conclude with a discussion in Sect. 6.

2. Preliminaries

2.1. Notation. We denote the Hilbert space associated to a system A by $\mathcal{H}_A$. We only consider finite-dimensional systems and denote the dimension of $\mathcal{H}_A$ by $|A|$. The set of linear operators on $\mathcal{H}$ is denoted by $\mathcal{L}(\mathcal{H})$ and the set of nonnegative operators on $\mathcal{H}$ by $\mathcal{P}(\mathcal{H})$. We define the sets of subnormalized states $\mathcal{S}_{\leqslant}(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : \operatorname{Tr}\rho \leqslant 1\}$ and normalized states $\mathcal{S}_{=}(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : \operatorname{Tr}\rho = 1\}$.

The tensor product of $\mathcal{H}_A$ and $\mathcal{H}_B$ is denoted by $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. For multipartite operators $\rho_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$, we write $\rho_A = \operatorname{Tr}_B(\rho_{AB})$ for the corresponding reduced operator. For $M_A \in \mathcal{L}(\mathcal{H}_A)$, we write $M_A = M_A \otimes \mathbb{1}_B$ for the enlargement on any $\mathcal{H}_{AB}$, where $\mathbb{1}_B$ denotes the identity in $\mathcal{P}(\mathcal{H}_B)$.

Completely positive maps from $\mathcal{L}(\mathcal{H}_A)$ to $\mathcal{L}(\mathcal{H}_B)$ are called CPMs and trace-preserving CPMs are called TPCPMs. For $\mathcal{H}_A$, $\mathcal{H}_B$ with orthonormal bases $\{|i\rangle_A\}_{i=1}^{|A|}$, $\{|i\rangle_B\}_{i=1}^{|B|}$ and $|A| = |B|$, the canonical identity mapping from $\mathcal{L}(\mathcal{H}_A)$ to $\mathcal{L}(\mathcal{H}_B)$ with respect to these bases is denoted by $\mathcal{I}_{A\to B}$, i.e., $\mathcal{I}_{A\to B}(|i\rangle\langle j|_A) = |i\rangle\langle j|_B$.

For $\rho \in \mathcal{P}(\mathcal{H})$, $\|\rho\|_\infty$ denotes the operator norm of ρ, which is equal to the maximum eigenvalue of ρ. The trace norm of $\rho \in \mathcal{L}(\mathcal{H})$ is defined as $\|\rho\|_1 = \operatorname{Tr}(\sqrt{\rho^\dagger\rho})$ and the induced metric on $\mathcal{S}_{\leqslant}(\mathcal{H})$ is called trace distance.[2] The fidelity between $\rho, \sigma \in \mathcal{S}_{\leqslant}(\mathcal{H})$ is defined as $F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$.

We will make use of the Choi–Jamiołkowski isomorphism, which relates CPMs to positive operators, and which we denote by J.

Lemma 2.1 [Jam72,Cho75]. *The Choi–Jamiołkowski map J takes maps $\mathcal{T}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ to operators $J(\mathcal{T}_{A\to B}) \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. It is defined as*

$$J(\mathcal{T}_{A\to B}) = (\mathcal{I}_A \otimes \mathcal{T}_{A'\to B})(|\Phi\rangle\langle\Phi|_{AA'}), \tag{4}$$

where $|\Phi\rangle_{AA'} = |A|^{-\frac{1}{2}} \sum_i |i\rangle_A \otimes |i\rangle_{A'}$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.[3] The map J bijectively maps the set of CPMs from $\mathcal{H}_A$ to $\mathcal{H}_B$ to the set $\mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$, and its inverse maps any $\gamma_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ to

$$\mathcal{T}_{A\to B} : M_A \mapsto |A| \cdot \operatorname{Tr}[\gamma_{AB} M_A^T], \tag{5}$$

where M_A^T denotes the transpose of M_A with respect to the basis $\{|i\rangle_A\}_{i=1}^{|A|}$.

2.2. Smooth entropies. The smooth entropy formalism [Ren05,RW04] has been introduced in (classical and quantum) information theory to study general one-shot scenarios, in which nothing needs to be assumed about the structure of the relevant probability distributions or quantum states (e.g., those modeling noise processes in a communication channel). The formalism therefore overcomes a limitation of the established theory, where it is usually assumed that the relevant processes can be modeled as asymptotic sequences of independent and identically distributed (iid) subprocesses.

[2] The trace distance is often defined with an additional factor $1/2$, which we omit here.

[3] The Choi–Jamiołkowski isomorphism is sometimes defined with an additional dimensional factor of $|A|$; we choose not to do this here.

In this section we provide the definitions of the underlying entropy measures, called smooth min- and max entropy, and state some of their basic properties. Further properties are summarized in Appendix A. For a more detailed discussion of the smooth entropy formalism we refer to [Tom12, Ren05, KRS09, TCR09, TCR10, Dat09].

Recall the following standard definitions. The von Neumann entropy of $\rho \in \mathcal{S}_=(\mathcal{H})$ is defined as[4] $H(\rho) = -\operatorname{Tr}(\rho \log \rho)$ and the conditional von Neumann entropy of A given B for $\rho_{AB} \in \mathcal{S}_=(\mathcal{H})$ is defined as $H(A|B)_\rho = H(AB)_\rho - H(B)_\rho$.

Definition 2.2. *Let $\rho_{AB} \in \mathcal{S}_\leqslant(\mathcal{H}_{AB})$. The conditional min-entropy of A given B is defined as*

$$H_{\min}(A|B)_\rho = \sup_{\sigma_B \in \mathcal{S}_=(\mathcal{H}_B)} \sup \left\{ \lambda \in \mathbb{R} : 2^{-\lambda} \cdot \mathbb{1}_A \otimes \sigma_B - \rho_{AB} \geqslant 0 \right\}. \tag{6}$$

The conditional max-entropy of A given B is defined as

$$H_{\max}(A|B)_\rho = \sup_{\sigma_B \in \mathcal{S}_=(\mathcal{H}_B)} \log F(\rho_{AB}, \mathbb{1}_A \otimes \sigma_B)^2. \tag{7}$$

In the special case where B is trivial (i.e., one-dimensional), we write $H_{\min}(A)_\rho$ and $H_{\max}(A)_\rho$ instead of $H_{\min}(A|B)_\rho$ and $H_{\max}(A|B)_\rho$, respectively, and it can be shown that $H_{\min}(A)_\rho = -\log \|\rho_A\|_\infty$ as well as $H_{\max}(A)_\rho = 2 \log \operatorname{Tr} \sqrt{\rho_A}$. Furthermore, for $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ the entropies can be ordered as [TCR09, Lemma 2]

$$H_{\min}(A|B)_\rho \leqslant H(A|B)_\rho \leqslant H_{\max}(A|B)_\rho. \tag{8}$$

The smooth conditional min- and max-entropy are defined by extremizing the non-smooth versions over a set of nearby states, where nearby is quantified by the purified distance.

Definition 2.3. *Let $\rho, \sigma \in \mathcal{S}_\leqslant(\mathcal{H})$. The purified distance between ρ and σ is defined as*

$$P(\rho, \sigma) = \sqrt{1 - \bar{F}(\rho, \sigma)^2}, \tag{9}$$

where $\bar{F}(\rho, \sigma) = F(\rho, \sigma) + \sqrt{(1 - \operatorname{Tr}[\rho])(1 - \operatorname{Tr}[\sigma])}$ denotes the generalized fidelity.

The purified distance is a metric on $\mathcal{S}_\leqslant(\mathcal{H})$ [TCR10, Lemma 5]. As its name indicates, $P(\rho, \sigma)$ corresponds to the minimum trace distance between purifications of ρ and σ. For more about the purified distance we refer to [TCR10].

Henceforth $\rho, \sigma \in \mathcal{S}_\leqslant(\mathcal{H})$ are called ε-close if $P(\rho, \sigma) \leqslant \varepsilon$ and this is denoted by $\rho \approx_\varepsilon \sigma$. We use the purified distance to specify an ε-ball around $\rho \in \mathcal{S}_\leqslant(\mathcal{H})$,

$$B^\varepsilon(\rho) = \{\rho' \in \mathcal{S}_\leqslant(\mathcal{H}) : \rho' \approx_\varepsilon \rho\}. \tag{10}$$

Definition 2.4. *Let $\varepsilon \geqslant 0$ and $\rho_{AB} \in \mathcal{S}_\leqslant(\mathcal{H}_{AB})$. The ε-smooth conditional min-entropy of A given B is defined as*

$$H^\varepsilon_{\min}(A|B)_\rho = \sup_{\hat{\rho}_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})} H_{\min}(A|B)_{\hat{\rho}}. \tag{11}$$

The ε-smooth conditional max-entropy of A given B is defined as

$$H^\varepsilon_{\max}(A|B)_\rho = \inf_{\hat{\rho}_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})} H_{\max}(A|B)_{\hat{\rho}}. \tag{12}$$

[4] All logarithms are taken to base 2.

We mention that the optimization problems defining the smooth conditional min- and max-entropy can be formulated as semi-definite programs [Tom12, Sect. 5.2.1]. This allows to efficiently compute them numerically.

The smooth conditional min- and max-entropy are dual to each other in the following sense.

Lemma 2.5 [TCR10, Lemma 16]. *Let $\varepsilon \geqslant 0$, $\rho_{AB} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB})$ and let $\rho_{ABC} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{ABC})$ be an arbitrary purification of ρ_{AB}. Then, we have that*

$$H_{\min}^{\varepsilon}(A|B)_{\rho} = -H_{\max}^{\varepsilon}(A|C)_{\rho}. \tag{13}$$

Smooth entropies satisfy various natural properties analogous to those known for the von Neumann entropy. One of them is the invariance under local isometries.

Lemma 2.6 [TCR10, Lemma 13/15]. *Let $\varepsilon \geqslant 0$, $\rho_{AB} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB})$, and let $\mathcal{U}_{A\to C}$ and $\mathcal{V}_{B\to D}$ be isometries from A to C and B to D, respectively. Then, we have that*

$$H_{\min}^{\varepsilon}(A|B)_{\rho} = H_{\min}^{\varepsilon}(C|D)_{\mathcal{V}\circ\mathcal{U}(\rho)} \tag{14}$$

$$H_{\max}^{\varepsilon}(A|B)_{\rho} = H_{\max}^{\varepsilon}(C|D)_{\mathcal{V}\circ\mathcal{U}(\rho)}. \tag{15}$$

Another important property is the data processing inequality.

Lemma 2.7 [TCR10, Theorem 18]. *Let $\varepsilon \geqslant 0$, $\rho_{AB} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB})$, and let $\mathcal{T}_{B\to C}$ be a TPCPM from B to C. Then, we have that*

$$H_{\min}^{\varepsilon}(A|B)_{\rho} \leqslant H_{\min}^{\varepsilon}(A|C)_{\mathcal{T}(\rho)} \tag{16}$$

$$H_{\max}^{\varepsilon}(A|B)_{\rho} \leqslant H_{\max}^{\varepsilon}(A|C)_{\mathcal{T}(\rho)}. \tag{17}$$

Smooth entropies are generalizations of the von Neumann entropy, in the sense that the von Neumann entropy can be retrieved as a special case via the quantum asymptotic equipartition property (AEP).

Lemma 2.8 [Tom12, Corollary 6.6 and 6.7]. *Let $0 < \varepsilon < 1$ and $\rho_{AB} \in \mathcal{S}_{=}(\mathcal{H}_{AB})$. Then, we have that*

$$\lim_{n\to\infty} \frac{1}{n} H_{\min}^{\varepsilon}(A|B)_{\rho^{\otimes n}} = H(A|B)_{\rho} \tag{18}$$

$$\lim_{n\to\infty} \frac{1}{n} H_{\max}^{\varepsilon}(A|B)_{\rho^{\otimes n}} = H(A|B)_{\rho}. \tag{19}$$

For more properties of smooth entropies we refer to the Appendix A and [Tom12, Ren05, KRS09, TCR09, TCR10, Dat09].

For technical reasons we will also need the following auxiliary quantities.

Definition 2.9. *Let $\rho_{AB} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB})$. The conditional collision entropy of A given B is defined as*

$$H_2(A|B)_{\rho} = \sup_{\sigma_B \in \mathcal{S}_{=}(\mathcal{H}_B)} -\log \operatorname{Tr}\left[\left((\mathbb{1}_A \otimes \sigma_B^{-1/4})\rho_{AB}(\mathbb{1}_A \otimes \sigma_B^{-1/4})\right)^2\right]. \tag{20}$$

Definition 2.10. *Let $\rho_{AB} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB})$ and $\sigma_B \in \mathcal{S}_{\leqslant}(\mathcal{H}_B)$. We define*

$$H_{\max}(A|B)_{\rho|\sigma} = \log F(\rho_{AB}, \mathbb{1}_A \otimes \sigma_B)^2. \tag{21}$$

It can be shown that $H_{\max}(A|B)_\rho = \sup_{\sigma \in \mathcal{S}_\leqslant(\mathcal{H}_B)} H_{\max}(A|B)_{\rho|\sigma}$.

Definition 2.11. *Let $\rho_{AB} \in \mathcal{S}_\leqslant(\mathcal{H}_{AB})$ and $\sigma_B \in \mathcal{S}_\leqslant(\mathcal{H}_B)$. We define*

$$H_{\min}(A|B)_{\rho|\sigma} = \sup\left\{\lambda \in \mathbb{R} : 2^{-\lambda} \cdot \mathbb{1}_A \otimes \sigma_B - \rho_{AB} \geqslant 0\right\}. \tag{22}$$

It can be shown that $H_{\min}(A|B)_\rho = \sup_{\sigma \in \mathcal{S}_\leqslant(\mathcal{H}_B)} H_{\min}(A|B)_{\rho|\sigma}$.

Finally, we note that, since all Hilbert spaces in this paper are assumed to have finite dimension, the infima and suprema in the expressions above can be replaced by minima and maxima, respectively.

3. Achievability

In this section, we present and prove a general decoupling theorem (Theorem 3.1), which corresponds to the achievability part of the criterion sketched informally in Sect. 1. The theorem subsumes and extends previous results in this direction.

3.1. Statement of the decoupling theorem. As explained in the introductory section (see Fig. 1), we consider a mapping from a system A to a system B. The mapping consists of a unitary on A, selected randomly according to the Haar measure over the unitary group on $\mathcal{H}_A$, followed by an arbitrary mapping $\mathcal{T} = \mathcal{T}_{A\to B}$. In applications, $\mathcal{T}$ often consists of a measurement or a partial trace (see Table 2 for examples). The decoupling theorem then tells us how well the output, B, of the mapping $\mathcal{T}$ is decoupled (on average over the choices of the unitary) from a reference system E.

Theorem 3.1 (Decoupling Theorem). *Let $\varepsilon > 0$, $\rho_{AE} \in \mathcal{S}_=(\mathcal{H}_{AE})$, and let $\mathcal{T}_{A\to B}$ be a CPM with Choi–Jamiołkowski representation $\tau_{AB} = J(\mathcal{T})$ such that $\mathrm{Tr}(\tau_{AB}) \leqslant 1$. Then, we have that*

$$\int_{\mathbb{U}(A)} \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 dU \leqslant 2^{-\frac{1}{2}H^\varepsilon_{\min}(A|E)_\rho - \frac{1}{2}H^\varepsilon_{\min}(A|B)_\tau} + 12\varepsilon, \tag{23}$$

where $\int \cdot\, dU$ denotes the integral over the Haar measure over the full unitary group on $\mathcal{H}_A$.

Here, the total CPM is of the form $\bar{\mathcal{T}} = \mathcal{T} \circ \mathcal{U}$ with the unitary channel $\mathcal{U}(\cdot) = U_A(\cdot)U_A^\dagger$ and U_A chosen at random. We note that, equivalently, we may think of $\bar{\mathcal{T}}$ as a channel that chooses at random a unitary U_A and outputs the choice of U_A, together with the output of $\mathcal{T}$.

The decoupling theorem (Theorem 3.1) provides a bound on the quality of decoupling that only depends on two entropic quantities, $H^\varepsilon_{\min}(A|E)_\rho$ and $H^\varepsilon_{\min}(A|B)_\tau$. The first is a measure for the correlations between A and E that are present in the initial state, ρ_{AE}. The second quantifies properties of the mapping $\mathcal{T}$, which is characterized by the bipartite state τ_{AB} obtained via the Choi–Jamiołkowski isomorphism J. Hence, in order to minimize the right hand side of (23), no channel ends up being better suited for some types of states than for others or vice-versa. Furthermore, as discussed in Sect. 4, the bound in (23) is essentially optimal in many cases of interest. We also note that, using

Markov's inequality, the expectation value over the unitaries U can be turned into a bound that holds for most unitaries. That is, for any $\mu > 0$,

$$\left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 \leqslant \frac{1}{\mu} \cdot 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A|E)_\rho - \frac{1}{2} H_{\min}^{\varepsilon}(A|B)_\tau} + \frac{12\varepsilon}{\mu} \tag{24}$$

holds with probability at least $1 - \mu$ (for U chosen according to the Haar measure).

Finally, as sketched in the introductory section, the decoupling theorem (Theorem 3.1) can also be phrased in another (but equivalent) way.

Corollary 3.2. *Let $\varepsilon > 0$, $\rho_{AE} \in \mathcal{S}_=(\mathcal{H}_{AE})$, and let $\mathcal{T}_{A\to B}$ be a CPM with Choi–Jamiołkowski representation $\tau_{AB} = J(\mathcal{T})$ such that* $\mathrm{Tr}(\tau_{AB}) \leqslant 1$. *Furthermore, assume that for every unitary channel $\mathcal{U}_A$ there exists a unitary channel $\mathcal{V}_B$ such that $\mathcal{V}_B \circ \mathcal{T}_{A\to B} = \mathcal{T}_{A\to B} \circ \mathcal{U}_A$. Then, we have that*

$$\|\mathcal{T}_{A\to B}(\rho_{AE}) - \tau_B \otimes \rho_E\|_1 \leqslant 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A|E)_\rho - \frac{1}{2} H_{\min}^{\varepsilon}(A|B)_\tau} + 12\varepsilon. \tag{25}$$

Proof. By the decoupling theorem (Theorem 3.1) for the map $\mathcal{T}_{A\to B}$, there exists a unitary U_A such that

$$\left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 \leqslant 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A|E)_\rho - \frac{1}{2} H_{\min}^{\varepsilon}(A|B)_\tau} + 12\varepsilon. \tag{26}$$

Since there exists by assumption a unitary V_B such that $\mathcal{V}_B \circ \mathcal{T}_{A\to B} = \mathcal{T}_{A\to B} \circ \mathcal{U}_A$, we get

$$\begin{aligned} \left\| \mathcal{T}_{A\to B}(\rho_{AE}) - V_B^\dagger \tau_B V_B \otimes \rho_E \right\|_1 &= \left\| V_B \mathcal{T}_{A\to B}(\rho_{AE}) V_B^\dagger - \tau_B \otimes \rho_E \right\|_1 \\ &\leqslant 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A|E)_\rho - \frac{1}{2} H_{\min}^{\varepsilon}(A|B)_\tau} + 12\varepsilon. \end{aligned} \tag{27}$$

Furthermore, again by assumption, there exists a unitary W_B such that $\mathcal{W}_B \circ \mathcal{T}_{A\to B} = \mathcal{T}_{A\to B} \circ \mathcal{U}_A^\dagger$, and hence

$$\mathcal{T}_{A\to B} = \mathcal{T}_{A\to B} \circ \mathcal{U}_A \circ \mathcal{U}_A^\dagger = \mathcal{V}_B \circ \mathcal{T}_{A\to B} \circ \mathcal{U}_A^\dagger = \mathcal{V}_B \circ \mathcal{W}_B \circ \mathcal{T}_{A\to B}. \tag{28}$$

This implies $\mathcal{V}_B^\dagger \circ \mathcal{T}_{A\to B} = \mathcal{W}_B \circ \mathcal{T}_{A\to B}$, and thus we get

$$V_B^\dagger \tau_B V_B = W_B \mathcal{T}_{A\to B}\left(\frac{\mathbb{1}_A}{|A|}\right) W_B^\dagger = \mathcal{T}_{A\to B}\left(U_A^\dagger \frac{\mathbb{1}_A}{|A|} U_A\right) = \mathcal{T}_{A\to B}\left(\frac{\mathbb{1}_A}{|A|}\right) = \tau_B. \tag{29}$$

Finally, we arrive at the claim by combining this with (27). □

To see why this alternative formulation (Corollary 3.2) is equivalent to the decoupling theorem (Theorem 3.1) we may think of the total map in Theorem 3.1 as a channel that chooses at random a unitary U_A and outputs the choice of U_A, together with the output of $\mathcal{T}$. By inspection, this total map then fulfills the assumption of Corollary 3.2.

Our first step in proving Theorem 3.1 is to prove a version involving non-smooth min-entropies (Theorem 3.3). Then, in a second step, we show that smoothing preserves the essence of the theorem. Note that Theorem 3.3 may be of interest in cases where no smoothing is required since it is slightly more general: it applies to any completely positive $\mathcal{T}$, not only trace-non-increasing ones.

Theorem 3.3 (Non-Smooth Decoupling Theorem). *Let $\rho_{AE} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AE})$ and let $\mathcal{T}_{A\to B}$ be a CPM with Choi–Jamiołkowski representation $\tau_{AB} = J(\mathcal{T})$. Then, we have that*

$$\int_{\mathbb{U}(A)} \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 dU \leqslant 2^{-\frac{1}{2}H_2(A|E)_\rho - \frac{1}{2}H_2(A|B)_\tau}, \tag{30}$$

where $\int \cdot\, dU$ denotes the integral over the Haar measure over the full unitary group on $\mathcal{H}_A$.

3.2. Technical ingredients to the proof. The proof of the non-smooth decoupling theorem (Theorem 3.3) is based on a few technical lemmas, which we state and prove in the following, and which may be of independent interest. We note that they partly generalize techniques developed in the context of privacy amplification [RK05, Ren05, TRSS10] as well as earlier work on decoupling (see, e.g., [HOW07]).

Lemma 3.4. *Let $M, N \in \mathcal{L}(\mathcal{H}_A)$. Then, we have that* $\mathrm{Tr}[(M \otimes N)F] = \mathrm{Tr}[MN]$*, where F swaps the two copies of the A subsystem.*

Proof. Write M and N in the standard basis for $\mathcal{H}_A$, that is, $M = \sum_{ij} m_{ij}|i\rangle\langle j|$ and $N = \sum_{kl} n_{kl}|k\rangle\langle l|$. Then, we have that

$$\begin{aligned}
\mathrm{Tr}[(M \otimes N)F] &= \mathrm{Tr}\left[\left(\sum_{ijkl} m_{ij} n_{kl} |i\rangle\langle j| \otimes |k\rangle\langle l|\right) F\right] \\
&= \mathrm{Tr}\left[\sum_{ijkl} m_{ij} n_{kl} |i\rangle\langle l| \otimes |k\rangle\langle j|\right] \\
&= \sum_{ij} m_{ij} n_{ji} \\
&= \mathrm{Tr}[MN].
\end{aligned} \tag{31}$$

□

The second lemma involves averaging over Haar distributed unitaries. While it would take us too far afield to formally introduce the Haar measure, it can simply be thought of as the uniform probability distribution over the set of all unitaries on a Hilbert space. The following then tells us the expected value of $U^{\otimes 2} M (U^\dagger)^{\otimes 2}$ with $M \in \mathcal{L}(\mathcal{H}_A^{\otimes 2})$ when U is selected "uniformly at random".

Lemma 3.5. *Let $M \in \mathcal{L}(\mathcal{H}_A^{\otimes 2})$. Then, we have that*

$$\mathbb{E}(M) = \int_{\mathbb{U}(A)} U^{\otimes 2} M (U^\dagger)^{\otimes 2} dU = \alpha \cdot \mathbb{1}_{AA'} + \beta \cdot F_A, \tag{32}$$

where F_A swaps the two copies of the A subsystem, α and β are such that $\mathrm{Tr}[M] = \alpha|A|^2 + \beta|A|$ *and* $\mathrm{Tr}[MF] = \alpha|A| + \beta|A|^2$*, and dU is the normalized Haar measure on $\mathbb{U}(A)$.*

Proof. This follows directly from a standard result in Schur-Weyl duality, e.g., [CS06, Proposition 2.2]. The latter states that $\mathbb{E} : \mathcal{L}(\mathcal{H}_A^{\otimes 2}) \to \mathcal{L}(\mathcal{H}_A^{\otimes 2})$ is an orthogonal projection onto $\mathrm{span}\{\mathbb{1}, F\}$ under the inner product $\langle A, B\rangle = \mathrm{Tr}[A^\dagger B]$. Hence, $\mathbb{E}(M)$ can be written as $\alpha \cdot \mathbb{1}_{AA'} + \beta \cdot F_A$ as claimed, and the conditions $\mathrm{Tr}[\mathbb{1}\mathbb{E}(M)] = \mathrm{Tr}[M]$ and $\mathrm{Tr}[F\mathbb{E}(M)] = \mathrm{Tr}[FM]$ must be fulfilled, and these lead to the two conditions on α and β. □

The following bounds the ratio of the purity of a bipartite state and the purity of the reduced state on one subsystem.

Lemma 3.6. *Let $\xi_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$. Then, we have that*

$$\frac{1}{|A|} \leqslant \frac{\mathrm{Tr}\left[{\xi_{AB}}^2\right]}{\mathrm{Tr}\left[{\xi_B}^2\right]} \leqslant |A|. \tag{33}$$

Proof. Letting A' be a system isomorphic to A, we first prove the left-hand side

$$\begin{aligned}
\mathrm{Tr}\left[{\xi_B}^2\right] &= \mathrm{Tr}\left[\mathrm{Tr}_A\left[\xi_{AB}\right]^2\right] \\
&= \mathrm{Tr}\left[\mathrm{Tr}_A\left[\xi_{AB}\right] \cdot \mathrm{Tr}_{A'}\left[\xi_{A'B}\right]\right] \\
&= \mathrm{Tr}\left[\xi_{AB}\left(\mathrm{Tr}_{A'}\left[\xi_{A'B}\right] \otimes \mathbb{1}_A\right)\right] \\
&= \mathrm{Tr}\left[(\xi_{AB} \otimes \mathbb{1}_{A'})(\xi_{A'B} \otimes \mathbb{1}_A)\right] \\
&\leqslant \sqrt{\mathrm{Tr}\left[(\xi_{AB} \otimes \mathbb{1}_{A'})^2\right] \cdot \mathrm{Tr}\left[(\xi_{A'B} \otimes \mathbb{1}_A)^2\right]} \\
&= \mathrm{Tr}\left[{\xi_{AB}}^2 \otimes \mathbb{1}_{A'}\right] \\
&= |A| \cdot \mathrm{Tr}\left[{\xi_{AB}}^2\right],
\end{aligned} \tag{34}$$

where the inequality is due to an application of Cauchy–Schwarz. The right-hand side follows from the fact that $\xi_{AB} \leqslant |A| \cdot \mathbb{1}_A \otimes \xi_B$. This can in turn be seen from the fact that we can write

$$|A| \cdot \mathbb{1}_A \otimes \xi_B = \sum_{i=1}^{|A|^2} U_A^i \xi_{AB} (U_A^i)^\dagger, \tag{35}$$

with unitaries U_A^i such that $\mathrm{Tr}\left[(U_A^i)^\dagger U_A^j\right] = 0$ for every $i \neq j$, and $U_A^1 = \mathbb{1}_A$. □

In the main proof, we will need to bound the trace distance between two states. The following lemma will allow us to do this.

Lemma 3.7. *Let $M \in \mathcal{L}(\mathcal{H}_A)$ and $\sigma \in \mathcal{P}(\mathcal{H}_A)$. Then, we have that*

$$\|M\|_1 \leqslant \sqrt{\mathrm{Tr}[\sigma] \cdot \mathrm{Tr}[\sigma^{-1/4} M \sigma^{-1/2} M^\dagger \sigma^{-1/4}]}. \tag{36}$$

In particular, if M is Hermitian then, we have that

$$\|M\|_1 \leqslant \sqrt{\mathrm{Tr}[\sigma] \cdot \mathrm{Tr}[(\sigma^{-1/4} M \sigma^{-1/4})^2]}. \tag{37}$$

This is a slight generalization of [Ren05, Lemma 5.1.3]. For completeness we give a different proof here.

Proof. We calculate

$$\begin{aligned}\|M\|_1 &= \max_U |\mathrm{Tr}[UM]| \\ &= \max_U \left|\mathrm{Tr}\left[(\sigma^{1/4}U\sigma^{1/4})(\sigma^{-1/4}M\sigma^{-1/4})\right]\right| \\ &\leqslant \max_U \sqrt{\mathrm{Tr}\left[(\sigma^{1/4}U\sigma^{1/4})(\sigma^{1/4}U^\dagger\sigma^{1/4})\right]\cdot \mathrm{Tr}\left[\sigma^{-1/4}M\sigma^{-1/2}M^\dagger\sigma^{-1/4}\right]} \\ &= \sqrt{\max_U \mathrm{Tr}[\sigma^{1/2}U\sigma^{1/2}U^\dagger]\cdot \mathrm{Tr}\left[\sigma^{-1/4}M\sigma^{-1/2}M^\dagger\sigma^{-1/4}\right]} \\ &= \sqrt{\mathrm{Tr}[\sigma]\cdot \mathrm{Tr}\left[\sigma^{-1/4}M\sigma^{-1/2}M^\dagger\sigma^{-1/4}\right]}, \end{aligned} \tag{38}$$

where the inequality results from an application of Cauchy–Schwarz, and the maximizations are over all unitaries on A. The last equality follows from

$$\begin{aligned}\max_U \mathrm{Tr}\left[\sigma^{1/2}U\sigma^{1/2}U^\dagger\right] &\leqslant \max_U \sqrt{\mathrm{Tr}\,[\sigma]\cdot \mathrm{Tr}\left[U\sigma^{1/2}U^\dagger U\sigma^{1/2}U^\dagger\right]} \\ &= \mathrm{Tr}[\sigma] \\ &\leqslant \max_U \mathrm{Tr}[\sigma^{1/2}U\sigma^{1/2}U^\dagger]. \end{aligned} \tag{39}$$

□

3.3. Proof of the non-smooth decoupling theorem (Theorem 3.3). Throughout the proof, we will denote with a prime the "twin" subsystems used when we take tensor copies of operators, and F_S denotes a swap between S and S'.

We first use Lemma 3.7 to bound the trace norm. For $\sigma_B \in \mathcal{S}_=(\mathcal{H}_B)$ and $\zeta_E \in \mathcal{S}_=(\mathcal{H}_E)$ we get

$$\begin{aligned}&\|\mathcal{T}_{A\to B}(U_A\rho_{AE}U_A^\dagger) - \tau_B\otimes\rho_E\|_1 \\ &\leqslant \sqrt{\mathrm{Tr}\left[\left((\sigma_B\otimes\zeta_E)^{-1/4}(\mathcal{T}_{A\to B}(U_A\rho_{AE}U_A^\dagger) - \tau_B\otimes\rho_E)(\sigma_B\otimes\zeta_E)^{-1/4}\right)^2\right]}. \end{aligned} \tag{40}$$

Now define the CPM $\tilde{\mathcal{T}}_{A\to B}(\cdot) = \sigma_B^{-1/4}\mathcal{T}_{A\to B}(\cdot)\sigma_B^{-1/4}$ and the operators $\tilde{\tau}_{A'B} = J(\tilde{\mathcal{T}})$ and $\tilde{\rho}_{AE} = \zeta_E^{-1/4}\rho_{AE}\zeta_E^{-1/4}$. We then rewrite the above as

$$\left\|\mathcal{T}_{A\to B}(U_A\rho_{AE}U_A^\dagger) - \tau_B\otimes\rho_E\right\|_1 \leqslant \sqrt{\mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger) - \tilde{\tau}_B\otimes\tilde{\rho}_E\right)^2\right]}. \tag{41}$$

Using Jensen's inequality we obtain

$$\begin{aligned}&\int \left\|\mathcal{T}_{A\to B}(U_A\rho_{AE}U_A^\dagger) - \tau_B\otimes\rho_E\right\|_1 dU \\ &\leqslant \sqrt{\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger) - \tilde{\tau}_B\otimes\tilde{\rho}_E\right)^2\right] dU}. \end{aligned} \tag{42}$$

We now simplify the integral

$$\begin{aligned}
&\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger)-\tilde{\tau}_B\otimes\tilde{\rho}_E\right)^2\right]dU\\
&=\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger)\right)^2\right]dU\\
&\quad-2\int \mathrm{Tr}\left[\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger)\,(\tilde{\tau}_B\otimes\tilde{\rho}_E)\right]dU+\mathrm{Tr}\left[(\tilde{\tau}_B\otimes\tilde{\rho}_E)^2\right]\\
&=\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger)\right)^2\right]dU\\
&\quad-2\,\mathrm{Tr}\left[\tilde{\mathcal{T}}_{A\to B}\left(\int U_A\tilde{\rho}_{AE}U_A^\dagger dU\right)(\tilde{\tau}_B\otimes\tilde{\rho}_E)\right]+\mathrm{Tr}\left[(\tilde{\tau}_B\otimes\tilde{\rho}_E)^2\right]\\
&=\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger)\right)^2\right]dU-\mathrm{Tr}\left[\tilde{\tau}_B^2\right]\cdot\mathrm{Tr}\left[\tilde{\rho}_E^2\right].
\end{aligned}\tag{43}$$

We rewrite the first term as follows

$$\begin{aligned}
\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger)\right)^2\right]dU&=\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}(U_A\tilde{\rho}_{AE}U_A^\dagger)\right)^{\otimes 2}F_{BE}\right]dU\\
&=\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}_{A\to B}^{\otimes 2}\left(U_A^{\otimes 2}\tilde{\rho}_{AE}^{\otimes 2}(U_A^\dagger)^{\otimes 2}\right)\right)F_{BE}\right]dU\\
&=\int \mathrm{Tr}\left[\tilde{\rho}_{AE}^{\otimes 2}\left(\left((U_A^\dagger)^{\otimes 2}(\tilde{\mathcal{T}}_{B\to A}^\dagger)^{\otimes 2}(F_B)U_A^{\otimes 2}\right)\otimes F_E\right)\right]dU\\
&=\mathrm{Tr}\left[\tilde{\rho}_{AE}^{\otimes 2}\left(\left(\int (U_A^\dagger)^{\otimes 2}(\tilde{\mathcal{T}}_{B\to A}^\dagger)^{\otimes 2}(F_B)U_A^{\otimes 2}dU\right)\otimes F_E\right)\right],
\end{aligned}\tag{44}$$

where we have used the swap trick (Lemma 3.4) with $F_{BE}=F_B\otimes F_E$ in the first equality, the definition of the adjoint of a superoperator in the third equality and the linearity of the trace in forth equality. We now compute the integral using a lemma about Haar distributed unitaries (Lemma 3.5)

$$\int (U_A^\dagger)^{\otimes 2}(\tilde{\mathcal{T}}_{B\to A}^\dagger)^{\otimes 2}(F_B)U_A^{\otimes 2}dU=\alpha\cdot\mathbb{1}_{AA'}+\beta\cdot F_A,\tag{45}$$

where α and β satisfy the following equations

$$\begin{aligned}
\alpha|A|^2+\beta|A|=\mathrm{Tr}\left[(\tilde{\mathcal{T}}_{B\to A}^\dagger)^{\otimes 2}(F_B)\right]=\mathrm{Tr}\left[F_B\tilde{\mathcal{T}}_{A\to B}^{\otimes 2}(\mathbb{1}_{AA'})\right]&=|A|^2\cdot\mathrm{Tr}\left[F_B\tilde{\tau}_B^{\otimes 2}\right]\\
&=|A|^2\cdot\mathrm{Tr}\left[\tilde{\tau}_B^2\right]
\end{aligned}\tag{46}$$

and

$$\begin{aligned}
\alpha|A|+\beta|A|^2&=\mathrm{Tr}\left[(\tilde{\mathcal{T}}_{B\to A}^\dagger)^{\otimes 2}(F_B)F_A\right]\\
&=\mathrm{Tr}\left[F_B\tilde{\mathcal{T}}_{A\to B}^{\otimes 2}(F_A)\right]\\
&=|A|^2\cdot\mathrm{Tr}\left[F_B\cdot\mathrm{Tr}_{AA'}\left[\tilde{\tau}_{AB}^{\otimes 2}(F_A\otimes\mathbb{1}_{BB'})\right]\right]
\end{aligned}$$

$$
\begin{aligned}
&= |A|^2 \cdot \operatorname{Tr}\left[(\mathbb{1}_{AA'} \otimes F_B)\tilde{\tau}_{AB}^{\otimes 2}(F_A \otimes \mathbb{1}_{BB'})\right] \\
&= |A|^2 \cdot \operatorname{Tr}\left[F_{AB}\tilde{\tau}_{AB}^{\otimes 2}\right] \\
&= |A|^2 \cdot \operatorname{Tr}\left[\tilde{\tau}_{AB}^2\right]. \qquad (47)
\end{aligned}
$$

In the third equality, we have used the fact that $\tilde{\tau}_{AB}$ is a Choi–Jamiołkowski representation of $\tilde{\mathcal{T}}$ (Lemma 2.1), and the fourth equality is due to the fact that the adjoint of the partial trace is tensoring with the identity. Solving this system of equations yields

$$
\alpha = \operatorname{Tr}\left[\tilde{\tau}_B^2\right] \cdot \left(\frac{|A|^2 - \frac{|A| \cdot \operatorname{Tr}\left[\tilde{\tau}_{AB}^2\right]}{\operatorname{Tr}\left[\tilde{\tau}_B^2\right]}}{|A|^2 - 1} \right) \qquad (48)
$$

$$
\beta = \operatorname{Tr}\left[\tilde{\tau}_{AB}^2\right] \cdot \left(\frac{|A|^2 - \frac{|A| \cdot \operatorname{Tr}\left[\tilde{\tau}_B^2\right]}{\operatorname{Tr}\left[\tilde{\tau}_{AB}^2\right]}}{|A|^2 - 1} \right). \qquad (49)
$$

By applying Lemma 3.6, we can simplify this to $\alpha \leqslant \operatorname{Tr}\left[\tilde{\tau}_B^2\right]$ and $\beta \leqslant \operatorname{Tr}\left[\tilde{\tau}_{AB}^2\right]$. Substituting this into (44) and using the swap trick twice (Lemma 3.4), and then substituting into (42) yields

$$
\int \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 dU \leqslant \sqrt{\operatorname{Tr}\left[\tilde{\tau}_{AB}^2\right] \cdot \operatorname{Tr}\left[\tilde{\rho}_{AE}^2\right]}. \qquad (50)
$$

Finally we get the theorem by using the definitions of $\tilde{\tau}_{AB}$, $\tilde{\rho}_{AE}$ and the definition of the conditional collision entropy (Definition 2.9). □

3.4. Proof of the main decoupling theorem (Theorem 3.1). We now prove our main result, which is obtained from the non-smooth decoupling theorem (Theorem 3.3) by replacing the conditional collision entropies by smooth conditional min-entropies.

First, note that the conditional collision entropy is always greater or equal to the conditional min-entropy (Lemma A.1) and therefore we are allowed to replace the H_2 terms on the right-hand side of the statement of Theorem 3.3 by $H_{\min}$ terms. Thus we only have to consider the smoothing.

Let $\widehat{\rho}^{AE} \in \mathcal{B}^\varepsilon(\rho_{AE})$ be such that $H_{\min}^\varepsilon(A|E)_\rho = H_{\min}(A|E)_{\widehat{\rho}}$ and $\widehat{\tau}_{AB} \in \mathcal{B}^\varepsilon(\tau_{AB})$ be such that $H_{\min}^\varepsilon(A|B)_\tau = H_{\min}(A|B)_{\widehat{\tau}}$.

Furthermore write $\widehat{\tau}_{AB} - \tau_{AB} = \Delta_{AB}^+ - \Delta_{AB}^-$, where $\Delta_{AB}^\pm \in \mathcal{P}(\mathcal{H}_{AB})$ have orthogonal support, and likewise, $\widehat{\rho}_{AE} - \rho_{AE} = \delta_{AE}^+ - \delta_{AE}^-$ with δ_{AE}^+ and δ_{AE}^- having orthogonal support as well as $\delta_{AE}^\pm \in \mathcal{P}(\mathcal{H}_{AE})$. By the equivalence of purified distance and trace distance (Lemma B.1) we have $\|\widehat{\tau}_{AB} - \tau_{AB}\|_1 \leqslant 2\varepsilon$ and hence $\left\|\Delta_{AB}^\pm\right\|_1 \leqslant 2\varepsilon$.

Moreover define $\widehat{\mathcal{T}}_{A\to B}$, $\mathcal{D}_{A\to B}^-$ and $\mathcal{D}_{A\to B}^+$ as the unique superoperators that are such that $\widehat{\tau}_{AB} = J(\widehat{\mathcal{T}}_{A\to B})$, $\Delta_{AB}^- = J(\mathcal{D}_{A\to B}^-)$ and $\Delta_{AB}^+ = J(\mathcal{D}_{A\to B}^+)$, respectively.

Using the non-smooth decoupling theorem (Theorem 3.3) we get

$$
\begin{aligned}
2^{-\frac{1}{2}H_{\min}^\varepsilon(A|B)_\tau - \frac{1}{2}H_{\min}^\varepsilon(A|E)_\rho} &\geqslant \int \left\| \widehat{\mathcal{T}}_{A\to B}(U_A \widehat{\rho}_{AE} U_A^\dagger) - \widehat{\tau}_B \otimes \widehat{\rho}_E \right\|_1 dU \\
&\geqslant \int \left\| \widehat{\mathcal{T}}_{A\to B}(U_A \widehat{\rho}_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 dU - 4\varepsilon
\end{aligned}
$$

$$\begin{aligned}
&\geqslant \int \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 dU \\
&\quad - \int \left\| \widehat{\mathcal{T}}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \widehat{\mathcal{T}}_{A\to B}(U_A \widehat{\rho}_{AE} U_A^\dagger) \right\|_1 dU \\
&\quad - \int \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \widehat{\mathcal{T}}_{A\to B}(U_A \rho_{AE} U_A^\dagger) \right\|_1 dU - 4\varepsilon,
\end{aligned} \tag{51}$$

where we have used the triangle inequality for the trace distance in the second inequality. We now deal with the second term above

$$\begin{aligned}
&\int \left\| \widehat{\mathcal{T}}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \widehat{\mathcal{T}}_{A\to B}(U_A \widehat{\rho}_{AE} U_A^\dagger) \right\|_1 dU \\
&= \int \left\| \widehat{\mathcal{T}}_{A\to B}(U_A (\delta^+_{AE} - \delta^-_{AE}) U_A^\dagger) \right\|_1 dU \\
&\leqslant \int \left\| \widehat{\mathcal{T}}_{A\to B}(U_A \delta^+_{AE} U_A^\dagger) \right\|_1 dU + \int \left\| \widehat{\mathcal{T}}_{A\to B}(U_A \delta^-_{AE} U_A^\dagger) \right\|_1 dU \\
&= \int \mathrm{Tr}\left[\widehat{\mathcal{T}}_{A\to B}(U_A \delta^+_{AE} U_A^\dagger) \right] dU + \int \mathrm{Tr}\left[\widehat{\mathcal{T}}_{A\to B}(U_A \delta^-_{AE} U_A^\dagger) \right] dU \\
&= \mathrm{Tr}\left[\widehat{\mathcal{T}}_{A\to B}\left(\frac{\mathbb{1}_A}{|A|} \right) \right] \cdot \left(\mathrm{Tr}\left[\delta^+_{AE} \right] + \mathrm{Tr}\left[\delta^-_{AE} \right] \right) \\
&\leqslant 4\varepsilon.
\end{aligned} \tag{52}$$

We deal with the third term in a similar fashion

$$\begin{aligned}
&\int \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \widehat{\mathcal{T}}_{A\to B}(U_A \rho_{AE} U_A^\dagger) \right\|_1 dU \\
&= \int \left\| (\mathcal{D}^+_{A\to B} - \mathcal{D}^-_{A\to B})(U_A \rho_{AE} U_A^\dagger) \right\|_1 dU \\
&\leqslant \int \left\| \mathcal{D}^+_{A\to B}(U_A \rho_{AE} U_A^\dagger) \right\|_1 dU + \int \left\| \mathcal{D}^-_{A\to B}(U_A \rho_{AE} U_A^\dagger) \right\|_1 dU \\
&= \int \mathrm{Tr}\left[\mathcal{D}^+_{A\to B}(U_A \rho_{AE} U_A^\dagger) \right] dU + \int \mathrm{Tr}\left[\mathcal{D}^-_{A\to B}(U_A \rho_{AE} U_A^\dagger) \right] dU \\
&= \mathrm{Tr}\left[\mathcal{D}^+_{A\to B}\left(\frac{\mathbb{1}_A}{|A|} \otimes \rho_E \right) \right] + \mathrm{Tr}\left[\mathcal{D}^-_{A\to B}\left(\frac{\mathbb{1}_A}{|A|} \otimes \rho_E \right) \right] \\
&= \mathrm{Tr}\left[\Delta^+_A \otimes \rho_E \right] + \mathrm{Tr}\left[\Delta^-_A \otimes \rho_E \right] \\
&\leqslant 4\varepsilon.
\end{aligned} \tag{53}$$

This results in

$$\begin{aligned}
&\int \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 dU \\
&\leqslant 2^{-\frac{1}{2} H^\varepsilon_{\min}(A|E)_\rho - \frac{1}{2} H^\varepsilon_{\min}(A|B)_\tau} + 12\varepsilon.
\end{aligned} \tag{54}$$

□

4. Converse

The main purpose of this section is to state and prove a theorem (Theorem 4.1) which implies that the achievability result of the previous section (Theorem 3.1) is essentially optimal for many natural choices of the mapping $\mathcal{T}$.

4.1. Statement of the converse theorem. According to Theorem 3.1, decoupling is achieved whenever the term $H_{\min}^{\varepsilon}(A|E)_{\rho} + H_{\min}^{\varepsilon}(A|B)_{\tau}$ is sufficiently larger than 0. Our converse now says that this is also a necessary condition (up to additive terms of the order $\log(1/\varepsilon)$ and the scaling of the smoothing parameter) if one replaces the smooth conditional min-entropy in the second term, $H_{\min}^{\varepsilon}(A|B)_{\tau}$ (which characterizes the channel), by a smooth conditional max-entropy.

Theorem 4.1 (Decoupling Converse). *Let $\rho_{AE} \in \mathcal{S}_{=}(\mathcal{H}_{AE})$, $\mathcal{T}_{A\to B}$ be a TPCPM, and suppose that*

$$\|\mathcal{T}_{A\to B}(\rho_{AE}) - \mathcal{T}_{A\to B}(\rho_A) \otimes \rho_E\|_1 \leqslant \varepsilon. \tag{55}$$

Then, we have for any $\varepsilon', \varepsilon'' > 0$ that

$$H_{\min}^{2\sqrt{6\varepsilon''+2\varepsilon}+2\sqrt{\varepsilon'}+\varepsilon''}(A|E)_{\rho} + H_{\max}^{\epsilon''}(A|B)_{\omega} \geqslant -\log\frac{1}{\varepsilon'}, \tag{56}$$

where $\omega_{AB} = \mathcal{T}_{A'\to B}(\rho_{AA'})$ with $\rho_{AA'} \in \mathcal{S}_{=}(\mathcal{H}_{AA'})$ a purification of ρ_A, and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.

Note that we could also write $\omega_{AB} = |A|\left(\sqrt{\rho_A}\right)^{\mathsf{T}} J(\mathcal{T}) \left(\sqrt{\rho_A}\right)^{\mathsf{T}}$. In our formulation of the converse theorem, the mapping $\mathcal{T}$ is not necessarily prepended by a unitary and the state that appears in the entropy term of the TPCPM is given by the more general expression $\omega_{AB} = \mathcal{T}_{A'\to B}(\rho_{AA'})$ (rather than $\tau_{AB} = J(\mathcal{T})$ as in Theorem 3.1, corresponding to the case where ρ_A is fully mixed). However, if we apply the converse to a TPCPM of the form $\bar{\mathcal{T}} = \mathcal{T} \circ \mathcal{U}$, where $\mathcal{U}$ corresponds to a random unitary channel applied to the input, Theorem 4.1 simplifies to the following.

Corollary 4.2. *For the same premises as in Theorem 4.1, but applied to the TPCPM $\bar{\mathcal{T}}_{A\to B} = \mathcal{T}_{A\to B} \circ \mathcal{U}_A$, where $\mathcal{U}_A$ corresponds to a Haar random unitary channel applied to the input, we have that*

$$H_{\min}^{2\sqrt{6\varepsilon''+2\varepsilon}+2\sqrt{\varepsilon'}+\varepsilon''}(A|E)_{\rho} + H_{\max}^{\epsilon''}(A|B)_{\tau} \geqslant -\log\frac{1}{\varepsilon'}, \tag{57}$$

where $\tau_{AB} = J(\mathcal{T})$.

Proof. By assumption we have

$$\int_{\mathbb{U}(A)} \left\| \mathcal{T}_{A\to B}(U_A \rho_{AE} U_A^{\dagger}) - \mathcal{T}_{A\to B}(U_A \rho_A U_A^{\dagger}) \otimes \rho_E \right\|_1 dU \leqslant \varepsilon. \tag{58}$$

and since the unitary U_A is chosen at random, this is equivalent to

$$\|\mathcal{T}_{A\to B} \circ \mathcal{F}_{A\to AU}(\rho_{AE}) - \mathcal{T}_{A\to B} \circ \mathcal{F}_{A\to AU}(\rho_A) \otimes \rho_E\|_1 \leqslant \varepsilon, \tag{59}$$

where $\mathcal{F}_{A\to AU}$ denotes the TPCPM that chooses at random a unitary U_A and outputs the choice of U_A. Now, let σ_{AUER} be a purification of $\sigma_{AUE} = \mathcal{F}_{A\to AU}(\rho_{AE})$ and note

that $\sigma_A = \frac{\mathbb{1}_A}{|A|}$ as well as $\sigma_E = \rho_E$. We apply Theorem 4.1 to (59) with the map $\mathcal{T}_{A\to B}$ and the state σ_{AUE} to get

$$H_{\min}^{\delta}(A|UE)_\sigma + H_{\max}^{\epsilon''}(UER|B)_{\mathcal{T}(\sigma)} \geqslant -\log\frac{1}{\varepsilon'}, \tag{60}$$

for $\delta = 2\sqrt{6\varepsilon'' + 2\varepsilon} + 2\sqrt{\varepsilon'} + \varepsilon''$. Since the state σ_{AUER} and the maximally entangled state $|\Phi\rangle\langle\Phi|_{AA'}$ are both purifications of $\frac{\mathbb{1}_A}{|A|}$, there exists by Uhlmann's theorem [Uhl76] an isometry $\mathcal{W}_{UER\to A'}$ such that $|\Phi\rangle\langle\Phi|_{AA'} = \mathcal{W}_{UER\to A'}(\sigma_{AUER})$. Hence, we have that $\mathcal{T}_{A\to B}(|\Phi\rangle\langle\Phi|_{AA'}) = \mathcal{W}_{UER\to A'} \circ \mathcal{T}_{A\to B}(\sigma_{AUER})$, and by the invariance of the smooth conditional max-entropy under local isometries (Lemma 2.6) we get

$$H_{\max}^{\epsilon''}(UER|B)_{\mathcal{T}(\sigma)} = H_{\max}^{\epsilon''}(A'|B)_{\mathcal{T}(|\Phi\rangle\langle\Phi|)} = H_{\max}^{\epsilon''}(A|B)_\tau. \tag{61}$$

Finally, we show that $H_{\min}^{\delta}(A|UE)_\sigma$ in (60) is upper bounded by $H_{\min}^{\delta}(A|E)_\rho$. Since the register U in σ_{AUE} is classical, we can copy U to another register U' resulting in the state $\sigma_{AUU'E}$. With Lemma A.7 we then have

$$H_{\min}^{\delta}(A|UE)_\sigma = H_{\min}^{\delta}(AU'|UE)_\sigma. \tag{62}$$

But now there exists an isometry $\mathcal{V}_{AU'\to A}$ that reverses the action of the TPCPM $\mathcal{F}$ such that $\mathcal{V}_{AU'\to A}(\sigma_{AU'E}) = \rho_{AE}$ (we let $\mathcal{V}$ act on the copy U' instead of U). Using the data processing inequality for the smooth conditional min-entropy (Lemma 2.7) and the invariance of the smooth conditional min-entropy under local isometries (Lemma 2.6), we conclude

$$H_{\min}^{\delta}(AU'|UE)_\sigma \leqslant H_{\min}^{\delta}(AU'|E)_\sigma = H_{\min}^{\delta}(A|E)_\rho. \tag{63}$$

□

It can also be verified that the two terms, $H_{\min}^{\varepsilon}(A|B)_\tau$ (from the achievability in Theorem 3.1) and $H_{\max}^{\varepsilon}(A|B)_\tau$ (from the converse in Corollary 4.2), coincide whenever the relevant states are essentially flat (i.e., proportional to projectors). This is the case for many channels used in applications (e.g., for state merging, cf. Sect. 5). Examples of such channels are given in Table 2. Furthermore, as we shall explain in the discussion section (Sect. 6), the two terms coincide asymptotically for iid channels.

4.2. Proof of the converse theorem (Theorem 4.1). Let ρ_{AER} be a purification of ρ_{AE}, $\mathcal{W}_{A\to BB'}$ a Stinespring dilation [Sti55] of $\mathcal{T}_{A\to B}$ and define

$$\tilde{\sigma}_{BB'ER} = |\tilde{\sigma}\rangle\langle\tilde{\sigma}|_{BB'ER} = \mathcal{W}_{A\to BB'}(\rho_{AER}). \tag{64}$$

We have by Uhlmann's theorem [Uhl76] that ω_{AB} and $\tilde{\sigma}_{BER}$ are related by an isometry $\mathcal{V}_{A\to ER}$, and hence by the invariance of the smooth conditional max-entropy under local isometries (Lemma 2.6) that

$$H_{\max}^{\epsilon''}(A|B)_\omega = H_{\max}^{\epsilon''}(ER|B)_{\tilde{\sigma}}. \tag{65}$$

Furthermore, let $\sigma_{BB'ER} = |\sigma\rangle\langle\sigma|_{BB'ER}$ be a subnormalized state with $P(\sigma, \tilde{\sigma}) \leqslant \epsilon''$ such that $H_{\max}(ER|B)_\sigma = H_{\max}^{\epsilon''}(A|B)_\omega$, as well as $\sigma_{BB'ER} = |\bar{\sigma}\rangle\langle\bar{\sigma}|_{BB'ER}$ such that $\bar{\sigma}_{BE} = \sigma_B \otimes \sigma_E$ and

$$F(\sigma_{BB'ER}, \bar{\sigma}_{BB'ER}) = F(\sigma_{BE}, \sigma_B \otimes \sigma_E). \tag{66}$$

Such a state exists by Uhlmann's theorem [Uhl76], and can be shown to satisfy $P(\bar{\sigma}, \sigma) \leqslant \sqrt{6\varepsilon'' + 2\varepsilon}$. The latter bound is obtained from

$$\begin{aligned}\|\sigma_{BE} - \bar{\sigma}_{BE}\|_1 &\leqslant \|\sigma_{BE} - \tilde{\sigma}_{BE}\|_1 + \|\tilde{\sigma}_{BE} - \bar{\sigma}_{BE}\|_1 \\ &\leqslant \|\sigma_{BE} - \tilde{\sigma}_{BE}\|_1 + \|\tilde{\sigma}_{BE} - \tilde{\sigma}_B \otimes \tilde{\sigma}_E\|_1 + \|\tilde{\sigma}_B \otimes \tilde{\sigma}_E - \sigma_B \otimes \sigma_E\|_1 \\ &\leqslant \varepsilon'' + \varepsilon + \|\tilde{\sigma}_B \otimes \tilde{\sigma}_E - \tilde{\sigma}_B \otimes \sigma_E\|_1 + \|\tilde{\sigma}_B \otimes \sigma_E - \sigma_B \otimes \sigma_E\|_1 \\ &\leqslant 3\varepsilon'' + \varepsilon, \end{aligned} \tag{67}$$

combined with the equivalence of purified distance and trace distance (Lemma B.1). Now, we know from a technical lemma about the conditional max-entropy (Lemma B.2) that

$$\sigma_{BB'ER} \leqslant 2^{H_{\max}(ER|B)_{\sigma|\sigma}} \cdot Y_{EBR} \otimes \mathbb{1}_{B'}, \tag{68}$$

where

$$Y_{BER} = 2^{-\frac{1}{2}H_{\max}(ER|B)_{\sigma|\sigma}} \cdot \sigma_B^{-1/2} \sqrt{\sigma_B^{1/2} \sigma_{BER} \sigma_B^{1/2}}\, \sigma_B^{-1/2}. \tag{69}$$

This implies that

$$\sigma_{BB'ER} \leqslant \frac{2^{H_{\max}(ER|B)_{\sigma|\sigma}}}{\varepsilon'} \cdot \left((1-\varepsilon') \cdot \sigma_B^{-1/2} \bar{\sigma}_{BER} \sigma_B^{-1/2} + \varepsilon' \cdot Y_{BER} \right) \otimes \mathbb{1}_{B'} \tag{70}$$

for any $\varepsilon' > 0$. Tracing out the R system, we get

$$\sigma_{BEB'} \leqslant \frac{2^{H_{\max}(ER|B)_{\sigma|\sigma}}}{\varepsilon'} \cdot \left((1-\varepsilon') \cdot \mathbb{1}_B \otimes \sigma_E + \varepsilon' \cdot Y_{BE} \right) \otimes \mathbb{1}_{B'}. \tag{71}$$

We now define $G_{BE} = \sqrt{1-\varepsilon'} \cdot \sigma_E^{1/2} ((1-\varepsilon') \cdot \mathbb{1}_B \otimes \sigma_E + \varepsilon' \cdot Y_{BE})^{-1/2}$. Note that G is a contraction, i.e., $\|G\|_\infty \leqslant 1$,

$$\begin{aligned} GG^\dagger &= (1-\varepsilon') \cdot \sigma_E^{1/2} \left((1-\varepsilon') \cdot \mathbb{1}_B \otimes \sigma_E + \varepsilon' \cdot Y_{BE} \right)^{-1} \sigma_E^{1/2} \\ &\leqslant (1-\varepsilon') \cdot \sigma_E^{1/2} \left((1-\varepsilon') \cdot \mathbb{1}_B \otimes \sigma_E \right)^{-1} \sigma_E^{1/2} \\ &= \mathbb{1}_{BE}, \end{aligned} \tag{72}$$

where we have used the operator monotonicity of $f(t) = -1/t$. At this point, we conjugate both sides of (70) by G_{BE} to get

$$\begin{aligned} G_{BE} \sigma_{BEB'} G_{BE}^\dagger &\leqslant \frac{(1-\varepsilon') \cdot 2^{H_{\max}(ER|B)_{\sigma|\sigma}}}{\varepsilon'} \cdot \sigma_E \otimes \mathbb{1}_{BB'} \\ &\leqslant \frac{2^{H_{\max}(ER|B)_{\sigma|\sigma}}}{\varepsilon'} \cdot \sigma_E \otimes \mathbb{1}_{BB'}. \end{aligned} \tag{73}$$

Let us now define $|\psi\rangle_{BERB'} = G_{BE}|\sigma\rangle_{BERB'}$ and note that $\psi_{BERB'} = |\psi\rangle\langle\psi|_{BERB'}$ is a subnormalized state since G is a contraction. Then, we can rewrite (73) as

$$\psi_{BEB'} \leqslant \frac{2^{H_{\max}(ER|B)_{\sigma|\sigma}}}{\varepsilon'} \cdot \sigma_E \otimes \mathbb{1}_{BB'}, \tag{74}$$

which implies

$$H_{\min}(BB'|E)_{\psi|\sigma} \geqslant -H_{\max}(ER|B)_{\sigma|\sigma} - \log(1/\varepsilon'). \tag{75}$$

We will now need to show that $\psi_{BEB'}$ is $(2\sqrt{6\varepsilon'' + 2\varepsilon} + 2\sqrt{\varepsilon'} + \varepsilon'')$-close to $\tilde{\sigma}_{BEB'}$, because the invariance of the smooth conditional min-entropy under local isometries (Lemma 2.6) then implies the claim

$$H_{\min}^{2\sqrt{6\varepsilon''+2\varepsilon}+2\sqrt{\varepsilon'}+\varepsilon''}(A|E)_\rho + H_{\max}^{\varepsilon''}(A|B)_\omega \geqslant -\log(1/\varepsilon'). \tag{76}$$

To this end, we shall define the following vectors

$$|\psi'\rangle_{BERB'} = G_{BE}^\dagger |\bar{\sigma}\rangle_{BERB'} \tag{77}$$

$$|\psi''\rangle_{BERB'} = G_{BE} |\bar{\sigma}\rangle_{BERB'} \tag{78}$$

$$|\tilde{\psi}\rangle_{BERB'} = \sqrt{1-\varepsilon'} \cdot G_{BE}^{-1} |\bar{\sigma}\rangle_{BERB'}. \tag{79}$$

We first show that all these vectors define subnormalized states such that the purified distance between them is well-defined. Since G_{BE} is a contraction, we immediately get that $\||\psi'\rangle_{BERB'}\| \leqslant 1$ and $\||\psi''\rangle_{BERB'}\| \leqslant 1$. Furthermore, we have that

$$\begin{aligned}
\left\| |\tilde{\psi}\rangle_{BERB'} \right\|^2 &= (1-\varepsilon') \cdot \langle \bar{\sigma} | {G_{BE}^{-1}}^\dagger G_{BE}^{-1} | \bar{\sigma} \rangle \\
&= \langle \bar{\sigma} | \sigma_E^{-1/2} \left((1-\varepsilon') \cdot \mathbb{1}_B \otimes \sigma_E + \varepsilon' Y_{BE} \right) \sigma_E^{-1/2} | \bar{\sigma} \rangle \\
&= 1 - \varepsilon' + \varepsilon' \cdot \langle \bar{\sigma} | \sigma_E^{-1/2} Y_{BE} \sigma_E^{-1/2} | \bar{\sigma} \rangle \\
&= 1 - \varepsilon' + \varepsilon' \cdot \operatorname{Tr}\left[Y_{BE} \sigma_E^{-1/2} \bar{\sigma}_{EB} \sigma_E^{-1/2} \right] \\
&= 1 - \varepsilon' + \varepsilon' \cdot \operatorname{Tr}\left[Y_{BE} \sigma_B \right] \\
&= 1.
\end{aligned} \tag{80}$$

We have $\langle \tilde{\psi} | \psi' \rangle = \sqrt{1-\varepsilon'}$, and

$$\begin{aligned}
\langle \bar{\sigma} | \tilde{\psi} \rangle &= \sqrt{1-\varepsilon'} \cdot \langle \bar{\sigma} | G_{BE}^{-1} | \bar{\sigma} \rangle \\
&= \operatorname{Tr}\left[(\sigma_B \otimes \sigma_E) \left((1-\varepsilon') \cdot \mathbb{1}_B \otimes \sigma_E + \varepsilon' Y_{BE} \right)^{1/2} \sigma_E^{-1/2} \right] \\
&= \operatorname{Tr}\left[(\sigma_B \otimes \sigma_E^{1/2}) \left((1-\varepsilon') \cdot \mathbb{1}_B \otimes \sigma_E + \varepsilon' Y_{BE} \right)^{1/2} \right] \\
&\geqslant \operatorname{Tr}\left[(\sigma_B \otimes \sigma_E^{1/2}) \cdot \sqrt{1-\varepsilon'} \cdot (\mathbb{1}_B \otimes \sigma_E^{1/2}) \right] \\
&= \sqrt{1-\varepsilon'} \cdot \operatorname{Tr}\left[\sigma_B \otimes \sigma_E \right] \\
&= \sqrt{1-\varepsilon'},
\end{aligned} \tag{81}$$

where the inequality is due to the operator monotonicity of the square-root function. Therefore, we have that $P(\psi', \bar{\sigma}) \leqslant 2\sqrt{\varepsilon'}$ and furthermore $P(\psi'', \bar{\sigma}) = P(\psi', \bar{\sigma})$, since

$$F(\psi'', \bar{\sigma}) = \langle \bar{\sigma} | G_{BE}^\dagger | \bar{\sigma} \rangle = F(\bar{\sigma}, \psi'). \tag{82}$$

Since conjugation by G is trace-non-increasing, we also have $P(\psi'', \psi) \leqslant P(\sigma, \bar{\sigma}) \leqslant \sqrt{6\varepsilon'' + 2\varepsilon}$. This implies

$$\begin{aligned}
P(\psi, \tilde{\sigma}) &\leqslant P(\psi, \psi'') + P(\psi'', \bar{\sigma}) + P(\bar{\sigma}, \sigma) + P(\sigma, \tilde{\sigma}) \\
&\leqslant \sqrt{6\varepsilon'' + 2\varepsilon} + 2\sqrt{\varepsilon'} + \sqrt{6\varepsilon'' + 2\varepsilon} + \varepsilon''.
\end{aligned} \tag{83}$$

□

5. One-Shot State Merging

As an example application of the decoupling theorem and its converse we discuss one-shot quantum state merging. This is a two-party task: its goal is to transfer the information contained in a quantum system, A, initially held by one party, Alice, to the other party, Bob. This should be achieved with only limited resources (such as entanglement or communication). It is taken into account that Bob may have access to a quantum system, B, correlated to A, which may be used to minimize the use of resources. The term one-shot is used to emphasize that the task is considered in the general one-shot scenario. As explained in the discussion section, the asymptotic iid results, where many independent copies of a given state are transferred, can be recovered as a special case.

The notion of quantum state merging has been introduced in [HOW05,HOW07] and a protocol has been proposed that achieves the task in the asymptotic iid scenario. The more general one-shot setup we consider here was first analyzed in [Ber08] and preliminary results appeared in [KRS09].

We start giving a formal definition of quantum state merging [HOW05,HOW07, Ber08]. Let ρ_{AB} be the joint initial state of Alice and Bob's systems. We can view this state as part of a larger pure state ρ_{ABE} that includes a reference system E. In this picture state merging means that Alice can send the A-part of ρ_{ABE} to Bob's side without altering the joint state. We consider the particular setting proposed in [HOW05] where classical communication from Alice to Bob is free, but no quantum communication is possible. Furthermore, Alice and Bob have access to a source of entanglement and their goal is to minimize the number of entangled bits consumed during the protocol (or maximize the number of entangled bits that can be generated).

Definition 5.1 (Quantum State Merging). *Let $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$, and let A_0B_0 be additional systems. A TPCPM $\mathcal{E} : AA_0 \otimes BB_0 \to A_1 \otimes B_1B'B$ is called quantum state merging of ρ_{AB} with error $\varepsilon \geqslant 0$, if it is a local operation and classical forward communication process for the bipartition $AA_0 \to A_1$ vs. $BB_0 \to B_1B'B$, and*

$$(\mathcal{E}_{AA_0BB_0\to A_1B_1B'B})(\Phi^K_{A_0B_0} \otimes \rho_{ABE}) \approx_\varepsilon \Phi^L_{A_1B_1} \otimes \rho_{BB'E}, \tag{84}$$

where $\rho_{BB'E} = (\mathcal{I}_{A\to B'} \otimes \mathcal{I}_{BE})\rho_{ABE}$ for a purification ρ_{ABE} of ρ_{AB}, and Φ^K, Φ^L are maximally entangled states on A_0B_0, A_1B_1 of Schmidt-rank K and L, respectively. The number

$$l^\varepsilon = \log K - \log L$$

is called entanglement cost.[5]

We are interested in quantifying the minimal entanglement cost for quantum state merging of ρ_{AB} with error ε. For this, we use the achievability and converse for decoupling (Theorems 3.1 and 4.1). These allow us to derive essentially tight (up to additive terms of the order $\log(1/\varepsilon)$ and the scaling of the smoothing parameter) bounds on the entanglement cost.

The basic idea underlying our analysis of quantum state merging is the observation that the desired situation after the protocol execution is necessarily such that Alice's system is decoupled from the reference. Furthermore, it follows from Uhlmann's theorem [Uhl76] that this decoupling is also sufficient.

[5] In the original references [HOW05,HOW07] quantum state merging was defined slightly differently, namely as a local operation and classical two-way communication process. However, their protocol for the achievability only uses classical forward communication.

Theorem 5.2 (Achievability for Quantum State Merging). *The minimal entanglement cost for quantum state merging of $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ with error $\varepsilon > 0$ is upper bounded by*

$$l^{\varepsilon} \leqslant H_{\max}^{\varepsilon^2/13}(A|B)_{\rho} + 4\log(1/\varepsilon) + 2\log 13. \tag{85}$$

Proof. Let ρ_{ABE} be a purification of ρ_{AB}. The intuition is as follows. In the first step of the protocol, Alice decouples her part from the reference (employing Theorem 3.1), where she chooses a rank-L projective measurement as the TPCPM, and she sends the measurement result to Bob. For all measurement outcomes the post-measurement state on Alice's side is then approximately given by $\frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \rho_E$ and Bob holds a purification of this. But $\frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \rho_E$ is the reduced state of $\Phi_{A_1B_1}^L \otimes \rho_{BB'E}$ as well and since all purifications are equal up to local isometries, there exists an isometry on Bob's side that transform the state into $\Phi_{A_1B_1}^L \otimes \rho_{BB'E}$ (by Uhlmann's theorem [Uhl76]); this is then the second step of the protocol.

More formally, choose K and L such that

$$\log K - \log L = H_{\max}^{\varepsilon^2/13}(A|B)_{\rho} + 4\log(1/\varepsilon) + 2\log 13, \tag{86}$$

which is the entanglement cost of the protocol.[6]

Choose N fixed orthogonal subspaces of dimension L on AA_0,[7] denote the projectors on these subspaces followed by a fixed unitary mapping it to A_1 by $P_{A_0A\to A_1}^x$ and define the isometry

$$W_{A_0A\to A_1X_AX_B} = \sum_x P_{A_0A\to A_1}^x \otimes |x\rangle_{X_A} \otimes |x\rangle_{X_B}. \tag{87}$$

Denote by U_{A_0A} a unitary selected randomly according to the Haar measure over the unitary group on $\mathcal{H}_{A_0A}$ and write

$$\theta_{A_0B_0ABE} = \Phi_{A_0B_0}^K \otimes \rho_{ABE} \tag{88}$$

$$\sigma_{A_0B_0ABE} = U_{A_0A}\theta_{A_0B_0ABE}U_{A_0A}^{\dagger}. \tag{89}$$

Now the first step of the protocol is to apply this unitary followed by the isometry (87), and to send the X_B system to Bob. In order to take into account that the channel is classical, we keep a copy X_A at Alice's side.

By the decoupling theorem (Theorem 3.1) we get for

$$\sigma_{A_1X_AX_BB_0BE} = (W_{A_0A\to A_1X_AX_B})\sigma_{A_0B_0ABE}(W_{A_0A\to A_1X_AX_B})^{\dagger}. \tag{90}$$

that

$$\|\sigma_{A_1X_AE} - \tau_{A_1X_A} \otimes \rho_E\|_1 \leqslant 2^{-1/2(H_{\min}^{\varepsilon^2/13}(A_0A|E)_{\theta} + H_{\min}^{\varepsilon^2/13}(A_0'A'|A_1X_A)_{\tau})} + \frac{12\varepsilon^2}{13}, \tag{91}$$

[6] Since we need $K, L \in \mathbb{N}$, we can not choose $\log K - \log L$ exactly equal to $H_{\max}^{\varepsilon^2/13}(A|B)_{\rho} + 4\log(1/\varepsilon) + 2\log 13$ in general. Rather, we need to choose $K, L \in \mathbb{N}$ such $\log K - \log L$ is minimal but still greater or equal than $H_{\max}^{\varepsilon^2/13}(A|B)_{\rho} + 4\log(1/\varepsilon) + 2\log 13$.

[7] For simplicity assume that $K \cdot |A|$ is divisible by L. In general one has to choose $N-1$ fixed orthogonal subspaces of dimension L and one of dimension $L' = K \cdot |A| - (N-1) \cdot L < L$. The proof remains the same, although some coefficients change.

where $A_0'A'$ is a copy of A_0A, and

$$|\tau\rangle_{A_0'A'A_1X_AX_B} = W_{A_0A\to A_1X_AX_B}|\Phi\rangle_{A_0'A'A_0A} \tag{92}$$

with

$$|\Phi\rangle_{A_0'A'A_0A} = \frac{1}{K\cdot|A|}\sum_i |i\rangle_{A_0'A'}\otimes|i\rangle_{A_0A}. \tag{93}$$

We can simplify this using the superadditivity of the smooth conditional min-entropy (Lemma A.2) and the duality between smooth conditional min- and max-entropy (Lemma 2.5)

$$H_{\min}^{\varepsilon^2/13}(A_0A|E)_\theta \geqslant H_{\min}^{\varepsilon^2/13}(A|E)_\rho + \log K = -H_{\max}^{\varepsilon^2/13}(A|B)_\rho + \log K. \tag{94}$$

Furthermore, because $\tau_{A_0'A'A_1X_A}$ is classical on X_A, we can use a lemma about the conditional min-entropy of classical-quantum states (Lemma A.5) and get

$$\begin{aligned} H_{\min}^{\varepsilon^2/13}(A_0'A'|A_1X_A)_\tau &\geqslant H_{\min}(A_0'A'|A_1X_A)_\tau \\ &= -\log\Big(\sum_x p_x\cdot 2^{-H_{\min}(A_0'A'|A_1)_{\tau^x}}\Big) \\ &\geqslant \min_x H_{\min}(A_0'A'|A_1)_{\tau^x}, \end{aligned} \tag{95}$$

where

$$\tau^x_{A_0'A'A_1} = \frac{1}{\sqrt{p_x}}P^x_{A_0A\to A_1}|\Phi\rangle_{A_0'A'A_0A} \tag{96}$$

$$p_x = \|P^x_{A_0A\to A_1}|\Phi\rangle_{A_0'A'A_0A}\|. \tag{97}$$

But since $P^x_{A_0A\to A_1}$ is a rank L projector, we can use a dimension lower bound of the conditional min-entropy (Lemma A.3) to conclude that for all x

$$H_{\min}(A_0'A'|A_1)_{\tau^x} \geqslant -\log L. \tag{98}$$

This together with (86), (91) and (94) implies

$$\begin{aligned} \left\|\sigma_{A_1X_AE} - \frac{\mathbb{1}_{A_1}}{|A_1|}\otimes\tau_{X_A}\otimes\rho_E\right\|_1 &= \left\|\sigma_{A_1X_AE} - \tau_{A_1X_A}\otimes\rho_E\right\|_1 \\ &\leqslant 2^{-1/2(\log K - \log L - H_{\max}^{\varepsilon^2/13}(A|B)_\rho)} + \frac{12\varepsilon^2}{13} \\ &= 2^{-1/2(4\log(1/\varepsilon)+2\log 13)} + \frac{12\varepsilon^2}{13} = \varepsilon^2, \end{aligned} \tag{99}$$

and hence $F(\sigma_{A_1X_AE}, \frac{\mathbb{1}_{A_1}}{|A_1|}\otimes\tau_{X_A}\otimes\rho_E) \geqslant 1-\varepsilon^2/2$ (by Lemma B.1).

In the second step of the protocol, Bob decodes the system to the state $\rho_{BB'E}\otimes\Phi_{A_1B_1}$. A suitable decoder can be shown to exist using Uhlmann's theorem [Uhl76]. There exists an isometry $\mathcal{V}_{BB_0X_B\to BB'B_1X_B}$ such that for

$$\eta_{A_1X_AX_BBB'B_1E} = \mathcal{V}_{BB_0X_B\to BB'B_1X_B}(\sigma_{A_1X_AX_BBB_0E}) \tag{100}$$

$$F(\sigma_{A_1X_AE}, \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \tau_{X_A} \otimes \rho_E) = F(\eta_{A_1X_AX_BBB'B_1E}, \tau_{X_AX_B} \otimes \Phi^L_{A_1B_1} \otimes \rho_{BB'E}), \tag{101}$$

and with that

$$F(\eta_{A_1X_AX_BBB'B_1E}, \tau_{X_AX_B} \otimes \Phi^L_{A_1B_1} \otimes \rho_{BB'E}) \geqslant 1 - \frac{\varepsilon^2}{2}. \tag{102}$$

Expressing this in the purified distance (with Lemma B.1) and discarding X_AX_B, we obtain a ε-error quantum state merging protocol for ρ_{ABE}. □

Theorem 5.3 (Converse for Quantum State Merging). *The minimal entanglement cost for quantum state merging of $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ with error $\varepsilon > 0$ is lower bounded by*

$$l^\varepsilon \geqslant H_{\max}^{4\sqrt{2\varepsilon}+3\varepsilon}(A|B)_\rho - 2\log\frac{1}{\varepsilon}. \tag{103}$$

Proof. We start with noting that any ε-error quantum state merging protocol for ρ_{AB} can be assumed to have the following form: applying local operations at Alice's side, then sending a classical register from Alice to Bob, and finally applying local operations at Bob's side. For a purified state ρ_{ABE}, the protocol produces a state ε-close to $\Phi^L_{A_1B_1} \otimes \rho_{BB'E}$.

As can be seen from the definition, it is a necessary step for any quantum state merging protocol to decouple Alice's part from the reference. The idea of the proof is to use the converse for decoupling (Theorem 4.1). This then results in the desired converse for quantum state merging.

More precisely, a general ε-error quantum state merging protocol for ρ_{ABE} has the following form. At first some TPCPM

$$\mathcal{T}_{A_0A\to A_1X_B}(.) = \sum_x M^x_{A_0A\to A_1}(.) \otimes |x\rangle\langle x|_{X_B} \tag{104}$$

is applied to the input state $\Phi^K_{A_0B_0} \otimes \rho_{ABE}$. By the Stinespring dilation [Sti55] we can think of this TPCPM as an isometry

$$W_{A_0A\to A_1A_GX_BX_A} = \sum_x M^x_{A_0A\to A_1A_G} \otimes |x\rangle_{X_A} \otimes |x\rangle_{X_B}, \tag{105}$$

where the $M^x_{A_0A\to A_1A_G}$ are partial isometries and A_G, X_A are additional 'garbage' registers on Alice's side that will be discarded in the end. The isometry W results in the state

$$|\gamma\rangle_{A_1A_GX_AX_BBB_0E} = \sum_x |\gamma^x\rangle_{A_1A_GBB_0E} \otimes |x\rangle_{X_A} \otimes |x\rangle_{X_B}, \tag{106}$$

with

$$|\gamma^x\rangle_{A_1A_GBB_0E} = M^x_{A_0A\to A_1A_G}(|\Phi^K\rangle_{A_0B_0} \otimes |\rho\rangle_{ABE}). \tag{107}$$

The next step of the protocol is then to send the classical register X_B to Bob.

Now let us analyze how the state $\gamma_{A_1A_GX_AE}$ has to look like. By the definition of quantum state merging (Definition 5.1) the state at the end of the protocol has to be ε-close to $\Phi^L_{A_1B_1} \otimes \rho_{BB'E}$. This implies that Alice's part A_1 has to be decoupled from

the reference. But because the state $\Phi^L_{A_1B_1} \otimes \rho_{BB'E}$ is pure this also implies that all additional registers, that we might have at the end of the protocol, have to be decoupled as well. Thus we need

$$\gamma_{A_1A_GX_AE} \approx_\varepsilon \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \gamma_{A_GX_A} \otimes \rho_E, \tag{108}$$

and in trace distance (using Lemma B.1) this reads

$$\left\| \gamma_{A_1A_GX_AE} - \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \gamma_{A_GX_A} \otimes \rho_E \right\|_1 \leqslant 2\varepsilon. \tag{109}$$

Using the converse for decoupling (Theorem 4.1) for the isometry $W_{A_0A\to A_1A_GX_BX_A}$ in (105) followed by the partial trace over X_B, we get that the decoupling condition (109) implies for any $\varepsilon', \varepsilon'' > 0$ that

$$H_{\min}^{2\sqrt{6\varepsilon''+2\varepsilon}+2\sqrt{\varepsilon'}+\varepsilon''}(A_0A|E)_\rho + H_{\max}^{\varepsilon''}(A_0'A'|A_1A_GX_A)_\omega \geqslant -\log\frac{1}{\varepsilon'}, \tag{110}$$

where

$$\omega_{A_0'A'A_1A_GX_A} = \mathrm{tr}_{X_B}\left[(W_{A_0A\to A_1A_GX_BX_A})\zeta_{A_0'A'A_0A}(W^\dagger_{A_0A\to A_1A_GX_BX_A})\right] \tag{111}$$

for $\zeta_{A_0'A'A_0A}$ a purification of $\frac{\mathbb{1}_{A_0}}{|A_0|} \otimes \rho_A$ with $A_0'A'$ a copy of A_0A. As a next step we simplify this in order to bring the converse into the desired form.

Choosing $\varepsilon' = \varepsilon^2$ and $\varepsilon'' = \varepsilon$, using a dimension upper bound for the smooth conditional min-entropy (Lemma A.4), and the duality between smooth conditional min- and max-entropy (Lemma 2.5) we obtain

$$\log K + H_{\max}^{\varepsilon}(A_0'A'|A_1A_GX_A)_\omega \geqslant H_{\max}^{4\sqrt{2\varepsilon}+3\varepsilon}(A|B)_\rho - 2\log\frac{1}{\varepsilon}. \tag{112}$$

By the decoupling criterion in purified distance (Eq. (108)), the state $\omega_{A_0'A'A_1A_GX_A}$ has to be ε-close to a state

$$\xi_{A_0'A'A_1A_GX_A} = \sum_x q_x \xi^x_{A_0'A'A_1A_G} \otimes |x\rangle\langle x|_{X_A} \tag{113}$$

where q_x is some probability distribution and $\xi^x_{A_0'A'A_1A_G}$ pure with $\xi^x_{A_1A_G} = \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \xi^x_{A_G}$ for all x. Hence

$$H_{\max}^{\varepsilon}(A_0'A'|A_1A_GX_A)_\omega \leqslant H_{\max}(A_0'A'|A_1A_GX_A)_\xi \tag{114}$$

and by a lemma about the conditional max-entropy of classical-quantum states (Lemma A.6)

$$H_{\max}(A_0'A'|A_1A_GX_A)_\xi = \log\left(\sum_x q_x \cdot 2^{H_{\max}(A_0'A'|A_1A_G)_{\xi^x}}\right). \tag{115}$$

Using the duality between conditional min- and max-entropy (Lemma 2.5) and a polar decomposition of $\xi^x_{A_0'A'A_1A_G}$, we get

$$H_{\max}(A_0'A'|A_1A_G)_{\xi^x} = -H_{\min}(A_0'A')_{\xi^x}$$

$$
\begin{aligned}
&= -H_{\min}(A_1 A_G)_{\xi^x} \\
&= -H_{\min}(A_1)_{\frac{\mathbb{1}}{|A_1|}} - H_{\min}(A_G)_{\xi^x} \\
&\leqslant -H_{\min}(A_1)_{\frac{\mathbb{1}}{|A_1|}} \\
&= -\log L. \qquad (116)
\end{aligned}
$$

Hence, the converse becomes

$$
\log K - \log L \geqslant H_{\max}^{4\sqrt{2\varepsilon}+3\varepsilon}(A|B)_\rho - 2\log\frac{1}{\varepsilon}. \qquad (117)
$$

□

6. Discussion

The main contribution of this work is a decoupling theorem, i.e., a sufficient (Theorem 3.1) and necessary (Theorem 4.1) criterion for decoupling in terms of smooth conditional entropies. These criteria can then be applied to obtain tight characterizations of various operational tasks. As outlined in Sect. 5 by means of state merging, such applications are often possible because of a duality between independence and maximum entanglement: given a pure state ρ_{BER} such that ρ_B is maximally mixed, the property that the subsystem B is independent of E and the property that B is fully entangled with R are equivalent.

A crucial property of our decoupling criterion is that it gives (nearly optimal) bounds in a one-shot scenario, where the decoupling map $\mathcal{T}$ may only be applied once (or, by replacing $\mathcal{T}$ by $\mathcal{T}^{\otimes k}$, any finite number of times). For a typical example, consider m qubits, A, and assume that A undergoes a reversible evolution, $\mathcal{U}$, after which we discard $m - m'$ qubits, corresponding to a partial trace, $\mathcal{T} = \mathrm{Tr}_{m-m'}$ (see last example of Table 2). Our decoupling theorem (Theorem 3.1) then shows that decoupling up to an error ε is achieved for most choices of $\mathcal{U}$ if

$$
m' \lessapprox \frac{1}{2}\left(m + H_{\min}^{\varepsilon}(A|E)_\rho\right). \qquad (118)
$$

In contrast to this, the original decoupling results [ADHW09], formulated in terms of smooth non-conditional entropies, only show that decoupling up to an error ε is achieved for most choices of $\mathcal{U}$ if

$$
m' \lessapprox \frac{1}{2}\left(m + H_{\min}^{\varepsilon}(AE)_\rho - H_{\max}^{\varepsilon}(E)_\rho\right). \qquad (119)
$$

To see that this latter bound may be arbitrarily weaker than the bound (118) that uses smooth conditional entropies, consider the following completely classical state. Let A and E be perfectly correlated, and let the marginal distribution of A (and E) have one value that is taken with probability 1/2, and be uniform over the remaining $2^m - 1$ values. Then we have (for $\varepsilon \geqslant 0$ close to zero)

$$
H_{\min}^{\varepsilon}(A|E)_\rho \approx 0 \quad \text{vs.} \quad H_{\min}^{\varepsilon}(AE)_\rho - H_{\max}^{\varepsilon}(E)_\rho \approx 1 - m. \qquad (120)
$$

The difference between these two bounds is conceptually relevant. An example illustrating this is the quantitative Landauer's principle derived recently in [FDOR12]. The result, which is based on the bound (118), shows that correlations between the inputs

and outputs of an irreversible mapping are relevant for the thermodynamic work cost of implementations of the mapping. These correlations would not be accounted for if a bound of the form (119) was used for the derivation of the principle.

In contrast to the original results on decoupling that are based on specific decoupling processes (where the mapping $\mathcal{T}$ is either a partial trace [ADHW09] or a projective measurement [HOW07]), our decoupling criterion is also applicable to general mappings $\mathcal{T}$. This extension is, e.g., employed in [Hut11, Sect. 5] in order to discuss the postulate of equal a priori probability in quantum statistical mechanics.

Our generalizations of the decoupling technique are crucial for other applications in physics as well, e.g., for the analysis of thermodynamic systems [dRAR+11], for finding an efficient classical description of 1D quantum states with an exponential decay of correlations [BH13], or for the study of black hole radiation [HP07, BP07, PZ13].

Information-theoretic applications other than state merging (cf. Sect. 5) have been investigated in the doctoral thesis of one of the authors [Dup09]. One of these applications is channel coding. Here, Alice wants to use a noisy quantum channel $\mathcal{N}^{A\to B}$ to send qubits to Bob with fidelity at least $1-\varepsilon$. The idea is that decoding is possible whenever a purification of the qubits Alice is sending is decoupled from the channel environment. One can therefore get a coding theorem directly from Theorem 3.1 by setting $\mathcal{T}$ to be the complementary channel of $\mathcal{N}$ (i.e., consider a Stinespring dilation [Sti55] $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ of $\mathcal{N}$, and set $\mathcal{T}_{A\to E}(\cdot) = \mathrm{Tr}_B[U_A \cdot U_A^\dagger]$). Unassisted channel coding [Llo97, Sho02, Dev05] can be obtained by choosing the input state $\rho_{AR} = \Phi_{AR}$ (where Φ_{AR} is a maximally entangled state between A and R). Similarly, entanglement-assisted channel coding [BSST02] corresponds to the input choice $\rho_{ABR} = \Phi_{A_R R} \otimes \Phi_{A_B B}$ (where $\mathcal{H}_A = \mathcal{H}_{A_R} \otimes \mathcal{H}_{A_B}$, with A_R containing the state to be transmitted and A_B the initial entanglement that Alice shares with Bob). Other choices of ρ_{ABR} correspond to different scenarios.

Another application where decoupling can be employed as a building block for constructing protocols is the simulation of noisy quantum channels using perfect classical channels together with pre-shared entanglement. The fully quantum reverse Shannon theorem asserts that this is possible using only a classical communication rate equal to the capacity of the channel to be simulated [BSST02, BDH+09]. In [BCR11], a proof of this theorem using one-shot decoupling has been proposed.

Our one-shot decoupling results contrast with (and are strictly more general than) the iid scenario[8] usually considered in information theory, where statements are proved asymptotically under the assumption that the underlying processes (such as channel uses) are repeated many times independently. We note that asymptotic iid statements can be easily retrieved from the general one-shot results using the quantum asymptotic equipartition property (AEP) for smooth entropies [Ren05, TCR09] (see Lemma 2.8). Consider decoupling with a map of the form $\tilde{\mathcal{T}} = \mathcal{T} \circ \mathcal{U}$ (with $\mathcal{U}$ a random unitary channel). If the map $\mathcal{T}$ as well as the initial state ρ_{AE} consist of many identical copies, i.e., $\mathcal{T}^{\otimes n}$ and $\rho_{AE}^{\otimes n}$, then the achievability bound of Theorem 3.1, i.e., the condition that is sufficient for decoupling, turns into the criterion

$$H(A|E)_\rho + H(A|B)_\tau \geqslant 0, \tag{121}$$

where H denotes the (conditional) von Neumann entropy. Analogously, the converse in Corollary 4.2 (i.e., the condition which is necessary for decoupling for maps of this form) turns into

$$H(A|E)_\rho + H(A|B)_\tau \leqslant 0. \tag{122}$$

[8] The abbreviation iid stands for independent and identically distributed.

In other words, in the iid scenario, the achievability bound (121) and the converse bound (122), taken together, imply an exact characterization of decoupling.

Acknowledgements. We thank Andreas Winter for insightful discussions and for his valuable contributions to [Ber08], which served as a starting point for this work. We also thank Patrick Hayden for enlightening discussions, as well as Oleg Szehr for fixing a bug regarding smoothing in the proof of Theorem 3.1, among other useful comments. We acknowledge support from the Swiss National Science Foundation (Grants No. 200021-119868 and 200020-135048), the National Centre of Competence in Research 'Quantum Science and Technology (QSIT)', and the European Research Council (Grant No. 258932). FD was supported by Canada's NSERC Postdoctoral Fellowship Program. MB was supported by the German Science Foundation (Grant CH 843/2-1), the Swiss National Science Foundation (Grants PP00P2-128455, 20CH21-138799 (CHIST-ERA project CQC)), and the Swiss State Secretariat for Education and Research supporting COST action MP1006. JW was funded by the UK EPSRC grant EP/E04297X/1 and the Canada–France NSERC-ANR project FREQUENCY. Parts of this work were done while JW was at the University of Bristol.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

A. Properties of smooth entropies

The conditional collision entropy is lower bounded by the conditional min-entropy.

Lemma A.1. *Let $\rho_{AB} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB})$. Then, we have that $H_2(A|B)_\rho \geqslant H_{\min}(A|B)_\rho$.*

Proof. Let $\sigma_B \in \mathcal{S}_=(\mathcal{H}_B)$ be such that $\rho_{AB} \leqslant 2^{-H_{\min}(A|B)_\rho} \cdot \mathbb{1}_A \otimes \sigma_B$. We then obtain

$$\begin{aligned} 2^{-H_2(A|B)_\rho} &= \min_{\omega_B} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \omega_B)^{-1/2}\rho_{AB}(\mathbb{1}_A \otimes \omega_B)^{-1/2}\rho_{AB}\right] \\ &\leqslant \operatorname{Tr}\left[(\mathbb{1}_A \otimes \sigma_B)^{-1/2}\rho_{AB}(\mathbb{1}_A \otimes \sigma_B)^{-1/2}\rho_{AB}\right] \\ &\leqslant 2^{-H_{\min}(A|B)_\rho} \cdot \operatorname{Tr}\left[\mathbb{1}_{AB}\rho_{AB}\right] \\ &\leqslant 2^{-H_{\min}(A|B)_\rho}. \end{aligned} \tag{123}$$

□

The smooth conditional min-entropy is superadditive.

Lemma A.2. *Let $\varepsilon, \varepsilon' \geqslant 0$, $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ and $\rho'_{A'B'} \in \mathcal{S}_=(\mathcal{H}_{A'B'})$. Then, we have that*

$$H_{\min}^{\varepsilon+\varepsilon'}(AA'|BB')_{\rho\otimes\rho'} \geqslant H_{\min}^{\varepsilon}(A|B)_\rho + H_{\min}^{\varepsilon'}(A'|B')_{\rho'}. \tag{124}$$

Proof. Let $\bar{\rho}_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})$ and $\bar{\rho}'_{A'B'} \in \mathcal{B}^\varepsilon(\rho'_{A'B'})$ such that $H_{\min}^{\varepsilon}(A|B)_\rho = H_{\min}(A|B)_{\bar{\rho}}$ and $H_{\min}^{\varepsilon'}(A'|B')_{\rho'} = H_{\min}(A'|B')_{\bar{\rho}'}$. By the triangle inequality for the purified distance [TCR10, Lemma 5] we have $\bar{\rho}_{AB} \otimes \bar{\rho}'_{A'B'} \in \mathcal{B}^{\varepsilon+\varepsilon'}(\rho_{AB} \otimes \rho'_{A'B'})$. Using the additivity of the conditional min-entropy [KRS09], we conclude

$$\begin{aligned} H_{\min}^{\varepsilon+\varepsilon'}(AA'|BB')_{\rho\otimes\rho'} &\geqslant H_{\min}(AA'|BB')_{\bar{\rho}\otimes\bar{\rho}'} \\ &= H_{\min}(A|B)_{\bar{\rho}} + H_{\min}(A'|B')_{\bar{\rho}'} \\ &= H_{\min}^{\varepsilon}(A|B)_\rho + H_{\min}^{\varepsilon'}(A'|B')_{\rho'}. \end{aligned} \tag{125}$$

□

We have the following dimension lower and upper bounds for the (smooth) conditional min-entropy.

Lemma A.3 [TCR10, Lemma 20]. *Let $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$. Then, we have that $H_{\min}(A|B)_\rho \geqslant -\log|B|$.*

Lemma A.4. *Let $\varepsilon \geqslant 0$ and $\rho_{ABC} \in \mathcal{S}_=(\mathcal{H}_{ABC})$. Then, we have that*

$$H^\varepsilon_{\min}(AB|C)_\rho \leqslant H^\varepsilon_{\min}(A|C)_\rho + \log|B|. \tag{126}$$

Proof. Let $\bar{\rho}_{ABC} \in \mathcal{B}^\varepsilon(\rho_{ABC})$, $\sigma_C \in \mathcal{S}_=(\mathcal{H}_C)$ and $\lambda \in \mathbb{R}$ such that

$$H^\varepsilon_{\min}(AB|C)_\rho = H_{\min}(AB|C)_{\bar{\rho}} = -\log\lambda, \tag{127}$$

that is, λ is minimal such that $\lambda \cdot \mathbb{1}_{AB} \otimes \sigma_C - \bar{\rho}_{ABC} \geqslant 0$. By taking the partial trace over B we get $\lambda \cdot |B| \cdot \mathbb{1}_A \otimes \sigma_C - \bar{\rho}_{AC} \geqslant 0$. Furthermore we have by the monotonicity of the purified distance [TCR10, Lemma 7] that $\bar{\rho}_{AC} \in \mathcal{B}^\varepsilon(\rho_{AC})$ and hence

$$H^\varepsilon_{\min}(A|C)_\rho \geqslant H_{\min}(A|C)_{\bar{\rho}} \geqslant -\log\mu, \tag{128}$$

where $\mu \in \mathbb{R}$ is minimal such that $\mu \cdot \mathbb{1}_A \otimes \sigma_C - \bar{\rho}_{AC} \geqslant 0$. Thus $\lambda \cdot |B| \geqslant \mu$ and therefore

$$H^\varepsilon_{\min}(AB|C)_\rho \leqslant H^\varepsilon_{\min}(A|C)_\rho + \log|B|. \tag{129}$$

□

The following lemma is about the conditional min-entropy of quantum-classical states.

Lemma A.5. *Let $\rho_{ABX} \in \mathcal{S}_=(\mathcal{H}_{ABX})$ with $\rho_{ABX} = \sum_x p_x \cdot \rho^x_{AB} \otimes |x\rangle\langle x|_X$ and $\rho^x_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ for all x. Then, we have that*

$$H_{\min}(A|BX)_\rho = -\log\Big(\sum_x p_x \cdot 2^{-H_{\min}(A|B)_{\rho^x}}\Big). \tag{130}$$

Proof. By the operational interpretation of the conditional min-entropy as the maximal achievable singlet fraction [KRS09, Theorem 2] we have

$$H_{\min}(A|BX)_\rho = -\log(|A| \cdot \max_{\mathcal{F}_{BX\to A'}} F^2((\mathcal{I}_A \otimes \mathcal{F}_{BX\to A'})(\rho_{ABX}), |\Phi\rangle\langle\Phi|_{AA'})), \tag{131}$$

where the maximum is taken over all TPCPMs $\mathcal{F}_{BX\to A'}$, $|\Phi\rangle_{AA'} = |A|^{-1/2}\sum_i |x\rangle_A \otimes |x\rangle_{A'}$, and $\mathcal{H}_{A'} \cong \mathcal{H}_A$. Writing out the conditional min-entropy terms on the right hand side of (130) in the same manner we obtain

$$H_{\min}(A|B)_{\rho^x} = -\log\left(|A| \cdot \max_{\mathcal{F}^x_{B\to A'}} F^2((\mathcal{I}_A \otimes \mathcal{F}^x_{B\to A'})(\rho^x_{AB}), |\Phi\rangle\langle\Phi|_{AA'})\right). \tag{132}$$

The claim is therefore equivalent to

$$\max_{\mathcal{F}_{BX\to A'}} F^2((\mathcal{I}_A \otimes \mathcal{F}_{BX\to A'})(\rho_{ABX}), |\Phi\rangle\langle\Phi|_{AA'})$$

$$= \sum_x p_x \cdot \max_{\mathcal{F}^x_{B\to A'}} F^2((\mathcal{I}_A \otimes \mathcal{F}^x_{B\to A'})(\rho^x_{AB}), |\Phi\rangle\langle\Phi|_{AA'}). \tag{133}$$

Now, because the state ρ_{ABX} is classical on X, the maximization on the left hand side can without loss of generality be restricted to TPCPMs that first measure on X in the basis $\{|x\rangle\}$ and then do some TPCPM $\mathcal{F}^x_{B\to A'}$ conditioned on the measurement outcome x. By the linearity of the square of the fidelity when one argument is pure, the claim then follows. □

The following lemma is about the conditional max-entropy of quantum-classical states.

Lemma A.6. *Let $\rho_{ABX} \in \mathcal{S}_=(\mathcal{H}_{ABX})$ with $\rho_{ABX} = \sum_x p_x \cdot \rho^x_{AB} \otimes |x\rangle\langle x|_X$ and $\rho^x_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ for all x. Then, we have that*

$$H_{\max}(A|BX)_\rho = \log(\sum_x p_x \cdot 2^{H_{\max}(A|B)_{\rho^x}}). \tag{134}$$

Proof. Let $\rho_{ABCXX'}$ be a purification of ρ_{ABX}. Then, we have by the duality of conditional min- and max-entropy (Lemma 2.5) and a lemma about the conditional min-entropy of quantum-classical states (Lemma A.5) that

$$\begin{aligned} H_{\max}(A|BX)_\rho &= -H_{\min}(A|CX')_\rho = \log(\sum_x p_x \cdot 2^{-H_{\min}(A|C)_{\rho^x}}) \\ &= \log(\sum_x p_x \cdot 2^{H_{\max}(A|B)_{\rho^x}}). \end{aligned} \tag{135}$$

□

The following lemma is property of the smooth conditional min-entropy of quantum-classical states.

Lemma A.7. *Let $\varepsilon \geqslant 0$ and $\rho_{ABXX'} \in \mathcal{S}_=(\mathcal{H}_{ABXX'})$ with $\rho_{ABXX'} = \sum_x p_x \cdot \rho^x_{AB} \otimes |x\rangle\langle x|_X \otimes |x\rangle\langle x|_{X'}$ and $\rho^x_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ for all x. Then, we have that*

$$H^\varepsilon_{\min}(A|BX)_\rho = H^\varepsilon_{\min}(AX'|BX)_\rho. \tag{136}$$

Proof. We first show the case $\varepsilon = 0$. By a property of the conditional min-entropy of quantum-classical states (Lemma A.5), the claim becomes equivalent to

$$H_{\min}(A|B)_{\rho^x} = H_{\min}(AX'|B)_{\rho^x \otimes |x\rangle\langle x|}. \tag{137}$$

But by the additivity of the conditional min-entropy [KRS09] this holds.

For $\varepsilon > 0$, let $\bar{\rho}_{ABXX'} \in \mathcal{B}^\varepsilon(\rho_{ABXX'})$ be classical on XX' with respect to the basis $\{|x\rangle \otimes |x\rangle\}_x$ such that $H^\varepsilon_{\min}(AX'|BX)_\rho = H_{\min}(AX'|BX)_{\bar{\rho}}$ (which is possible by [Tom12, Proposition 5.8]). Since the purified distance is monotone under trace non-increasing CPMs [TCR10, Lemma 7], we have $\bar{\rho}_{ABX} \in \mathcal{B}^\varepsilon(\rho_{ABX})$ and hence

$$H^\varepsilon_{\min}(AX'|BX)_\rho \leqslant H^\varepsilon_{\min}(A|BX)_\rho. \tag{138}$$

For the inequality in the other direction, let $\hat{\rho}_{ABX} \in \mathcal{B}^\varepsilon(\rho_{ABX})$ be classical on X with respect to the basis $\{|x\rangle\}_x$ such that $H^\varepsilon_{\min}(A|BX)_\rho = H_{\min}(A|BX)_{\hat{\rho}}$ (which is possible by [Tom12, Proposition 5.8]). By [TCR10, Corollary 9] and the monotonicity of the purified distance under trace non-increasing CPMs [TCR10, Lemma 7] there exists an

extension $\hat{\rho}_{ABXX'} \in \mathcal{B}^{\varepsilon}(\rho_{ABXX'})$ of $\hat{\rho}_{AXB}$ that is classical on XX' with respect to the basis $\{|x\rangle \otimes |x\rangle\}_x$. Thus, we conclude

$$H_{\min}^{\varepsilon}(A|BX)_{\rho} \leqslant H_{\min}^{\varepsilon}(AX'|BX)_{\rho}. \tag{139}$$

□

We have the following chain rule for the smooth conditional min-entropy.

Lemma A.8. *Let $\varepsilon > 0$, $\varepsilon', \varepsilon'' \geqslant 0$ and $\rho_{ABC} \in \mathcal{S}_{=}(\mathcal{H}_{ABC})$. Then, we have that*

$$H_{\min}^{\varepsilon+2\varepsilon'+\varepsilon''}(AB|C)_{\rho} \geqslant H_{\min}^{\varepsilon'}(A|BC)_{\rho} + H_{\min}^{\varepsilon''}(B|C)_{\rho} - \log\frac{2}{\varepsilon^2}. \tag{140}$$

Proof. Let $\rho'_{ABC} \in \mathcal{B}^{\varepsilon'}(\rho_{ABC})$ such that $H_{\min}^{\varepsilon'}(A|BC)_{\rho} = H_{\min}(A|BC)_{\rho'}$ and let ρ'_{ABCE} be a purification of ρ'_{ABC}. Furthermore let $\rho''_{BC} \in \mathcal{B}^{\varepsilon''}(\rho_{BC})$, $\sigma_C \in \mathcal{S}_{=}(\mathcal{H}_{BC})$ and $\lambda \in \mathbb{R}$ such that $H_{\min}^{\varepsilon''}(B|C)_{\rho} = H_{\min}(B|C)_{\rho''} = -\log\lambda$, that is, λ is minimal such that

$$\lambda \cdot \mathbb{1}_B \otimes \sigma_C - \rho''_{BC} \geqslant 0. \tag{141}$$

By [TRSS10, Lemma 21] there exists a projector P_{AE} such that

$$\bar{\rho}'_{ABCE} = (P_{AE} \otimes \mathbb{1}_{BC})\rho'_{ABCE}(P_{AE} \otimes \mathbb{1}_{BC}) \in \mathcal{B}^{\varepsilon}(\rho'_{ABCE}), \tag{142}$$

and

$$2^{-H_{\min}^{\varepsilon'}(A|BC)_{\rho}+\log\frac{2}{\varepsilon^2}} \cdot \mathbb{1}_A \otimes \rho'_{BC} - \bar{\rho}'_{ABC} \geqslant 0. \tag{143}$$

Now let T_{BC} be defined as in Lemma B.3 with $\rho''_{BC} = T_{BC}\rho'_{BC}T_{BC}^{\dagger}$ and consider the state

$$\bar{\rho}''_{ABCE} = (\mathbb{1}_{AE} \otimes T_{BC})\bar{\rho}'_{ABCE}(\mathbb{1}_{AE} \otimes T_{BC}^{\dagger}) = (P_{AE} \otimes T_{BC})\rho'_{ABCE}(P_{AE} \otimes T_{BC}^{\dagger}). \tag{144}$$

Applying T_{BC} to (143) we obtain

$$2^{-H_{\min}^{\varepsilon'}(A|BC)_{\rho}+\log\frac{2}{\varepsilon^2}} \cdot \mathbb{1}_A \otimes \rho''_{BC} - \bar{\rho}''_{ABC} \geqslant 0. \tag{145}$$

Together with (141) this yields

$$2^{-H_{\min}^{\varepsilon'}(A|BC)_{\rho}+\log\frac{2}{\varepsilon^2}-H_{\min}^{\varepsilon''}(B|C)_{\rho}} \cdot \mathbb{1}_{AB} \otimes \sigma_C - \bar{\rho}''_{ABC} \geqslant 0. \tag{146}$$

This implies

$$H_{\min}(AB|C)_{\bar{\rho}''} \geqslant H_{\min}^{\varepsilon'}(A|BC)_{\rho} + H_{\min}^{\varepsilon''}(B|C)_{\rho} - \log\frac{2}{\varepsilon^2}. \tag{147}$$

But by the monotonicity of the purified distance [TCR10, Lemma 7] and the definition of T_{BC} we have

$$\begin{aligned} P(\bar{\rho}''_{ABC}, \bar{\rho}'_{ABC}) \leqslant P((&P_{AE} \otimes T_{BC})\rho'_{ABCE}(P_{AE} \otimes T_{BC}^{\dagger}), \\ &(P_{AE} \otimes \mathbb{1}_{BC})\rho'_{ABCE}(P_{AE} \otimes \mathbb{1}_{BC})) \end{aligned}$$

$$\leqslant P((\mathbb{1}_{AE} \otimes T_{BC})\rho'_{ABCE}(\mathbb{1}_{AE} \otimes T_{BC}^{\dagger}), \rho'_{ABCE})$$
$$= P(\rho''_{BC}, \rho'_{BC}), \tag{148}$$

and hence

$$P(\bar{\rho}''_{ABC}, \bar{\rho}'_{ABC}) \leqslant P(\rho''_{BC}, \rho_{BC}) + P(\rho_{BC}, \rho'_{BC}) \leqslant \varepsilon'' + \varepsilon'. \tag{149}$$

Finally we obtain

$$P(\bar{\rho}''_{ABC}, \rho_{ABC}) \leqslant P(\bar{\rho}''_{ABC}, \bar{\rho}'_{ABC}) + P(\bar{\rho}'_{ABC}, \rho'_{ABC}) + P(\rho'_{ABC}, \rho_{ABC})$$
$$\leqslant \varepsilon'' + \varepsilon' + \varepsilon + \varepsilon' = \varepsilon + 2\varepsilon' + \varepsilon', \tag{150}$$

and thus together with (147) that

$$H_{\min}^{\varepsilon+2\varepsilon'+\varepsilon'}(AB|C)_\rho \geqslant H_{\min}^{\varepsilon'}(A|BC)_\rho + H_{\min}^{\varepsilon''}(B|C)_\rho - \log \frac{2}{\varepsilon^2}. \tag{151}$$

□

B. Technical lemmas

Lemma B.1 [TCR10, Lemma 6]. *Let $\rho, \sigma \in \mathcal{S}_{\leqslant}(\mathcal{H})$. Then, we have that*

$$\bar{D}(\rho, \sigma) \leqslant P(\rho, \sigma) \leqslant \sqrt{2\bar{D}(\rho, \sigma)} \leqslant \sqrt{2\|\rho - \sigma\|_1} \tag{152}$$

$$\frac{1}{2}P(\rho, \sigma)^2 \leqslant \bar{D}(\rho, \sigma) \leqslant P(\rho, \sigma), \tag{153}$$

where $\bar{D}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1 + \frac{1}{2}|\operatorname{Tr}[\rho] - \operatorname{Tr}[\sigma]|$.

Lemma B.2. *Let $\rho_{ABC} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{ABC})$ be pure. Then, we have that for any $\sigma_B \in \mathcal{S}_{=}(\mathcal{H}_B)$ with full rank,*

$$\rho_{ABC} \leqslant Z_{AB} \otimes \mathbb{1}_C, \tag{154}$$

where $Z_{AB} = 2^{\frac{1}{2}H_{\max}(A|B)_{\rho|\sigma}} \cdot \sigma_B^{-1/2}\sqrt{\sigma_B^{1/2}\rho_{AB}\sigma_B^{1/2}}\sigma_B^{-1/2}$. Furthermore, Z_{AB} has the property that $\operatorname{Tr}[Z_{AB}\sigma_B] = 2^{H_{\max}(A|B)_{\rho|\sigma}}$.

Proof. Consider the following semidefinite program (for a introduction to semidefinite programs presented in this manner, see for instance [Wat08]):

	Primal		Dual
maximize:	$\operatorname{Tr}[\rho_{ABC}X_{ABC}]$	minimize:	$\operatorname{Tr}[(\mathbb{1}_A \otimes \sigma_B)Z_{AB}]$
subject to:	$\operatorname{Tr}_C[X_{ABC}] = \mathbb{1}_A \otimes \sigma_B$	subject to:	$\rho_{ABC} \leqslant Z_{AB} \otimes \mathbb{1}_C$.
	$X_{ABC} \geqslant 0$		

From the definition of the conditional max-entropy (Definition 2.10) and Uhlmann's theorem [Uhl76] it is clear that the optimal value of the primal problem is $2^{H_{\max}(A|B)_{\rho|\sigma}}$. One can also easily show that strong duality holds (i.e., that the optimal value of the dual problem is equal to that of the primal problem). One simply needs to show that there exists a Z_{AB} such that $Z_{AB} \otimes \mathbb{1}_C > \rho_{ABC}$, which holds for $Z_{AB} = 2 \cdot \mathbb{1}_{AB}$.

Now, we need to show that the optimal Z_{AB} for this problem has the form given in the lemma statement. First, note that by Uhlmann's theorem [Uhl76], there must exist an optimal X_{ABC} which has rank 1, assuming we consider the system C to be large enough.

Let $X_{ABC} = |\varphi\rangle\langle\varphi|_{ABC}$ and let $\rho_{ABC} = |\rho\rangle\langle\rho|_{ABC}$, and consider the complementary slackness condition for X and Z to be optimal: $\rho_{ABC}X_{ABC} = (Z_{AB} \otimes \mathbb{1}_C)X_{ABC}$. We can rewrite this as

$$\langle\rho|\varphi\rangle|\rho\rangle\langle\varphi| = (Z_{AB} \otimes \mathbb{1}_C)|\varphi\rangle\langle\varphi|, \tag{155}$$

and therefore

$$\langle\rho|\varphi\rangle|\rho\rangle = (Z_{AB} \otimes \mathbb{1}_C)|\varphi\rangle, \tag{156}$$

as well as

$$F(\rho,\varphi)^2|\rho\rangle\langle\rho| = (Z_{AB} \otimes \mathbb{1}_C)|\varphi\rangle\langle\varphi|(Z_{AB} \otimes \mathbb{1}_C). \tag{157}$$

Tracing out C and using the fact that $F(\rho,\varphi)^2 = 2^{H_{\max}(A|B)_{\rho|\sigma}}$, we get

$$2^{H_{\max}(A|B)_{\rho|\sigma}} \cdot \rho_{AB} = Z_{AB}(\mathbb{1}_A \otimes \sigma_B)Z_{AB}. \tag{158}$$

Now, conjugating both sides by $\sigma_B^{1/2}$ and taking square roots on both sides, we get that

$$2^{\frac{1}{2}H_{\max}(A|B)_{\rho|\sigma}} \cdot \sqrt{\sigma_B^{1/2}\rho_{AB}\sigma_B^{1/2}} = \sigma_B^{1/2}Z_{AB}\sigma_B^{1/2}. \tag{159}$$

If σ_B has full rank, we get the expression for Z_{AB} by conjugating both sides by $\sigma_B^{-1/2}$. Finally, the fact that $\mathrm{Tr}[Z_{AB}\sigma_B] = 2^{H_{\max}(A|B)_{\rho|\sigma}}$ can simply be computed from the expression for Z. □

Lemma B.3. *Let $\rho_{AB} \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB})$ and $\sigma_A \in \mathcal{S}_{\leqslant}(\mathcal{H}_A)$. Then, there exists $T_A \in \mathcal{L}(\mathcal{H}_A)$ with*

$$\sigma_{AB} = (T_A \otimes \mathbb{1}_B)\rho_{AB}(T_A^\dagger \otimes \mathbb{1}_B) \in \mathcal{S}_{\leqslant}(\mathcal{H}_{AB}) \tag{160}$$

an extension of σ_A such that $P(\rho_{AB},\sigma_{AB}) = P(\rho_A,\sigma_A)$.

Proof. Define $X_A = \sigma_A^{\frac{1}{2}}\rho_A^{\frac{1}{2}}$ and polar decompose $X_A = V_A(X_A^\dagger X_A)^{1/2}$. Furthermore define $T_A = \sigma_A^{\frac{1}{2}}V_A\rho_A^{-\frac{1}{2}}$, where the inverse is a generalized inverse.[9] We have

$$\mathrm{Tr}_B((T_A \otimes \mathbb{1}_B)\rho_{AB}(T_A^\dagger \otimes \mathbb{1}_B)) = T_A\rho_A T_A^\dagger = \sigma_A^{\frac{1}{2}}V_A V_A^\dagger\sigma_A^{\frac{1}{2}} = \sigma_A, \tag{161}$$

which shows that $\sigma_{AB} = (T_A \otimes \mathbb{1}_B)\rho_{AB}(T_A^\dagger \otimes \mathbb{1}_B)$ is an extension of σ_A. Thus it remains to prove that $P(\rho_{AB},\sigma_{AB}) = P(\rho_A,\sigma_A)$.

For this we first assume that ρ_{AB} is pure and normalized, i.e., $\rho_{AB} = |\rho\rangle\langle\rho|_{AB} \in \mathcal{S}_{=}(\mathcal{H}_{AB})$. Then, we have that

$$\begin{aligned} P(\rho_{AB},\sigma_{AB}) &= \sqrt{1 - |\langle\rho|\sigma\rangle|^2} \\ &= \sqrt{1 - |\mathrm{Tr}\left[(T_A \otimes \mathbb{1}_B)\rho_{AB}\right]|^2} \\ &= \sqrt{1 - \left|\mathrm{Tr}\left[(\sigma_A^{1/2}V_A\rho_A^{-1/2} \otimes \mathbb{1}_B)\rho_{AB}\right]\right|^2} \end{aligned}$$

[9] For $M \in \mathcal{P}$, M^{-1} is a generalized inverse of M if $MM^{-1} = M^{-1}M = \mathrm{supp}(M) = \mathrm{supp}(M^{-1})$, where $\mathrm{supp}(\cdot)$ denotes the support.

$$
\begin{aligned}
&= \sqrt{1 - \left|\mathrm{Tr}\left[\sigma_A^{1/2} V_A \rho_A^{1/2}\right]\right|^2} \\
&= \sqrt{1 - \left|\mathrm{Tr}\left[\rho_A^{1/2} \sigma_A^{1/2} V_A\right]\right|^2} \\
&= \sqrt{1 - \left|\mathrm{Tr}\left[\sqrt{\rho_A^{1/2} \sigma_A \rho_A^{1/2}}\right]\right|^2} \\
&= \sqrt{1 - F^2(\rho_A, \sigma_A)} \\
&= P(\rho_A, \sigma_A). \qquad (162)
\end{aligned}
$$

If $\rho_{AB} = |\rho\rangle\langle\rho|_{AB}$ is not normalized we obtain analogously

$$
\begin{aligned}
P(\rho_{AB}, \sigma_{AB}) &= \sqrt{1 - [F(\rho_{AB}, \sigma_{AB}) + \sqrt{(1 - \mathrm{Tr}[\rho_{AB}])(1 - \mathrm{Tr}[\sigma_{AB}])}]^2} \\
&= \sqrt{1 - \left(F(\rho_A, \sigma_A) + \sqrt{(1 - \mathrm{Tr}[\rho_A])(1 - \mathrm{Tr}[\sigma_A])}\right)^2} \\
&= P(\rho_A, \sigma_A). \qquad (163)
\end{aligned}
$$

The statement for a general ρ_{AB} (not necessarily pure) follows by the monotonicity of the purified distance [TCR10, Lemma 7] under partial trace. □

References

[Abe13] Aberg, J.: Truly work-like work extraction via a single-shot analysis. Nat. Commun. **4**, 1925 (2013)

[ADHW09] Abeyesinghe, A., Devetak, I., Hayden, P., Winter, A.: The mother of all protocols: restructuring quantum information's family tree. Proc. Roy. Soc. A **465**, 2537 (2009)

[BBCM95] Bennett, C.H., Brassard, G., Crépeau, C., Maurer, U.: Generalized privacy amplification. IEEE Trans. Info. Theory **41**, 1915 (1995)

[BCR11] Berta, M., Christandl, M., Renner, R.: The quantum reverse Shannon theorem based on one-shot information theory. Commun. Math. Phys. **306**, 579 (2011)

[BDH+09] Bennett Charles, H., Devetak, I., Harrow, A.W., Shor, P.W., Winter, A.: Quantum reverse Shannon theorem. (2009). arXiv:0912.5537v2

[Ber08] Berta, M.: Single-shot quantum state merging. Diploma Thesis, ETH Zurich, (2008). arXiv:0912.4495v1

[BH13] Brandao, F.G.S.L., Horodecki, M.: An area law for entanglement from exponential decay of correlations. Nat. Phys. **9**, 721 (2013)

[BP07] Braunstein, S.L., Pati, A.K.: Quantum information cannot be completely hidden in correlations: implications for the black-hole information paradox. Phys. Rev. Lett. **98**, 080502 (2007)

[BRW07] Berta, M., Renner, R., Winter, A.: Tightness of decoupling by projective measurements. Unpublished manuscript; the technical proof appeared as part of [Ber08] (2007)

[BSST02] Bennett, C.H., Shor, P.W., Smolin, J.A., Thapliyal, A.V.: Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. IEEE Trans. Info. Theory **48**, 2637 (2002)

[Bus09] Buscemi, F.: Private quantum decoupling and secure disposal of information. New J. Phys. **11**, 123002 (2009)

[Cho75] Choi, M.-D.: Completely positive linear maps on complex matrices. Linear Algebra Appl. **10**, 285 (1975)

[CS06] Collins , B., Śniady, P.: Integration with respect to the haar measure on unitary, orthogonal and symplectic group. Commun. Math. Phys. **264**, 773 (2006)

[Dat09] Datta, N.: Min- and max- relative entropies and a new entanglement monotone. IEEE Trans. Info. Theory **55**, 2816 (2009)

[Dev05] Devetak, I.: The private classical capacity and quantum capacity of a quantum channel. IEEE Trans. Info. Theory **51**, 44 (2005)

[dRAR+11] Lídia, del R., Åberg, J., Renner, R., Dahlsten, O., Vedral, V. The thermodynamic meaning of negative entropy. Nature, **474**, 61 (2011)
[DRRV09] Dahlsten, O., Renner, R., Rieper, E., Vedral, V.: Inadequacy of von Neumann entropy for characterizing extractable work. New J. Phys. **13**, 053015 (2009)
[Dup09] Dupuis, F.: The decoupling approach to quantum information theory. PhD thesis, Université de Montréal, (2009). arXiv:1004.1641v1
[FDOR12] Faist, P., Dupuis, F., Oppenheim, J., Renner, R.: A quantitative Landauer's principle (2012). arXiv:1211.1037v1
[GPW05] Groisman, B., Popescu, S., Winter, A.: Quantum, classical, and total amount of correlations in quantum state. Phys. Rev. A **72**, 032317 (2005)
[HHWY08] Hayden, P., Horodecki, M., Winter, A., Yard, J.: A decoupling approach to the quantum capacity. Open Syst. Info. Dynam. **15**, 7 (2008)
[HO13] Horodecki, M., Oppenheim, J.: Fundament limitations for quantum and nanoscale thermodynamics. Nat. Commun. **4**, 2059 (2013)
[HOW05] Horodecki, M., Oppenheim, J., Winter, A.: Partial quantum information. Nature, **436**, 673 (2005)
[HOW07] Horodecki, M., Oppenheim, J., Winter, A.: Quantum state merging and negative information. Commun. Math. Phys. **269**, 107 (2007)
[HP07] Hayden, P., Preskill, J.: Black holes as mirrors: quantum information in random subsystems. J. High Energy Phys. **07**, 120 (2007)
[Hut11] Hutter, A.: Understanding thermalization from decoupling. Master Thesis, ETH Zurich, (2011). http://www.quantumlah.org/media/thesis/NCQT_AdrianHutter_MSc2011.pdf
[Jam72] Jamiołkowski, A.: Linear transformations which preserve trace and positive semidefiniteness of operators. Reports Math. Phys. **3**, 275 (1972)
[KRS09] König, R., Renner, R., Schaffner, C.: The operational meaning of min- and max-entropy. IEEE Trans. Info. Theory **55**, 4337 (2009)
[Llo97] Lloyd, S.: Capacity of the noisy quantum channel. Phys. Rev. A **55**, 1613 (1997)
[LPSW09] Linden, N., Popescu, S., Short, A.J., Winter, A.: Quantum mechanical evolution towards thermal equilibrium. Phys. Rev. E **79**, 061103 (2009)
[Par89a] Partovi, M.H.: Irreversibility, reduction, and entropy increase in quantum measurements. Phys. Lett. A **137**, 445 (1989)
[Par89b] Partovi, M.H.: Quantum thermodynamics. Phys. Lett. A **137**, 440 (1989)
[PZ13] Braunstein Stefano Pirandola, S.L., Zyczkowski, K.: Better late than never: information retrieval from black holes. Phys. Rev. Lett. **110**, 101301 (2013)
[Ren05] Renner, R.: Security of quantum key distribution. PhD thesis, ETH Zurich, (2005). http://www.worldscientific.com/doi/abs/10.1142/S0219749908003256
[Ren09] Renner, R.: Optimal decoupling. Proc. Intern. Congr. Math. Phys, p. 541, (2009)
[RK05] Renner, R., Robert, K.: Universally composable privacy amplification against quantum adversaries. In: Second Theory of Cryptography Conference TCC, vol. 3378 of Lecture Notes in Computer Science, p. 407. Springer, (2005)
[RW04] Renner, R., Stefan, W.: Smooth Rényi entropy and applications. In: Proceedings International Symposium on Information Theory, p. 233, (2004)
[Sho02] Peter, S.: The quantum channel capacity and coherent information. Lecture notes, MSRI workshop on quantum computation, (2002). http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/
[Sti55] Stinespring, W.F.: Positive function on C*-algebras. Proc. Amer. Math. Soc. **6**, 211 (1955)
[TCR09] Marco, T., Roger, C., Renner, R.: A fully quantum asymptotic equipartition property. IEEE Trans. Info. Theory **55**, 5840 (2009)
[TCR10] Marco, T., Roger, C., Renner, R.: Duality between smooth min- and max-entropies. IEEE Trans. Info. Theory **56**, 4674 (2010)
[Tom12] Marco, T.: A framework for non-asymptotic quantum information theory. PhD thesis, ETH Zurich, (2012). arXiv:1203.2142v2
[TRSS10] Marco, T., Renner, R., Christian, S., Adam, S.: Leftover hashing against quantum side information. In: Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, p. 2703, (2010)
[Uhl76] Uhlmann, A.: The 'transition probability' in the state space of a *-algebra. Reports Math. Phys. **9**, 273 (1976)
[Wat08] John, W.: Theory of quantum information—Lecture notes from fall 2008. (2008). http://www.cs.uwaterloo.ca/~watrous/quant-info/

Communicated by M. B. Ruskai