

# Catalytic Decoupling of Quantum Information

Christian Majenz,<sup>1,\*</sup> Mario Berta,<sup>2</sup> Frédéric Dupuis,<sup>3</sup> Renato Renner,<sup>4</sup> and Matthias Christandl<sup>1</sup>

<sup>1</sup>Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø.

<sup>2</sup>Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA 91125, USA.

<sup>3</sup>Faculty of Informatics, Masaryk University, Brno, Czech Republic.

<sup>4</sup>Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland.

(Dated: May 3, 2016)

The decoupling technique is a fundamental tool in quantum information theory with applications ranging from quantum thermodynamics to quantum many body physics to the study of black hole radiation. In this work we introduce the notion of catalytic decoupling, that is, decoupling in the presence of an uncorrelated ancilla system. This removes a restriction on the standard notion of decoupling, which becomes important for structureless resources, and yields a tight characterization in terms of the max-mutual information. Catalytic decoupling naturally unifies various tasks like the erasure of correlations and quantum state merging, and leads to a resource theory of decoupling.

*Introduction.* Erasing correlations between quantum systems via local operations, decoupling, is a task that was first studied in the context of quantum information theory [1] (see [2] for an introductory tutorial). In particular, decoupling has been crucial for understanding how to distribute quantum information between different parties [3–7] and for understanding how to send quantum information over noisy quantum channels [8–11]. In that context, the idea of decoupling has also been made use of in quantum cryptography [12]. The concept is, however, also very useful in physics (as, e.g., outlined in [13]). Applications range from quantum thermodynamics [14–18], to the study of black hole radiation [19–21], and solid state physics [22].

*Standard decoupling.* The basic idea behind decoupling is the following: if a mixed bipartite quantum state  $\rho_{AE}$  is only weakly correlated, then it should suffice to erase a small part of  $A$  to approximately decouple  $A$  from  $E$ , i.e., to get an approximate product state (see Figure 1). More precisely, we say that a bipartite quantum state  $\rho_{AE}$  is  $\varepsilon$ -decoupled by the partial trace map  $T_{A \rightarrow A_1}(\cdot) = \text{Tr}_{A_2}[\cdot]$  with  $A = A_1 A_2$  if there exists an unitary operation  $U_A$  such that,

$$\min_{\omega_{A_1} \otimes \omega_E} P(\mathcal{T}_{A \rightarrow A_1}(U_A \rho_{AE} U_A^\dagger), \omega_{A_1} \otimes \omega_E) \leq \varepsilon, \quad (1)$$

where the minimum is over all product quantum states  $\omega_{A_1} \otimes \omega_E$ , and  $P(\beta, \gamma) := (1 - \|\sqrt{\beta} \sqrt{\gamma}\|_1^2)^{1/2}$  denotes the purified distance [23]. The  $A_1$ -system is called the decoupled system and the  $A_2$ -system the remainder system. Now, the fundamental question that we want to discuss is how large we have to choose the remainder system  $A_2$  in order to achieve  $\varepsilon$ -decoupling. We denote the minimal remainder system size, i.e., the logarithm of the minimal remainder system dimension, for  $\varepsilon$ -decoupling  $A$  from  $E$  in a state  $\rho_{AE}$  by  $R^\varepsilon(A; E)_\rho$ .

*Converse.* We first show quite naturally that  $R^\varepsilon(A; E)_\rho$  has to be at least of the size of the smooth

max-mutual information  $I_{\max}^\varepsilon(E; A)_\rho$  present in the initial state  $\rho_{AE}$ . This measure is defined as [11],

$$I_{\max}^\varepsilon(E; A)_\rho := \min_{\bar{\rho}} I_{\max}(E; A)_{\bar{\rho}} \quad \text{with} \quad (2)$$

$$I_{\max}(E; A)_{\bar{\rho}} := \min_{\sigma_A} \min \left\{ \lambda \in \mathbb{R} \mid 2^\lambda \cdot \sigma_A \otimes \bar{\rho}_E \geq \bar{\rho}_{AE} \right\}, \quad (3)$$

where the minimum in (2) is over all bipartite quantum states with  $P(\rho_{AE}, \bar{\rho}_{AE}) \leq \varepsilon$  [24], and the minimum in (3) is over all quantum states  $\sigma_A$ . We note that the definition of the smooth max-information is a priori not symmetric in  $A : E$ . However, we have [25],

$$I_{\max}^\varepsilon(E; A)_\rho =_\varepsilon I_{\max}^\varepsilon(A; E)_\rho, \quad (4)$$

where  $=_\varepsilon$  stands for equality up to terms  $\mathcal{O}(\log(1/\varepsilon))$ . For the converse we exploit that the smooth max-mutual information is invariant under local unitary operations and that it has the so-called non-locking property (see [26] about information locking). That is, just like the quantum mutual information it fulfills the inequality [11, Lemma B.12],

$$I_{\max}^\varepsilon(E; A_1 A_2)_\rho \leq I_{\max}^\varepsilon(E; A_1)_\rho + 2 \log |A_2|, \quad (5)$$

where  $|A_2|$  denotes the dimension of  $A_2$ . Since the final state is a product state, its smooth max-mutual information  $I_{\max}^\varepsilon(E; A_1)_{\omega \otimes \omega}$  becomes zero. This means that in order to erase the initial correlations  $I_{\max}^\varepsilon(E; A)_\rho$  we need at least a remainder system of size [27],

$$R^\varepsilon(A; E)_\rho \geq \frac{1}{2} I_{\max}^\varepsilon(E; A)_\rho. \quad (6)$$

*Previous works.* Most of the aforementioned decoupling references only give good achievability bounds for states of the form  $\rho_{A^n E^n} = \rho_{AE}^{\otimes n}$  in the asymptotic limit  $n \rightarrow \infty$ . Whereas this setting is relevant in quantum Shannon theory, it is often a severe restriction for

applications in physics. For typical physical situations (e.g., in thermodynamics), there is usually not even a natural decomposition of a large system in  $n$  subsystems. A notable exception concerning achievability results is reference [13], where the authors show that

$$R^\varepsilon(A; E)_\rho \leq_\varepsilon \frac{1}{2} \left( H_{\max}^{\varepsilon'}(A)_\rho - H_{\min}^{\varepsilon'}(A|E)_\rho \right) \text{ with } \varepsilon' = \frac{\varepsilon}{5}, \quad (7)$$

where  $\leq_\varepsilon$  means up to terms  $\mathcal{O}(\log(1/\varepsilon))$ . (We give a proof of this particular statement in the supplemental material). Here,  $H_{\max}^\varepsilon$  and  $H_{\min}^\varepsilon$  denote the smooth conditional max- and min-entropy whose exact definitions can be found in the supplemental material (or see the textbook [23]). In fact, the results from [13] also show that not only decoupling in the sense of (1) is achieved, but moreover that the decoupled system is also randomized. That is, there exists a quantum state  $\omega_E$  and a unitary operation  $U_A$  such that the decoupled system is left in the fully mixed state:

$$P \left( \mathcal{T}_{A \rightarrow A_1} (U_A \rho_{AE} U_A^\dagger), \frac{1_{A_1}}{|A_1|} \otimes \omega_E \right) \leq \varepsilon. \quad (8)$$

However, it turns out that there can be an arbitrary big gap between the converse (6) and the achievability result (7). This is best seen for an example with trivial system  $E$ . In that case the achievability bound (7) reduces to the difference between the smooth max- and min-entropy and it is known that this can become roughly as big as  $\log |A|$  (we provide an explicit example in the supplemental material). In order to achieve the converse from (6) we propose in the following a generalized notion of decoupling.

*Catalytic decoupling.* A natural question to ask at this point is if decoupling can be achieved more efficiently in the presence of an already uncorrelated ancilla system (see Figure 1). Formally, we say that  $\varepsilon$ -decoupling can be achieved catalytically for a bipartite quantum state  $\rho_{AE}$  if there exists an ancilla state  $\rho_{A'}$  and a decomposition  $AA' \cong A_1 A_2$  such that

$$\min_{\omega_{A_1} \otimes \omega_E} P(\rho_{A_1 E}, \omega_{A_1} \otimes \omega_E) \leq \varepsilon \quad \text{where} \quad (9)$$

$$\rho_{A_1 A_2 E} = \rho_{AA'E} = \rho_{AE} \otimes \rho_{A'}. \quad (10)$$

Again, we call the  $A_1$ -system the decoupled system and the  $A_2$ -system the remainder system. The term catalytic means that the share of the initially uncorrelated ancilla system  $A'$  that becomes part of the decoupled system  $A_1$  stays decoupled (see Figure 1).

Now, we are interested in the minimal size of the remainder system  $A_2$  in order to achieve  $\varepsilon$ -decoupling catalytically. We denote the optimal remainder system size for catalytically decoupling  $A$  from  $E$  in a

state  $\rho_{AE}$  by  $R_c^\varepsilon(A; E)_\rho$ . Clearly, we have  $R_c^\varepsilon(A; E)_\rho \leq R^\varepsilon(A; E)_\rho$ , as we can always choose a trivial ancilla. Moreover, since appending with an ancilla does not increase the smooth max-mutual information (see supplemental material), the same converse as in (6) still holds.

These concepts can naturally be phrased as a resource theory of decoupling. A quantum system  $A$  coupled to the environment  $E$  can yield a decoupled system  $A_1$  of a certain size through standard decoupling. That is, in the resource theory language of [28] we have  $\langle \rho_{AE} \rangle \geq_\varepsilon (\log |A| - R^\varepsilon(A; E)_\rho)[d]$ . Here,  $[d]$  denotes a decoupled qbit and  $\geq_\varepsilon$  stands for up to error  $\varepsilon$  (see also [29]). Now, our novel paradigm makes use of the possibility that if we already have decoupled qbits at hand, then we might be able to decouple a larger system,

$$\langle \rho_{AE} \rangle + n[d] \geq_\varepsilon (n + \log |A| - R_c^\varepsilon(A; E)_\rho)[d] \quad \text{for } n \text{ large enough.} \quad (11)$$

*Achievability.* In contrast to standard decoupling as in (1), catalytic decoupling can be achieved with a remainder system size that is essentially equal to the smooth max-mutual information.

**Theorem 1** (Catalytic decoupling). *For any bipartite quantum state  $\rho_{AE}$  and  $0 < \delta \leq \varepsilon \leq 1$  we have:*

$$R_c^\varepsilon(A; E)_\rho \lesssim \frac{1}{2} I_{\max}^{\varepsilon-\delta}(E; A)_\rho \quad (12)$$

where  $\lesssim$  stands for smaller or equal up to terms  $\mathcal{O}(\log \log |A| + \log(1/\delta))$ . We also have the converse

$$R_c^\varepsilon(A; E)_\rho \geq \frac{1}{2} I_{\max}^\varepsilon(E; A)_\rho. \quad (13)$$

In fact, we not only show that catalytic decoupling in the sense of (9) is achieved, but moreover that the decoupled system ends up in the marginal of the original state:

$$P(\rho_{A_1 E}, \rho_{A_1} \otimes \omega_E) \leq \varepsilon \quad \text{for some quantum state } \omega_E. \quad (14)$$

In particular, and in contrast to the standard decoupling results leading to (7), our catalytic decoupling scheme does not randomize the decoupled system but leaves it invariant (up to the approximation error  $\varepsilon$ ). We can even choose  $A_1 = AA'_1$  such that the decoupled system contains the marginal of the input state (plus part of the catalyst).

In the supplemental material we give two conceptually different proofs for Theorem 1. The first proof is based on the standard decoupling techniques from [11, 13] combined with the use of embezzling entangled quantum states [30]. For (12) this yields a difference of

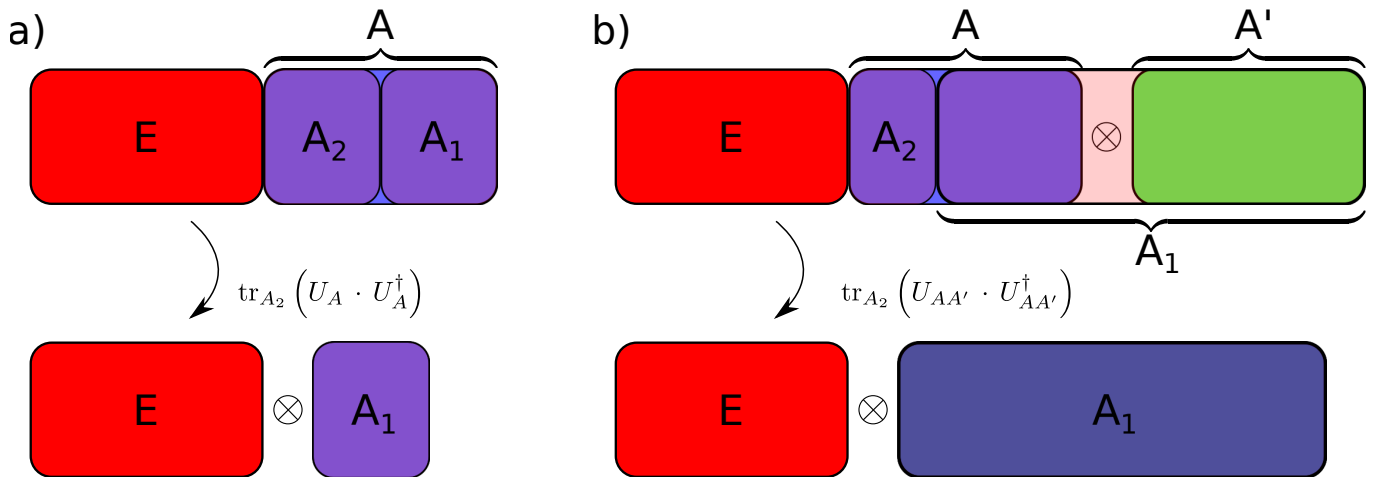


FIG. 1. Schematic representation of a) standard and b) catalytic decoupling: tracing out a system  $A_2$  leaves the remaining state decoupled. While there is no ancilla for standard decoupling as in a), catalytic decoupling as in b) allows to make use of an additional, already decoupled system  $A'$ . The basic question is how large we have to choose the system  $A_2$  such that the remaining system  $A_1$  is decoupled from  $E$ .

size at most  $\log \log |A| + \mathcal{O}(\log(1/\delta))$  [31]. The second proof is based on the convex splitting technique of Anshu *et al.* [32]. It allows to upper bound the difference in (12) with the tighter bound

$$\frac{1}{2} I_{\max}^{\varepsilon-\delta}(E; A)_\rho - R_c^\varepsilon(A; E)_\rho \leq \frac{1}{2} \left\{ \log \log I_{\max}^{\varepsilon-\delta}(E; A)_\rho \right\}_+ + \mathcal{O}(\log(1/\delta)), \quad (15)$$

where  $\{\cdot\}_+ := \max\{0, \cdot\}$ . Moreover, this argument is also constructive and hence leads to an explicit scheme for decoupling. This improves on the standard decoupling bounds which are achieved using the probabilistic technique [33] (as, e.g., the previously best known bound (7) from [13]).

*Discussion.* The achievability result (12) together with the converse (13) establish an operational interpretation of the smooth max-information as twice the minimal size of the remainder system to achieve  $\varepsilon$ -decoupling. We note that the approximation error as well as the smoothing parameter can be made arbitrarily close in (13) and (12) with only a logarithmic penalty. Following the information-theoretic arguments outlined in [34], we find that for states of the form  $\rho_{A^n E^n} = \rho_{AE}^{\otimes n}$  and large  $n \rightarrow \infty$ ,

$$\frac{1}{n} R_c^\varepsilon(A^n; E^n)_{\rho^{\otimes n}} = \frac{1}{2} \left( I(A : E)_\rho + \sqrt{\frac{V(A : E)_\rho}{n}} \Phi^{-1}(\varepsilon) \right) + \mathcal{O}\left(\frac{\log n}{n}\right), \quad (16)$$

with the mutual information  $I(A : E)_\rho = H(A)_\rho + H(E)_\rho - H(AE)_\rho$  featuring the von Neumann entropy

$H(A)_\rho = -\text{Tr}(\rho_A \log \rho_A)$ , and the mutual information variance  $V(A : E)_\rho$ , as well as the cumulative normal distribution function  $\Phi$  specified in the supplemental material. We note that no such tight (second-order) asymptotic expansion is known for standard decoupling. However, the achievability (7) together with the converse (6) imply that (using the asymptotic equipartition property from [23]),

$$\lim_{n \rightarrow \infty} \frac{1}{n} R_c^\varepsilon(A^n; E^n)_{\rho^{\otimes n}} = \frac{1}{2} I(A : E)_\rho. \quad (17)$$

Thus, we can conclude that catalytic decoupling and standard decoupling become equivalent in the first order rate asymptotically: the mutual information quantifies the minimal size of the remainder system.

*Applications.* Groisman *et al.* [35] introduced an operational approach to quantifying the total correlations that are present in a quantum state. In analogy to Landauer's erasure principle [36], they characterize the strength of correlations by the amount of randomness that has to be injected locally to decorrelate the state. This randomizing is done by a random-unitary channel on one of the systems (called local unitary randomizing, A-LUR in [35]):

$$\Lambda(\cdot) = \sum_{i=1}^N p_i U_i(\cdot) U_i^\dagger. \quad (18)$$

We say that the correlations between  $A$  and  $E$  in a state  $\rho_{AE}$  can be  $\varepsilon$ -erased by a local mixture of  $N$  unitaries on  $A$  up to an error  $\varepsilon$ , if  $\Lambda_A$   $\varepsilon$ -decouples  $A$  from  $E$ . That is, if there exists a quantum channel  $\Lambda_A$  of the

form (18) such that

$$\min_{\omega_A \otimes \omega_E} P(\Lambda_A(\varrho_{AE}), \omega_A \otimes \omega_E) \leq \varepsilon. \quad (19)$$

We denote the minimal number of unitaries needed for  $\varepsilon$ -erasing the correlations between  $A$  and  $E$  in a state  $\varrho_{AE}$  by  $R_U^\varepsilon(A; E)_\varrho$ . Groisman *et al.* show that for states of the form  $\varrho_{A^n E^n} = \varrho_{AE}^{\otimes n}$  for large  $n \rightarrow \infty$ :

$$\lim_{n \rightarrow \infty} \frac{1}{n} R_U^\varepsilon(A^n; E^n)_{\varrho^{\otimes n}} = I(A : E)_\varrho. \quad (20)$$

In the following we will see that the task of *catalytic* local erasure of correlation becomes equivalent to catalytic decoupling. We therefore define  $R_{U,c}^\varepsilon(A; E)_\varrho := \inf R_U^\varepsilon(AA' : E)_{\varrho \otimes \sigma}$ , where the infimum is taken over all ancilla systems.

**Proposition 2** (Erasure of correlations). *For any bipartite quantum state  $\varrho_{AE}$  we have  $\frac{1}{2} R_{U,c}^\varepsilon(A; E)_\varrho = R_c^\varepsilon(A; E)_\varrho$ . Hence, we get (with the notation from Theorem 1),*

$$I_{\max}^\varepsilon(E; A)_\varrho \leq R_{U,c}^\varepsilon(A; E)_\varrho \lesssim I_{\max}^{\varepsilon-\delta}(E; A)_\varrho. \quad (21)$$

The same asymptotic expansion as in (16) holds.

This is the generalization of the results in [35] to arbitrary (structureless) states. It gives an alternative operational characterization of the smooth max-mutual information as the minimal number of unitaries needed for  $\varepsilon$ -erasing the correlations between  $A$  and  $E$ . The proof of Proposition 2 proceeds as follows. Suppose we have a way of decoupling  $A$  from  $E$  with remainder system  $A_2$ , and let  $|A_2| = 2^k$  for some  $k \in \mathbb{N}$ . Then, we can think of  $A_2$  as  $k$  qbits and erase each of them applying a uniform mixture of the Pauli matrices and the identity. This is a mixture of  $4^k = 2^{2k}$  unitaries. Conversely, suppose we have a uniform mixture of  $N = 2^{2k}$  unitaries on  $A$  that erase the correlations to  $E$ . We take the fully mixed ancilla state  $1_{A'_1 A'_2} / |A'_1 A'_2|$  with  $A'_i \cong \mathbb{C}^{2^k}$ . Now, we apply the unitaries controlled on an orthonormal basis of maximally entangled states of  $A'_1 A'_2$ . Then,  $A'_1 A$  are decoupled from  $E$ , i.e., we achieved catalytic decoupling with remainder system size  $\log |A'_2| = k$ .

As a second application we discuss quantum state merging [1] in whose context decoupling was originally introduced [3, 4]. Any catalytic decoupling theorem naturally leads to a quantum state merging protocol. Since the catalytic decoupling theorem is the abstraction of the work on quantum state merging in [11, 32], inserting the bounds from Theorem 1, we recover the following optimal result for the communication cost of quantum state merging.

**Proposition 3** (Coherent quantum state merging). *Let  $\varrho_{ABR}$  be a pure tripartite quantum state shared between Alice, Bob and a Referee. If Alice and Bob have arbitrary entanglement assistance at hand, then Alice can send her system  $A$  to Bob up to error  $\varepsilon > 0$  in purified distance using*

$$q^\varepsilon(A)B)_\varrho \lesssim \frac{1}{2} I_{\max}^{\varepsilon/3}(R; A)_\varrho \quad (22)$$

qbits of quantum communication (with the same notation as in Theorem 1).

We note that in the asymptotic limit standard decoupling is sufficient to obtain,

$$\lim_{n \rightarrow \infty} \frac{1}{n} q^\varepsilon(A^n)B^n)_{\varrho^{\otimes n}} = \frac{1}{2} I(R : A)_\varrho, \quad (23)$$

which is also optimal [4]. However, for the general setup there is an issue known as entanglement spread [37], and for the proof of Proposition 3 we make use of catalytic decoupling and Uhlmann's theorem [38]. In the following we present a proof sketch but defer the full argument to the supplemental material. Setting  $\delta = \varepsilon/6$  in Theorem 1 shows that there exists an ancilla state  $\varrho_{A'}$  and a unitary  $U_{AA' \rightarrow A_1 A_2}$  such that  $A_1$  is  $\varepsilon/2$  decoupled from  $R$  and

$$\log |A_2| \lesssim \frac{1}{2} I_{\max}^{\varepsilon/3}(R : A)_\varrho \quad (24)$$

Now, Alice and Bob take a pure entangled state  $\varrho_{A'B'}$  where Alice's part  $A'$  is in state  $\varrho_{A'}$ . She applies the unitary  $U_{AA' \rightarrow A_1 A_2}$  and sends  $A_2$  to Bob. The decoupling condition and the triangle inequality for the purified distance imply that  $P(\varrho_{A_1 R}, \varrho_{A_1} \otimes \varrho_R) \leq \varepsilon$ , so by Uhlmann's theorem there exists a unitary  $U_{A_2 B \rightarrow A B B_1}$  such that

$$P(U \varrho_{A_1 A_2 R} U^\dagger, \varrho_{A_1 B_1} \otimes \varrho_{ABR}) \leq \varepsilon, \quad (25)$$

where  $\varrho_{A_1 B_1}$  is a purification of  $\varrho_{A_1}$  and we omitted the subscript of  $U$ . This implies that Bob has systems  $AB$  after applying  $U$ .

Finally, we also show in the supplemental material that catalytic decoupling directly implies the achievability bound for quantum state redistribution of Anshu *et al.* [32] (see [39, 40] for alternative bounds).

*Extensions.* So far we have analyzed how well the partial trace map  $T_{A \rightarrow A_1}(\cdot) = \text{Tr}_{A_2}[\cdot]$  decouples. However, as originally suggested in [13], we can also study quantum channels  $T_{A \rightarrow B}(\cdot)$  that add noise in an arbitrary way in order to achieve decoupling. To further clarify the important difference between standard decoupling and catalytic decoupling, as well as to correct the faulty [13, Corollary 4.2], we now give a converse for the decoupling behavior of general quantum channels.

**Proposition 4** (Correction of Corollary 4.2 from [13]). *If for a bipartite quantum state  $\rho_{AE}$  and a quantum channel  $\mathcal{T}_{A \rightarrow B}$ ,*

$$\int dU_A P \left( \mathcal{T}_{A \rightarrow B}(U_A \rho_{AE} U_A^\dagger), T_{A \rightarrow B} \left( \frac{1_A}{|A|} \right) \otimes \rho_E \right) \leq \varepsilon, \quad (26)$$

then we have

$$H_{\min}^{\varepsilon'}(A|E)_\rho + H_{\max}^\varepsilon(A'|B)_\tau \gtrsim 0 \quad \text{with } \varepsilon' = 15\sqrt{\varepsilon}, \quad (27)$$

where  $\tau_{A'B} = \mathcal{T}_{A \rightarrow B}(\phi_{A'A}^+)$  is the Choi-Jamiołkowski state.

In the supplemental material we prove Proposition 4 starting from [13, Theorem 4.1] (from which also the faulty [13, Corollary 4.2] was derived). The crucial difference of Proposition 4 to the erroneous version is the assumption that not only decoupling, but decoupling and randomizing is achieved:

$$T_{A \rightarrow B}(\rho_A) \otimes \rho_E \quad \text{vs.} \quad T_{A \rightarrow B} \left( \frac{1_A}{|A|} \right) \otimes \rho_E. \quad (28)$$

For example, a product state  $\rho_{AE} = \rho_A \otimes \rho_E$  with  $\rho_A$  pure has  $H_{\min}^{\varepsilon'}(A|E)_\rho \approx 0$ . It is, however, already perfectly decoupled by the identity map on  $A$ , which yields  $H_{\max}^\varepsilon(A|B)_\tau \approx -\log |A|$ .

In turn, applying the converse bound (27) to the partial trace map  $T_{A \rightarrow A_1}(\cdot) = \text{Tr}_{A_2}[\cdot]$  shows that the standard decoupling bound (7) in terms of a difference of smooth max- and min-entropy is natural if we ask for decoupling and randomizing. However, if we are not interested in randomizing but only in decoupling, then our main result about catalytic decoupling (Theorem 1) shows that the smooth max-mutual information is the relevant measure.

*Conclusion.* In this work we introduced the notion of catalytic decoupling. As our main result we established that the optimal remainder system size for decoupling is given by one-half times the smooth max-mutual information. In contrast to standard decoupling results our decoupling scheme is explicit and does not randomize the decoupled system. Moreover, we have shown that catalytic decoupling for general (structureless) states naturally quantifies the resources needed in the erasure of correlation model from [35] and for quantum state merging as in [11]. All of this strengthens the smooth max-mutual information as the one-shot generalization of the quantum mutual information. Finally, given that standard decoupling has already proven useful in various areas of physics (see the references in the introduction), we believe that catalytic decoupling has manifold applications that remain to be explored.

*Acknowledgments.* MC and CM acknowledge financial support from the European Research Council (ERC Grant Agreement no 337603), the Danish Council for Independent Research (Sapere Aude) and the Swiss National Science Foundation (project no PP00P2-150734). MB acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028). Additional funding support was provided by the ARO grant for Research on Quantum Algorithms at the IQIM (W911NF-12-1-0521). FD acknowledges the support of the Czech Science Foundation (GA ĆR) project no GA16-22211S and FD and RR acknowledge the support of the EU FP7 under grant agreement no 323970 (RAQUEL).

\* majenz@math.ku.dk

- [1] M. Horodecki, J. Oppenheim, and A. Winter, *Nature* **436**, 673 (2005).
- [2] P. Hayden, Tutorial QIP Singapore (2011).
- [3] M. Horodecki, J. Oppenheim, and A. Winter, *Communications in Mathematical Physics* **269**, 107 (2007).
- [4] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, *Proceedings of the Royal Society A* **465**, 2537 (2009).
- [5] Z. Luo and I. Devetak, *Information Theory, IEEE Transactions on* **55**, 1331 (2010).
- [6] J. T. Yard and I. Devetak, *Information Theory, IEEE Transactions on* **55**, 5339 (2009).
- [7] I. Devetak and J. Yard, *Physical Review Letters* **100**, 230501 (2008).
- [8] P. Hayden, M. Horodecki, A. Winter, and J. Yard, *Open Systems and Information Dynamics* **15**, 7 (2008).
- [9] F. Dupuis, *The decoupling approach to quantum information theory*, Ph.D. thesis, Université de Montréal (2009).
- [10] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, *Information Theory, IEEE Transactions on* **60**, 2926 (2014).
- [11] M. Berta, M. Christandl, and R. Renner, *Communications in Mathematical Physics* **306**, 579 (2011).
- [12] M. Berta, O. Fawzi, and S. Wehner, *Information Theory, IEEE Transactions on* **60**, 1168 (2014).
- [13] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, *Communications in Mathematical Physics* **328**, 251 (2014).
- [14] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, *Nature* **474**, 61 (2011).
- [15] J. Åberg, *Nature Communications* **4**, 1925 (2013).
- [16] A. Hutter, *Understanding Thermalization from Decoupling*, Master's thesis, ETH Zurich (2011).
- [17] R. Chaves and T. Fritz, *Physical Review A* **85**, 032113 (2012).
- [18] F. G. Brandão and M. Horodecki, *Nature Physics* **9**, 721 (2013).

- [19] P. Hayden and J. Preskill, *Journal of High Energy Physics* **07**, 120 (2007).
- [20] S. L. B. S. Pirandola and K. Życzkowski, *Physical Review Letters* **110**, 101301 (2013).
- [21] S. L. Braunstein and A. K. Pati, *Physical Review Letters* **98**, 080502 (2007).
- [22] F. G. Brandão, M. Christandl, and J. Yard, *Communications in Mathematical Physics* **306**, 805 (2011).
- [23] M. Tomamichel, *Quantum Information Processing with Finite Resources — Mathematical Foundations* (Springer International Publishing, 2016).
- [24] More precisely, this minimum is taken over sub-normalized states, see supplemental material.
- [25] N. Ciganovic, N. J. Beaudry, and R. Renner, *Information Theory, IEEE Transactions on* **60**, 1573 (2014).
- [26] D. P. DiVincenzo, M. Horodecki, D. W. Leung, J. A. Smolin, and B. M. Terhal, *Phys. Rev. Lett.* **92**, 067902 (2004).
- [27] For the definition (3), it is convention in the literature to write  $E; A$  (and not  $A; E$ ).
- [28] I. Devetak, A. Harrow, and A. Winter, *IEEE Transactions on Information Theory* **54**, 4587 (2008).
- [29] N. Datta and M.-H. Hsieh, *New Journal of Physics* **13**, 093042 (2011).
- [30] W. van Dam and P. Hayden, *Physical Review A* **67**, 060302 (2003).
- [31] The term  $\log \log |A|$  can be improved to be logarithmic in the smooth max-information, when accepting a slightly worse leading order term.
- [32] A. Anshu, V. K. Devabathini, and R. Jain, preprint arXiv:1410.3031 (2014).
- [33] A partial derandomization can be achieved using (approximate) unitary 2-designs [41].
- [34] M. Tomamichel and M. Hayashi, *Information Theory, IEEE Transactions on* **59**, 7693 (2013).
- [35] B. Groisman, S. Popescu, and A. Winter, *Physical Review A* **72**, 032317 (2005).
- [36] R. Landauer, *IBM Journal of Research and Development* **5**, 183 (1961).
- [37] A. W. Harrow, *Proc. XVI Int. Cong. Math. Phys* **536** (2009).
- [38] A. Uhlmann, *Annals of Physics* **497**, 524 (1985).
- [39] M. Berta, M. Christandl, and D. Touchette, *Information Theory, IEEE Transactions on* **62**, 1425 (2016).
- [40] N. Datta, M.-H. Hsieh, and J. Oppenheim, preprint arXiv:1409.4352 (2014).
- [41] O. Szehr, F. Dupuis, M. Tomamichel, and R. Renner, *New Journal of Physics* **15**, 053022 (2013).
- [42] M. Tomamichel, R. Colbeck, and R. Renner, *Information Theory, IEEE Transactions on* **56**, 4674 (2010).
- [43] It is shown in [42] that the generalized trace distance and the generalized purified distance are metrics.
- [44] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
- [45] R. Renner, "Security of quantum key distribution,".
- [46] R. König, R. Renner, and C. Schaffner, *Information Theory, IEEE Transactions on* **55**, 4337 (2009).
- [47] This is counterintuitive and due to the fact that we only project but do not renormalize the state.
- [48] R. Renner and S. Wolf, in *IEEE International Symposium on Information Theory* (2004) pp. 233–233.
- [49] The original concept was defined using the trace distance instead of the purified distance [30]. We use the purified distance here as it fits our task, the definitions are equivalent up to a square according to Supplemental Lemma 2.
- [50] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner, arXiv preprint arXiv:1205.5231 (2012).
- [51] The limit  $\delta \rightarrow 0$  exists, as the min-entropy term that depends on  $\delta$  is nondecreasing in  $\delta$  and bounded from below.
- [52] H. Umegaki, in *Kodai Mathematical Seminar Reports*, Vol. 14 (Department of Mathematics, Tokyo Institute of Technology, 1962) pp. 59–85.

SUPPLEMENTAL MATERIAL

Additional notation, definitions and lemmas

All Hilbert spaces considered here are finite-dimensional. Given a Hilbert space  $\mathcal{H}$  we denote the set of endomorphisms on this Hilbert space by  $\text{End}(\mathcal{H})$ . The set of normalized quantum states on a Hilbert space  $\mathcal{H}$  is denoted by  $\mathcal{S}(\mathcal{H}) = \{\varrho \in \text{End}(\mathcal{H}) \mid \text{Tr}\varrho = 1, \varrho \geq 0\}$ , the set of sub-normalized quantum states by  $\mathcal{S}_{\leq}(\mathcal{H}) = \{\varrho \in \text{End}(\mathcal{H}) \mid \text{Tr}\varrho \leq 1, \varrho \geq 0\}$ . The identity is denoted by  $\mathbb{1}$ , and the maximally mixed state by  $\tau = \mathbb{1} / \dim \mathcal{H}$ . The unitary group on this Hilbert space is denoted by  $\text{U}(\mathcal{H})$ . We will make use of two matrix norms, the trace norm and the operator norm, defined by

$$\begin{aligned} \|A\|_1 &= \text{Tr}\sqrt{A^\dagger A} \\ \|A\|_\infty &= \max_{|\phi\rangle \in \mathcal{H}} \|A|\phi\rangle\|_2, \end{aligned}$$

for an operator  $A \in \text{End}(\mathcal{H})$ , where  $\| |\phi\rangle \|_2 = \sqrt{\langle \phi | \phi \rangle}$ .

*Distance measures*

We need two different metrics on  $\mathcal{S}_{\leq}(\mathcal{H})$ , the trace distance and the purified distance. These are defined as follows.

**Supplemental Definition 1** (Generalized trace distance and purified distance [42]). *For two sub-normalized quantum states  $\varrho, \sigma \in \mathcal{S}_{\leq}(\mathcal{H})$ , the trace distance is defined as*

$$\delta(\varrho, \sigma) = \frac{1}{2} (\|\varrho - \sigma\|_1 + |\text{Tr}(\varrho - \sigma)|).$$

*Their purified distance is defined as*

$$P(\varrho, \sigma) = \sqrt{1 - F(\varrho, \sigma)^2}, \quad \text{where} \quad F(\varrho, \sigma) = \|\sqrt{\varrho}\sqrt{\sigma}\|_1 + \sqrt{(1 - \text{Tr}\varrho)(1 - \text{Tr}\sigma)}$$

*is the generalized fidelity. We extend these definitions to apply to pairs of probability distributions by considering the corresponding diagonal density matrices.  $B_\varepsilon(\varrho)$  denotes the purified distance ball of radius  $\varepsilon$  around  $\varrho$ ,  $B_\varepsilon^{\text{Tr}}(\varrho)$  the trace distance ball.*

Note that the generalized trace distance coincides with the standard definition for normalized states, and the generalized fidelity coincides with the standard fidelity if at least one of the states is normalized.

The two metrics [43] are equivalent and respect the following inequalities.

**Supplemental Lemma 2** (Equivalence of trace distance and purified distance).

$$\delta(\varrho, \sigma) \leq P(\varrho, \sigma) \leq \sqrt{2\delta(\varrho, \sigma)}$$

Forgetting the eigenbases of two states does not increase their trace distance.

**Supplemental Lemma 3.** [44, Box 11.2] *We have*

$$\delta(\varrho, \sigma) \geq \delta(\text{spec}(\varrho), \text{spec}(\sigma)),$$

*where  $\text{spec}(A)$  denotes the ordered spectrum of a Hermitian matrix  $A$ .*

The following Lemma is a direct consequence of Supplemental Lemma 2 and Hölder's inequality.

**Supplemental Lemma 4.** *Let  $\varrho \in \mathcal{S}(\mathcal{H})$  be a quantum state and  $U \in \text{U}(\mathcal{H})$  a unitary. Then, we have*

$$P(U\varrho U^\dagger, \varrho) \leq \sqrt{2\|U - \mathbb{1}\|_\infty}.$$

We also need a lemma about low rank approximations of a quantum state.

**Supplemental Lemma 5.** *Let  $\varrho, \varrho' \in \mathcal{S}(\mathcal{H}_A)$  be quantum states. Then, we have*

$$P(\varrho, \varrho') \geq P(\varrho, \Pi\varrho\Pi/\text{Tr}(\Pi\varrho)),$$

where  $\Pi$  is the projection onto the support of  $\varrho'$

*Proof.* Let  $|\varrho\rangle_{AB}$  be a purification of  $\varrho$ . Then, we have

$$F(\varrho, \varrho') = \max_{|\varrho'\rangle} |\langle \varrho | \varrho' \rangle| = \max_{|\varrho'\rangle} |\langle \varrho | \Pi | \varrho' \rangle| = F(\Pi\varrho\Pi, \varrho'),$$

where the maximum is taken over purifications of  $\varrho'$  on  $AB$ . But by the Cauchy-Schwarz inequality the normalized vector with maximum inner product with  $\Pi\varrho\Pi$  is  $\Pi\varrho\Pi/\text{Tr}(\Pi\varrho)$ , which implies the claimed inequality.  $\square$

### Entropies

In this section we collect additional definitions of entropic quantities that are needed in the proofs given in this supplemental material.

In analogy to the conditional mutual information given in terms of the Shannon entropy, the quantum conditional mutual information is defined in terms of the von Neumann entropy.

**Supplemental Definition 6** (Quantum conditional mutual information). *The quantum conditional mutual information of a tripartite state  $\varrho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$  is defined as*

$$I(A; B|C)_\varrho = H(AC)_\varrho + H(BC)_\varrho - H(ABC)_\varrho - H(C)_\varrho,$$

where  $H(A)_\varrho = H(\varrho_A) = -\text{Tr}(\varrho_A \log \varrho_A)$  denotes the von Neumann entropy.

In addition to the max-mutual information defined in the main paper we use the following one-shot entropic quantities.

**Supplemental Definition 7** (Max-relative entropy). *The max-relative entropy of a state  $\varrho \in \mathcal{S}_\leq(\mathcal{H})$  with respect to a state  $\sigma \in \mathcal{S}(\mathcal{H})$  is defined as*

$$D_{\max}(\varrho||\sigma) = \min \left\{ \lambda \in \mathbb{R} \mid 2^\lambda \sigma \geq \varrho \right\}.$$

**Supplemental Definition 8** (Smooth conditional min- and max-entropy, [42, 45]). *The conditional min-entropy of a positive semidefinite matrix  $\varrho_{AB} \in \text{End}(\mathcal{H}_A \otimes \mathcal{H}_B)$  is defined as*

$$\begin{aligned} H_{\min}(A|B)_\varrho &= \max_{\sigma} \max \left\{ \lambda \mid 2^{-\lambda} \mathbb{1}_A \otimes \sigma_B \geq \varrho_{AB} \right\} \\ &= \max_{\sigma} \left( -D_{\max}(\varrho_{AB} || \mathbb{1}_A \otimes \sigma_B) \right), \end{aligned}$$

where the maximum is taken over all normalized quantum states. The conditional max-entropy is defined as the dual of the conditional min-entropy in the sense that

$$H_{\max}(A|B)_\varrho = -H_{\min}(A|C)_\varrho,$$

where  $\varrho_{ABC}$  is a purification of  $\varrho_{AB}$ . The smooth conditional min- and max-entropies are defined by maximizing and minimizing over a ball of sub-normalized states  $\tilde{\varrho}_{AB}$ , respectively,

$$H_{\min}^e(A|B)_\varrho = \max_{\tilde{\varrho} \in B_\varepsilon(\varrho)} H_{\min}(A|B)_{\tilde{\varrho}}, \quad \text{and} \quad H_{\max}^e(A|B)_\varrho = \min_{\tilde{\varrho} \in B_\varepsilon(\varrho)} H_{\max}(A|B)_{\tilde{\varrho}}.$$

The conditional max-entropy can be expressed in terms of the fidelity.



**Supplemental Lemma 9.** [46, Theorem 3] We have

$$H_{\max}(A|B)_\varrho = \max_{\sigma \in \mathcal{S}(\mathcal{H}_B)} 2 \log F(\varrho_{AB}, \mathbb{1}_A \otimes \sigma_B).$$

The unconditional min- and max-entropy are defined as their conditional counterparts with a trivial conditioning system.

**Supplemental Lemma 10.** [46] The min and max-entropy are given by

$$H_{\min}(\varrho) = -\log \|\varrho\|_\infty \quad \text{and} \quad H_{\max}(\varrho) = 2 \log \text{Tr} \sqrt{\varrho}.$$

The min-entropy does not decrease under projections[47].

**Supplemental Lemma 11.** Let  $\varrho_{AB}$  be a bipartite quantum state and  $\Pi_A$  a projection on  $\mathcal{H}_A$ . Then, we have

$$H_{\min}(A|B)_\varrho \leq H_{\min}(A|B)_{\Pi_A \varrho \Pi_A}.$$

*Proof.* Let  $\sigma_B$  be a quantum state such that  $2^{-H_{\min}(A|B)_\varrho} \mathbb{1}_A \otimes \sigma_B \geq \varrho_{AB}$ . Applying  $\Pi$  on both sides yields

$$2^{-H_{\min}(A|B)_\varrho} \mathbb{1}_A \otimes \sigma_B \geq 2^{-H_{\min}(A|B)_\varrho} \Pi_A \otimes \sigma_B \geq \Pi_A \varrho_{AB} \Pi_A.$$

This is a valid point in the maximization defining  $H_{\min}(A|B)_{\Pi_A \varrho \Pi_A}$ , implying the result.  $\square$

The fact that min- and max-entropy are invariant under local isometries [42, Lemma 13], implies that the max-mutual information has the same property.

**Supplemental Lemma 12.** For a bipartite quantum state  $\varrho_{AB}$  and isometries  $V_{A \rightarrow A'}$ ,  $W_{B \rightarrow B'}$ ,

$$I_{\max}^\varepsilon(A; B)_\varrho = I_{\max}^\varepsilon(A'; B')_{\tilde{\varrho}},$$

where  $\tilde{\varrho}_{A'B'} = V \otimes W \varrho_{AB} V^\dagger \otimes W^\dagger$

*Proof.* Suppose first that  $\varrho_A$  is invertible. Then, it follows directly from the definitions of the max-mutual information and the conditional min-entropy, that

$$I_{\max}^\varepsilon(A; B)_\varrho = -H_{\min}^\varepsilon(A|B)_{\varrho_{B|A}},$$

where  $\varrho_{B|A} = \varrho_A^{-1/2} \varrho_{AB} \varrho_A^{-1/2}$ . This implies, together with the case  $\varepsilon = 0$  of [42, Lemma 13] that the non-smooth max information is invariant under isometries.

Now, we treat the smooth case. There exists a state  $\tilde{\varrho}_{AB}$  such that  $I_{\max}^\varepsilon(A; B)_\varrho = I_{\max}^\varepsilon(A; B)_{\tilde{\varrho}}$ , i.e.

$$I_{\max}^\varepsilon(A; B)_\varrho = I_{\max}^\varepsilon(A; B)_{\tilde{\varrho}} = I_{\max}^\varepsilon(A'; B')_{\tilde{\varrho}} \geq I_{\max}^\varepsilon(A'; B')_{\tilde{\varrho}},$$

where  $\tilde{\varrho}_{A'B'} = V \otimes W \tilde{\varrho}_{AB} V^\dagger \otimes W^\dagger$ . The other inequality is proven in a way similar to the one in [42, Lemma 13]. Let  $\tau_{A'B'} \in B_\varepsilon(\tilde{\varrho})$ ,  $\eta_{B'}$  be quantum state such that  $2^\lambda \tau_{A'} \otimes \eta_{B'} \geq \tau_{A'B'}$ , where  $\lambda = I_{\max}^\varepsilon(A'; B')_{\tilde{\varrho}}$ . Let  $\Pi_V, \Pi_W$  be the projections onto the ranges of  $V$  and  $W$ . It follows that  $2^\lambda \tilde{\tau}_{A'} \otimes \tilde{\eta}_{B'} \geq \tilde{\tau}_{A'B'}$ , where  $\tilde{\tau}_{A'B'} = \Pi_V \otimes \Pi_W \tau_{A'B'} \Pi_V \otimes \Pi_W$  and  $\tilde{\eta}_{B'} = \Pi_W \eta_{B'} \Pi_W$ . It follows from the fact that the purified distance contracts under projections that  $\tilde{\tau}_{A'B'} \in B_\varepsilon(\tilde{\varrho})$  and therefore

$$I_{\max}^\varepsilon(A'; B')_{\tilde{\varrho}} \geq \min \left\{ \lambda \in \mathbb{R} \mid 2^\lambda \tilde{\tau}_A \otimes \tilde{\eta}_B \geq \tilde{\tau}_{AB} \right\} \geq I_{\max}^\varepsilon(A; B)_\varrho,$$

where  $\tilde{\tau}_{AB} = V^\dagger \otimes W^\dagger \tilde{\tau}_{A'B'} V \otimes W$  and  $\tilde{\eta}_B = W^\dagger \tilde{\eta}_{B'} W$ . The observation that the minimum can be replaced by an infimum over invertible states in the definition of the smooth max-mutual information finishes the proof.  $\square$

Tensoring a local ancilla does not change the max-mutual information.

**Supplemental Lemma 13.** Let  $\varrho_{AB}, \sigma_C$  be quantum states. The smooth max-mutual information is invariant under adding local ancillas,

$$I_{\max}^\varepsilon(A; B)_\varrho = I_{\max}^\varepsilon(A; BC)_{\varrho \otimes \sigma}$$

*Proof.* According to [11, Lemma B.17] the max-mutual information decreases under local CPTP maps. But both adding and removing an ancilla is such a map, which implies the claimed invariance.  $\square$

There are several ways to define the max-mutual information [25], one of the alternative definitions will be useful for catalytic decoupling.

**Supplemental Definition 14.** *An alternative max-mutual information of a quantum state  $\rho_{AB}$  is defined by*

$$I_{\max}(A : B)_{\rho, \rho} = D_{\max}(\rho \| \rho_A \otimes \rho_B).$$

The smooth version  $I_{\max}^\varepsilon(A : B)_{\rho, \rho}$  is defined analogously to  $I_{\max}^\varepsilon(A : B)_\rho$ ,

$$I_{\max}^\varepsilon(A : B)_{\rho, \rho} = \min_{\tilde{\rho} \in B_\varepsilon(\rho)} I_{\max}(A : B)_{\tilde{\rho}, \tilde{\rho}}$$

This alternative definition has some disadvantages, in particular the non-smooth version is not bounded from above for a fixed Hilbert space dimension. The two different smooth max-mutual informations, however, are quite similar, in particular they can be approximated up to a dimension independent error.

**Supplemental Lemma 15** ([25], Theorem 3). *For a bipartite quantum state  $\rho_{AB}$ ,*

$$I_{\max}^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'}(A : B)_\rho \lesssim I_{\max}^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'}(A : B)_{\rho, \rho} \lesssim I_{\max}^{\varepsilon'}(A : B)_\rho, \quad (29)$$

where the notation  $\lesssim$  hides errors of order  $\log(1/\varepsilon)$  as in the main text.

As an auxiliary quantity we also need the unconditional Rényi entropy of order 0.

**Supplemental Definition 16.** *For a quantum state  $\rho_A \in \mathcal{S}(\mathcal{H}_A)$  the Rényi entropy of order 0 is defined by*

$$H_0(A)_\rho = \log \text{rk}(\rho_A),$$

where  $\text{rk}(X)$  denotes the rank of a matrix  $X$ . Like in the case of the max-entropy, the smoothed version is defined by minimizing over an epsilon ball,

$$H_0^\varepsilon(A)_\rho = \min_{\tilde{\rho} \in B_\varepsilon(\rho)} H_0(A)_{\tilde{\rho}}.$$

The smoothed 0-entropy is almost equal to the smoothed max-entropy.

**Supplemental Lemma 17.** [48, Lemma 4.3] *We have*

$$H_{\max}^{2\varepsilon}(\rho) \leq H_0^{2\varepsilon}(A)_\rho \leq H_{\max}^\varepsilon(\rho) + 2\log(1/\varepsilon).$$

### Examples and proofs

Here we give proofs for the theorems given and claims made in the main paper, and explicit examples.

#### Catalytic decoupling

Here we present two proofs Theorem 1 in the main text, the achievability of catalytic decoupling.

The following is the key lemma of [32] and called convex split lemma by the authors.

**Supplemental Lemma 18.** [32, Lemma 3.1] *Let  $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  and  $\sigma \in \mathcal{S}(\mathcal{H}_E)$  be quantum states,  $k = D_{\max}(\rho_{AE} \| \rho_A \otimes \sigma_E)$  and  $0 < \delta < \frac{1}{6}$ . Define*

$$n = \begin{cases} 1 & k \leq 3\delta \\ \left\lceil \frac{8 \cdot 2^k \log(\frac{k}{\delta})}{\delta^3} \right\rceil & \text{else} \end{cases}.$$

For the state

$$\tau_{AE_1 \dots E_n} = \frac{1}{n} \sum_{j=1}^n \rho_{AE_j} \otimes \left( \sigma^{\otimes(n-1)} \right)_{E_j^c} \quad (30)$$

$E$  is decoupled from  $A$  in the following sense:

$$I(A; E_1 \dots E_n)_\tau \leq 3\delta \quad \text{as well as} \quad P(\tau_A \otimes \tau_{E_1 \dots E_n}, \tau_{AE_1 \dots E_n}) \leq \sqrt{6\delta},$$

where  $E_j^c$  denotes  $\{E_i\}_{i \neq j}$ .

**Theorem 1** (Catalytic decoupling). Let  $\hat{\rho}_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  be a quantum state. Then, for any  $0 < \delta \leq \varepsilon$  catalytic decoupling with error  $\varepsilon$  can be achieved with remainder system size

$$\log |A_2| \leq \frac{1}{2} \left( I_{\max}^{\varepsilon-\delta}(E; A)_{\hat{\rho}} + \left\{ \log \log I_{\max}^{\varepsilon-\delta}(E; A)_{\hat{\rho}} \right\}_+ \right) + \mathcal{O}(\log \frac{1}{\delta}),$$

where we define  $\{x\}_+$  to be equal to  $x$  if  $x \in \mathbb{R}_{\geq 0}$  and 0 otherwise.

*Proof.* Let  $\gamma = \varepsilon - \delta$ . Take  $\rho \in B_\gamma(\hat{\rho})$  such that  $I_{\max}(E; A)_\rho = I_{\max}^\gamma(E; A)_{\hat{\rho}}$ . Let  $\sigma_A$  be the minimizer in

$$k = I_{\max}(E; A)_\rho = \min_{\sigma_A \in \mathcal{S}(\mathcal{H}_A)} D_{\max}(\rho_{AE} \| \sigma_A \otimes \rho_E).$$

If  $k \leq \frac{\delta^2}{2}$  the state is already decoupled according to Supplemental Lemma 18 and the statement is trivially true, so let us assume  $k > \frac{\delta^2}{2}$ . We want to use Supplemental Lemma 18 so let

$$n = \left\lceil \frac{8 \cdot 2^k \log \left( \frac{k}{\delta'} \right)}{\delta'^3} \right\rceil$$

with  $\delta' = \frac{\delta^2}{6}$ ,  $\mathcal{H}_{A'} = \mathcal{H}_A^{\otimes(n-1)} \otimes \mathcal{H}_{\bar{A}}$  with  $\mathcal{H}_{\bar{A}} \cong \mathbb{C}^n$  and define the state  $\tilde{\rho}_{A^{(2)} \dots A^{(n)} \bar{A}} = \sigma^{\otimes(n-1)} \otimes \tau_{\bar{A}}$ , where  $\tau_{\bar{A}} = \mathbb{1}_{\bar{A}} / |\bar{A}|$  denotes the maximally mixed state on  $\mathcal{H}_{\bar{A}}$ . We can now define a unitary that permutes the  $A$ -systems conditioned on  $\bar{A}$  and thus creates an extension of the state  $\tau$  from Equation (30) when applied to  $\rho_{AE} \otimes \tilde{\rho}_{A'}$ ,

$$U_{AA'}^{(1)} = \sum_{j=1}^n (1j)_{A^{(1)} \dots A^{(n)}} \otimes |j-1\rangle\langle j-1|_{A'},$$

where  $(1j)_{A^{(1)} \dots A^{(n)}}$  is the transposition  $(1j) \in S_n$  under the representation  $S_n \curvearrowright \mathcal{H}_A^{\otimes n}$  of the symmetric group that acts by permuting the tensor factors, and  $(11) = 1_{S_n}$ . Now, we are almost done, as Supplemental Lemma 18 implies that

$$P(\xi_{EA^{(1)} \dots A^{(n)}}, \xi_E \otimes \xi_{A^{(1)} \dots A^{(n)}}) \leq \delta,$$

where  $\xi = U_{AA'}^{(1)} \rho_{AE} \otimes \tilde{\rho}_{A'} \left( U_{AA'}^{(1)} \right)^\dagger$ . The register  $\bar{A}$ , however, is still a factor of two larger than the claimed bound for  $|A_2|$ . We can win this factor of two by using superdense coding, as  $\bar{A}$  is classical. Let us therefore slightly enlarge  $\mathcal{H}_{\bar{A}}$  such that  $\dim(\mathcal{H}_{\bar{A}}) = m^2$  for  $m = \lceil \sqrt{n} \rceil$ . We now rotate the standard basis of  $\mathcal{H}_M$  into a Bell basis

$$|\psi_{kl}\rangle = \frac{1}{\sqrt{m}} \sum_{s=0}^{m-1} e^{\frac{2\pi i k s}{m}} |s\rangle \otimes |s+l \pmod{m}\rangle$$

of  $\mathcal{H}_{\bar{A}_1} \otimes \mathcal{H}_{\bar{A}_2}$ , with  $\mathcal{H}_{\bar{A}_i} \cong \mathbb{C}^m$ . That is done by the unitary

$$U_{\bar{A}}^{(2)} : \mathcal{H}_{\bar{A}} \rightarrow \mathcal{H}_{\bar{A}_1} \otimes \mathcal{H}_{\bar{A}_2} \quad \text{with} \quad U_{\bar{A}}^{(2)} = \sum_{k,l=0}^{m-1} |\psi_{kl}\rangle\langle mk+l|.$$

As  $\text{Tr}_{\bar{A}_2} |\psi_{kl}\rangle\langle\psi_{kl}| = \tau_{\bar{A}_1}$  for all  $k, l \in \{0, \dots, m-1\}$ , the unitary  $V_{AA' \rightarrow A_1 A_2} = U_{\bar{A}}^{(2)} U_{AA'}^{(1)}$  and the definitions  $\mathcal{H}_{A_2} = \mathcal{H}_{\bar{A}_2}$  and  $\mathcal{H}_{A_1} = \mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_{\bar{A}_1}$  achieve  $P(\eta_{A_1 E}, \eta_{A_1} \otimes \varrho_E) \leq \delta$ , with  $\eta = V_{AA' \rightarrow A_1 A_2} \varrho \otimes \tilde{\varrho} V_{AA' \rightarrow A_1 A_2}^\dagger$ . Using the triangle inequality for the purified distance we finally arrive at

$$P(\hat{\xi}_{A_1 E}, \eta_{A_1} \otimes \varrho_E) \leq \gamma + \delta = \varepsilon,$$

for  $\hat{\xi} = V_{AA' \rightarrow A_1 A_2} \hat{\varrho} \otimes \tilde{\varrho} V_{AA' \rightarrow A_1 A_2}^\dagger$ . The size of the remainder system is

$$\log |A_2| = \frac{1}{2} \log n \leq \frac{1}{2} (I_{\max}^\gamma(E; A)_{\hat{\varrho}} + \{\log \log I_{\max}^\gamma(E; A)_{\hat{\varrho}}\}_+) + \mathcal{O}\left(\log \frac{1}{\delta}\right).$$

□

**Remark.** Using the alternative definition of the max-mutual information, Supplemental Definition 14, we can prove in the same way that

$$P(\hat{\xi}_{A_1 E}, \eta_{A_1} \otimes \varrho_E) \leq \varepsilon$$

with  $\hat{\xi} = V_{AA' \rightarrow A_1 A_2} \hat{\varrho}_{AE} \otimes \varrho_A^{\otimes n} V_{AA' \rightarrow A_1 A_2}^\dagger$  and  $\eta = V_{AA' \rightarrow A_1 A_2} \varrho_{AE} \otimes \varrho_A^{\otimes n} V_{AA' \rightarrow A_1 A_2}^\dagger$  in this case, and  $n$  defined in the same way as above, just with  $k = I_{\max}^\varepsilon(A : E)_{\hat{\varrho}, \hat{\varrho}}$ . This achieves a stronger notion of decoupling, as a large part of the catalyst can be approximately handed back in the same state,

$$\eta_{A_1} = \varrho_A^{\otimes n} \otimes \tau_{\bar{A}_1}.$$

By Supplemental Lemma 15 this still implies

$$\begin{aligned} \log |A_2| &= \frac{1}{2} \log n \leq \frac{1}{2} (I_{\max}^{\varepsilon-\delta}(E; A)_{\hat{\varrho}, \hat{\varrho}} + \{\log \log I_{\max}^{\varepsilon-\delta}(E; A)_{\hat{\varrho}, \hat{\varrho}}\}_+) + \mathcal{O}\left(\log \frac{1}{\delta}\right) \\ &\leq \frac{1}{2} (I_{\max}^{\varepsilon-2\delta-2\sqrt{\delta}}(E; A)_{\hat{\varrho}} + \{\log \log I_{\max}^{\varepsilon-2\delta-2\sqrt{\delta}}(E; A)_{\hat{\varrho}}\}_+) + \mathcal{O}\left(\log \frac{1}{\delta}\right) \\ &\leq \frac{1}{2} (I_{\max}^{\varepsilon-\delta'}(E; A)_{\hat{\varrho}} + \{\log \log I_{\max}^{\varepsilon-\delta'}(E; A)_{\hat{\varrho}}\}_+) + \mathcal{O}\left(\log \frac{1}{\delta'}\right), \end{aligned}$$

having defined  $\delta' = 2\delta + 2\sqrt{\delta}$ .

The second proof is based on the state splitting protocol in [11]. This uses *embezzling states* [30].

**Supplemental Definition 19** (Embezzling state [30]). *A state  $|\mu\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  is called universal  $(d, \delta)$ -embezzling state if for any state  $|\psi\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$  with  $\dim \mathcal{H}_A = \dim \mathcal{H}_B \leq d$  there exists an isometry  $V_{\psi, X} : \mathcal{H}_X \rightarrow \mathcal{H}_X \otimes \mathcal{H}_{X'}$ ,  $X = A, B$  such that [49]*

$$P(V_{\psi, A} \otimes V_{\psi, B} |\mu\rangle, |\mu\rangle \otimes |\psi\rangle) \leq \delta.$$

**Supplemental Proposition 20.** [30] *Universal  $(d, \delta)$ -embezzling states exist for all  $d$  and  $\varepsilon$ .*

The proof also uses the one-shot version of standard decoupling.

**Supplemental Lemma 21.** [11, Theorem 3.1], [13, Table 2] *Let  $\varrho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  be a quantum state. Then, there exists a decomposition  $\mathcal{H}_A \cong \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2}$  with*

$$\log(|A_2|) \leq \frac{1}{2} (\log(|A|) - H_{\min}(A|E)_\varrho) + 2 \log \frac{1}{\varepsilon} + 1.$$

such that

$$P\left(\varrho_{A_1 E}, \frac{1_{A_1}}{|A_1|} \otimes \varrho_E\right) \leq \varepsilon.$$

The difference between the bound given here and the bound from [11, Theorem 3.1] stems from the fact that we define decoupling using the purified distance.

We include the following alternative proof to show how catalytic decoupling unifies different techniques from one-shot quantum communication. As a first step we prove the following non-smooth theorem.

**Supplemental Theorem 22** (Non-smooth catalytic decoupling from standard decoupling and embezzling states). *Let  $\varrho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  be a quantum state. Then,  $\varepsilon$ -catalytic decoupling can be achieved with remainder system size*

$$\log |A_2| \leq \frac{1}{2} I_{\max}(A; E)_\varrho + \log H_0(A)_\varrho + \mathcal{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right).$$

*In addition, if we allow for the use of isometries instead of unitaries, the ancilla systems final state is  $\varepsilon$  close to its initial state.*

*Proof.* For notational convenience let  $|\varrho\rangle_{AER}$  be a purification of  $\varrho$ . Also in slight abuse of notation we replace  $\mathcal{H}_A$  by  $\text{supp}(\varrho_A)$  so that  $|A| \leq 2^{H_0(A)_\varrho}$ . The idea is to decompose the Hilbert space  $\mathcal{H}_A$  into a direct sum of subspaces where the spectrum of  $\varrho_A$  is almost flat. Let  $Q = \lceil \log |A| + 2 \log\left(\frac{1}{\varepsilon}\right) - 1 \rceil$  and define the projectors  $P_i$ ,  $i = 0, \dots, Q+1$  such that  $P_{Q+1}$  projects onto the eigenvectors of  $\varrho_A$  with eigenvalues in  $[0, 2^{-(Q+1)}]$  and  $P_i$  projects onto the eigenvectors of  $\varrho_A$  with eigenvalues in  $[2^{-(i+1)}, 2^{-i}]$  for  $i = 0, \dots, Q$ . We can now write the approximate state  $|\bar{\varrho}\rangle_{AER} = \frac{1}{\sqrt{\alpha}} (\mathbb{1}_A - P_{Q+1}) |\varrho\rangle_{AER}$ ,  $\alpha = \text{Tr}(\mathbb{1}_A - P_{Q+1})\varrho$  as a superposition of states with almost flat marginal spectra on  $A$ ,

$$|\bar{\varrho}\rangle = \sum \sqrt{p_i} |\varrho^{(i)}\rangle,$$

with  $p_i = \text{Tr} \bar{\varrho} P_i$  and  $|\varrho^{(i)}\rangle = \frac{1}{\sqrt{p_i}} P_i |\varrho\rangle$ . This decomposition corresponds to the direct sum decomposition

$$\mathcal{H}_A \cong \bigoplus_{i=0}^{Q+1} \mathcal{H}_{A^{(i)}},$$

where  $\mathcal{H}_{A^{(i)}} = \text{supp}(P_i)$ . Note that we have  $P(\varrho, \bar{\varrho}) = \sqrt{1 - \alpha}$  and

$$1 - \alpha \leq |A| 2^{-(\log |A| + 2 \log(\frac{1}{\varepsilon}))} = \varepsilon^2,$$

i.e.  $P(\varrho, \bar{\varrho}) \leq \varepsilon$ . Now, we have a family of states,  $\{\varrho_{A^{(i)}E}^{(i)}\}$  to each of which we apply Supplemental Lemma 21. This yields decompositions  $\mathcal{H}_{A^{(i)}} \cong \mathcal{H}_{A_1^{(i)}} \otimes \mathcal{H}_{A_2^{(i)}}$  such that

$$P\left(\varrho_{A_1^{(i)}E}^{(i)}, \tau_{A_1^{(i)}} \otimes \varrho_E^{(i)}\right) \leq \varepsilon \quad (31)$$

and

$$\log(|A_2^{(i)}|) \geq \frac{1}{2} \left( \log(|A^{(i)}|) + H_{\min}(A|E)_{\varrho^{(i)}} \right) + 2 \log\left(\frac{1}{\varepsilon}\right) + 1,$$

where  $\tau_A = \frac{\mathbb{1}_A}{|A|}$  is the maximally mixed state on a quantum system  $A$ .

At this stage of the protocol the situation can be described as follows. Conditioned on  $i$ ,  $A_1^{(i)}$  is decoupled from  $E$ . If  $\varrho_E^{(i)} \neq \varrho_E^{(j)}$  and  $|A^{(i)}| \neq |A^{(j)}|$ , however, there are still correlations left between  $A_1$  and  $E$ . To get rid of this problem, we hide the maximally mixed states of different dimensions in an embezzling state by "un-embezzling" them. Let us therefore first isometrically embed all these states in the same Hilbert space. To do that, define

$$d_2 = \max_i |A_2^{(i)}| \quad \text{and} \quad d_1 = \max \left( \max_i |A_1^{(i)}|, \left\lceil \frac{|A^{(Q+1)}|}{d_2} \right\rceil \right).$$

Now, let  $\mathcal{H}_{\tilde{A}_\alpha} \cong \mathbb{C}^{d_\alpha}$  and choose isometries  $U_{A_\alpha^{(i)} \rightarrow \tilde{A}_\alpha}^{(\alpha,i)}$  for  $\alpha = 1, 2$ , define  $U_{A^{(i)} \rightarrow \tilde{A}_1 \otimes \tilde{A}_2}^{(i)} = U_{A_1^{(i)} \rightarrow \tilde{A}_1}^{(1,i)} \otimes U_{A_2^{(i)} \rightarrow \tilde{A}_2}^{(2,i)}$  for  $i = 1, \dots, Q$ . In addition, choose an isometry  $U_{A^{(Q+1)} \rightarrow \tilde{A}_1 \otimes \tilde{A}_2}^{Q+1}$ . Let  $|\mu\rangle_{A'B'}$   $\in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$  be a  $(d_1, \varepsilon)$ -embezzling state, and let  $\sigma_{A'} = \text{Tr}_{B'} |\mu\rangle\langle\mu|$ . Define the isometries  $\tilde{V}_{A' \rightarrow A' \tilde{A}_1}^{(i)}$  that would embezzle a state  $\tau_{\tilde{A}_1}^{(i)} = U_{A_1^{(i)} \rightarrow \tilde{A}_1}^{(1,i)} \tau_{A_1}^{(i)} \left( U_{A_1^{(i)} \rightarrow \tilde{A}_1}^{(1,i)} \right)^\dagger$  from  $\sigma_{A'}$ . Taking some state  $|0\rangle_{\tilde{A}_1} \in \mathcal{H}_{\tilde{A}_1}$  we can pad these embezzling isometries to unitaries  $V_{A' \tilde{A}_1}^{(i)}$  such that

$$P \left( V_{A' \tilde{A}_1}^{(i)} \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \left( V_{A' \tilde{A}_1}^{(i)} \right)^\dagger, \sigma_{A'} \otimes \tau_{\tilde{A}_1}^{(i)} \right) \leq \varepsilon. \quad (32)$$

We can combine the above isometries and unitaries now to un-embezzle the states that are approximately equal to  $\tau_{A_1}^{(i)}$  conditioned on  $i$ . Define  $\tilde{A}_3 \cong \mathbb{C}^{Q+1}$  and

$$W_{A \rightarrow \tilde{A}_1 \tilde{A}_2 \tilde{A}_3}^I = \sum_i U_{A^{(i)} \rightarrow \tilde{A}_1 \otimes \tilde{A}_2}^{(i)} P_i \otimes |i\rangle_{\tilde{A}_3}, \quad W_{A' \tilde{A}_1 \tilde{A}_3}^{II} = \sum_i \left( V_{A' \tilde{A}_1}^{(i)} \right)^\dagger \otimes |i\rangle\langle i|_{\tilde{A}_3}.$$

The final state of our decoupling protocol is

$$\varrho_{A_1 A_2 E}^f = W^{II} W^I \varrho_{AE} \otimes \sigma_{A'} \left( W^I \right)^\dagger \left( W^{II} \right)^\dagger,$$

where we omitted the subscripts of the  $V$ s for compactness and have defined  $A_1 = A' \tilde{A}_1$  and  $A_2 = \tilde{A}_2 \tilde{A}_3$ . Let us show that this protocol actually decouples  $A_1$  from  $E$ . We bound, omitting the subscripts of unitaries and isometries,

$$\begin{aligned} P \left( \varrho_{A_1 E}^f \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \otimes \varrho_E \right) &= \sum_i p_i P \left( \left( V^{(i)} \right)^\dagger U^{(1,i)} \sigma_{A'} \otimes \varrho_{A_1^{(i)} E}^{(i)} \left( U^{(1,i)} \right)^\dagger V^{(i)}, \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \otimes \varrho_E \right) \\ &= \sum_i p_i P \left( U^{(1,i)} \sigma_{A'} \otimes \varrho_{A_1^{(i)} E}^{(i)} \left( U^{(1,i)} \right)^\dagger, V^{(i)} \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \otimes \varrho_E \left( V^{(i)} \right)^\dagger \right) \\ &\leq \sum_i p_i P \left( U^{(1,i)} \sigma_{A'} \otimes \varrho_{A_1^{(i)} E}^{(i)} \left( U^{(1,i)} \right)^\dagger, \sigma_{A'} \otimes \tau_{\tilde{A}_1}^{(i)} \otimes \varrho_E \right) \\ &\quad + \sum_i p_i P \left( \sigma_{A'} \otimes \tau_{\tilde{A}_1}^{(i)} \otimes \varrho_E, V^{(i)} \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \otimes \varrho_E \left( V^{(i)} \right)^\dagger \right) \\ &\leq \sum_i p_i P \left( U^{(1,i)} \sigma_{A'} \otimes \varrho_{A_1^{(i)} E}^{(i)} \left( U^{(1,i)} \right)^\dagger, \sigma_{A'} \otimes \tau_{\tilde{A}_1}^{(i)} \otimes \varrho_E \right) + \varepsilon. \end{aligned}$$

The first inequality is the triangle inequality, the second one is Equation (32). It remains to bound the first summand,

$$\begin{aligned} \sum_i p_i P \left( U^{(1,i)} \sigma_{A'} \otimes \varrho_{A_1^{(i)} E}^{(i)} \left( U^{(1,i)} \right)^\dagger, \sigma_{A'} \otimes \tau_{\tilde{A}_1}^{(i)} \otimes \varrho_E \right) &= \sum_i p_i P \left( \sigma_{A'} \otimes \varrho_{A_1^{(i)} E}^{(i)}, \sigma_{A'} \otimes \left( \left( U^{(1,i)} \right)^\dagger \tau_{\tilde{A}_1}^{(i)} U^{(1,i)} \right) \otimes \varrho_E \right) \\ &= \sum_i p_i P \left( \sigma_{A'} \otimes \varrho_{A_1^{(i)} E}^{(i)}, \sigma_{A'} \otimes \tau_{\tilde{A}_1}^{(i)} \otimes \varrho_E \right) \\ &\leq \sum_i p_i P \left( \varrho_{A_1^{(i)} E}^{(i)}, \tau_{\tilde{A}_1}^{(i)} \otimes \varrho_E \right) \\ &\leq \varepsilon, \end{aligned}$$

where the first inequality is the triangle inequality again, and the second one is Equation (31). This shows that we achieved  $2\varepsilon$ -decoupling, i.e.

$$P \left( \varrho_{A_1 E}^f \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \otimes \varrho_E \right) \leq 2\varepsilon. \quad (33)$$

We also have to bound  $\log |A_2|$ , i.e. we need to make sure that

$$\max_{i=1, \dots, Q} \left( H_0(A)_{\varrho^{(i)}} - H_{\min}(A|E)_{\varrho^{(i)}} \right) \leq I_{\max}(E; A)_\varrho + \mathcal{O} \left( \log \left( \frac{1}{\varepsilon} \right) \right).$$

This is shown in [11] in the last part of the proof of Theorem 3.10. Thereby the size of the remainder system is bounded by

$$\log |A_2| = \frac{1}{2} I_{\max}(A; E)_{\hat{\rho}} + \log H_0(A)_{\hat{\rho}} + \mathcal{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right).$$

If we only want to use unitaries, we can complete all involved isometries to unitaries by adding an appropriate additional pure ancilla system.  $\square$

As an easy corollary we can derive a bound on the remainder system that involves the smooth max-mutual information in a way that is fit for deriving an the asymptotic expansion of Equation (15) in the main text.

**Theorem 1'** (Catalytic decoupling). Let  $\hat{\rho}_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  be a quantum state. Then,  $\varepsilon$ -catalytic decoupling can be achieved with remainder system size

$$\log |A_2| \leq \frac{1}{2} I_{\max}^{\varepsilon-\delta}(A; E)_{\rho} + \log H_0(A)_{\rho} + \mathcal{O}\left(\log\left(\frac{1}{\delta}\right)\right).$$

In addition, if we allow for the use of isometries instead of unitaries, the ancilla systems final state is  $\varepsilon$  close to its initial state.

*Proof.* Let  $\hat{\rho} \in B_{\eta}(\rho)$  with  $\eta = \varepsilon - \delta$  such that

$$I_{\max}^{\eta}(A; E)_{\rho} = I_{\max}(A; E)_{\hat{\rho}}. \quad (34)$$

Define  $\rho' = \Pi \hat{\rho} \Pi$ , where  $\Pi$  is the orthogonal projector onto the support of  $\rho$ . It follows from Uhlmann's theorem that  $P(\rho, \rho') \leq P(\rho, \hat{\rho})$ . As the max-mutual information is non-increasing under projections (cf. [11, Lemma B.19]), it follows that

$$I_{\max}^{\eta}(A; E)_{\rho} = I_{\max}(A; E)_{\rho'} \quad (35)$$

as well. Applying Supplemental Theorem 22 to  $\rho'$  and an application of the triangle inequality yields the claimed bound.  $\square$

If we accept a slightly worse smoothing parameter for the leading order term, i.e. the max-mutual information, we can smooth the second term as well and replace the renyi-0 entropy by the max-entropy.

**Corollary 5.** Let  $\rho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  be a quantum state. Then,  $\varepsilon$ -catalytic decoupling can be achieved with remainder system size

$$\log |A_2| \leq \frac{1}{2} I_{\max}^{\varepsilon'}(A; E)_{\rho} + \log H_{\max}^{\varepsilon'^2/2}(A)_{\rho} + \mathcal{O}(\log \varepsilon'),$$

where  $\varepsilon' = \varepsilon/6$ .

*Proof.* To get the bound involving smooth entropy measures we will find a state  $\hat{\rho} \in B_{2\varepsilon'}(\rho)$  such that  $I_{\max}(E; A)_{\hat{\rho}} \leq I_{\max}^{\varepsilon'}(E; A)_{\rho}$  and  $H_0(A)_{\hat{\rho}} \leq H_0^{\varepsilon'^2/2}(A)_{\rho}$ . Let  $\rho'_{AE} \in B_{\varepsilon'}(\rho_{AE})$  such that  $I_{\max}(E; A)_{\rho'} = I_{\max}^{\varepsilon'}(E; A)_{\rho}$ . Let  $\Pi_A$  be a projection of minimal rank such that  $H_0^{\varepsilon'^2/2}(A)_{\rho} \geq H_0(A)_{\rho''}$ , with  $\rho'' = \Pi_A \rho_{AE} \Pi_A \in B_{\varepsilon'}(\rho_{AE})$ . To see why such a projection exists, note that Supplemental Lemma 3 implies that there exists a state  $\rho''_A$  such that

$$H_0(A)_{\rho''_A} = H_0^{\varepsilon', \text{Tr}}(A)_{\rho} \leq H_0^{\varepsilon'^2/2}(A)_{\rho} \quad (36)$$

and  $[\rho_A, \rho''_A] = 0$ , where the inequality is due to the equivalence lemma 2 of the trace distance and the purified distance. But for the case of commuting density matrices, i.e. the classical case, it is clear that the density matrix in a given trace distance neighborhood of  $\rho$ , that has minimal rank, is just equal to  $\rho$  with the smallest eigenvalues set to zero. This implies that  $\rho''_A$  can be chosen to have the form  $\rho''_A = \Pi_A \rho_A \Pi_A$ . It is easy to see that  $P(\rho_{AE}, \rho''_{AE}) \leq \varepsilon'$  where  $\rho''_{AE} = \Pi_A \rho_{AE} \Pi_A$ : Pick a purification  $|\rho''\rangle_{AER} = \Pi_A |\rho\rangle_{AER}$  and observe that

$$F(\rho_{AE}, \rho''_{AE}) = \max_{|\sigma\rangle_{AER}} |\langle \sigma | \rho'' \rangle| = \max_{|\sigma\rangle_{AER}} |\langle \sigma | \Pi | \rho \rangle| = \text{Tr} \Pi \rho = F(\rho_A, \rho''_A),$$

where the first equation is Uhlmann's theorem and the third equation follows from the saturation of the Cauchy-Schwarz inequality. We also use that  $[q_A, q_A''] = 0$  in the last equation. Now, we define  $\hat{q} = \Pi_A q' \Pi_A$  and bound

$$P(\hat{q}_{AE}, q_{AE}) = P(\Pi_A q'_{AE} \Pi_A, q_{AE}) = P(q'_{AE}, \Pi_A q_{AE} \Pi_A) = P(q'_{AE}, q''_{AE}) \leq P(q'_{AE}, q_{AE}) + P(q_{AE}, q''_{AE}) \leq 2\varepsilon'. \quad (37)$$

The second equation follows easily by Uhlmann's theorem. According to [11, Lemma B.19] the max-mutual-information decreases under projections, i.e. we have

$$I_{\max}(E; A)_{\hat{q}} \leq I_{\max}(E; A)_{q'} = I_{\max}^{\varepsilon'}(E; A)_q.$$

Our choice of  $\Pi_A$  gives

$$H_0^{\varepsilon'^2/2}(A)_q \geq H_0(A)_{q'_A} = \log \text{rk} \Pi_A \geq H_0(A)_{\hat{q}},$$

where the first inequality is Equation (36). Now, we apply Supplemental Theorem 22, to  $\hat{q}_{AE}$ . Let  $q_{A_1 A_2 E}^{(f)}$  be the final state when applying the resulting protocol to  $q_{AE}$ . Then, we get

$$P\left(q_{A_1 E}^{(f)}, \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \otimes q_E\right) \leq 6\varepsilon'$$

by using Equations (37), (33), the triangle inequality and the monotonicity of the purified distance under CPTP maps.  $\square$

#### Standard decoupling and comparison

Let us look at an example of a state where the smooth min-entropy is almost zero and the smooth max-entropy is almost maximal to illustrate the significance of the randomization condition that is usually demanded for standard decoupling. To bound the max-entropy in the following example we need

**Supplemental Lemma 23.** *Let  $0 < q < 1$  and  $0 < \varepsilon^2 < 1 - \sqrt{1 - q}$ , and let  $|\Phi\rangle_{AB}$  be a maximally entangled state with  $\dim A = \dim B = d$ . Then, we have that*

$$H_{\min}^{\varepsilon}(A|B)_{q\Phi} \leq -\log d + \log \frac{1}{1 - \varepsilon^2 - \sqrt{1 - q}}.$$

*Proof.* We can modify the SDP for the smooth min-entropy from [50, Proof of Lemma 5] to work with subnormalized states, by adding an extra dimension. The result is that given a state  $q_{AB}$  with  $\text{Tr}[q] = p$ , the value of the following SDP is  $2^{-H_{\min}^{\varepsilon}(A|B)_q}$ :

#### Primal problem:

$$\begin{aligned} & \text{minimize } \text{Tr}[\sigma_B] \\ & \text{subject to } \tilde{q}_{AB} \leq \mathbb{1}_A \otimes \sigma_B \\ & \text{Tr} \left[ X \begin{pmatrix} q_{ABC} & \sqrt{1-p}|q\rangle \\ \sqrt{1-p}\langle q| & 1-p \end{pmatrix} \right] \geq 1 - \varepsilon^2 \\ & \text{Tr}[X] \leq 1 \\ & X = \begin{pmatrix} \tilde{q}_{ABC} & |\psi\rangle \\ \langle \psi| & x \end{pmatrix}. \end{aligned}$$

#### Dual problem:



$$\begin{aligned}
& \text{maximize } (1 - \varepsilon^2)\mu - \lambda \\
& \text{subject to } \mu \begin{pmatrix} \varrho_{ABC} & \sqrt{1-p}|\varrho\rangle \\ \sqrt{1-p}\langle\varrho| & 1-p \end{pmatrix} \leq \begin{pmatrix} E_{AB} \otimes \mathbb{1}_C & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \lambda \mathbb{1}_{ABC} & 0 \\ 0 & \lambda \end{pmatrix} \\
& \quad \text{Tr}_A[E_{AB}] \leq \mathbb{1}_B.
\end{aligned}$$

In the above,  $|\varrho\rangle_{ABC}$  is some fixed purification of  $\varrho_{AB}$ .

Now, to get the bound for  $\varrho = q\Phi$ , we can choose  $\mu = d$ ,  $E_{AB} = d\Phi_{AB}$ , and  $\lambda = d\sqrt{1-q}$ . The value of the dual problem for this choice of variables is then  $d(1 - \varepsilon^2) - d\sqrt{1-q}$ . This is therefore a lower bound on  $2^{-H_{\min}^\varepsilon(A|B)_{q\Phi}}$  and concludes the proof.  $\square$

**Example 6.** Define a probability distribution on  $\{0, 1, \dots, n\}$  by  $p(0) = p_0$  and  $p(i) = \frac{1-p_0}{n}$  for  $i \neq 0$ . Supplemental Lemma 2 shows that  $H_{\min}^\varepsilon(p) \leq H_{\min}^{\varepsilon, \text{Tr}}(p)$ , where the superscript Tr indicates that the non-smooth quantity is optimized over the trace distance ball instead of the purified distance ball. Considering that the min- and max-entropy are functions of the spectrum we can optimize over probability distributions only. The non-smooth min-entropy of  $p$  is  $H_{\min}(p) = -\log p_0$ . Assume  $p_0(1 - \varepsilon + 1/n) - 1/n \geq 0$ . Then, the best we can do for increasing this is obviously to reduce the probability of the outcome 0. Take a sub-normalized probability distribution  $q$  with  $q(0) = q_0$  and  $q(i) = p(i)$ ,  $i > 0$ . Then, we have  $\delta(p, q) = p_0 - q_0$ , i.e. by Supplemental Lemma 10

$$H_{\min}^{\varepsilon, \text{Tr}}(p) = -\log(p_0 - \varepsilon),$$

Assuming  $\varepsilon \leq p_0 - (1 - p_0)/n$ .

Using Supplemental Lemma 23 we get, assuming  $\varepsilon^2 \leq 1 - \sqrt{p_0}$ , that

$$H_{\max}^\varepsilon(p) \geq H_{\max}^\varepsilon((1 - p_0)U(n)) = -H_{\min}^\varepsilon(A|B)_{(1-p_0)\Phi} \geq \log n - \log \frac{1}{1 - \varepsilon^2 - \sqrt{p_0}},$$

where  $U(n)$  denotes the uniform distribution on  $n$  symbols. Putting in  $p_0 = 1/2$  and  $\varepsilon < 1/15$  yields, after some calculations,

$$\begin{aligned}
H_{\max}^\varepsilon(p) - H_{\min}^\varepsilon(p) & \geq H_{\max}^\varepsilon(p) - H_{\min}^{\varepsilon, \text{Tr}}(p) \geq \log(n) - \log \left( \frac{1}{(1 - \sqrt{p_0} - \varepsilon^2)(p_0 - \varepsilon)} \right) \\
& \geq \log n - \log \left( \frac{10}{1 - 15\varepsilon} \right).
\end{aligned}$$

The next theorem is a one-shot decoupling theorem for the partial trace with a bound on the remainder system involving smooth entropies. Plugging in the partial trace map into Theorem 3.1 in [13] yields a priori the non-smooth  $\log |A_2| \geq \frac{1}{2}(\log |A| - H_{\min}^\varepsilon(A|E))$  for the remainder system when decoupling  $A$  from  $E$  in a state  $\varrho_{AE}$  despite the smoothness of the term depending on the map. This can be understood considering the fact that the Choi-Jamiołkowski state of the partial trace is a tensor product of states with flat marginals, such that smoothing doesn't change much. For convenience we use [11, Theorem 3.1] as a basic decoupling theorem.

**Supplemental Theorem 24.** Let  $\varrho_{AE}$  be a bipartite quantum state, and let  $\mathcal{H}_A \cong \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2}$  such that

$$\log |A_2| \geq \frac{1}{2} (H_{\max}^\varepsilon(A)_\varrho - H_{\min}^\varepsilon(A|E)_\varrho) - 3 \log \frac{1}{\varepsilon}.$$

Then, we have

$$\int_{\mathcal{U}(\mathcal{H}_A)} P \left( \text{Tr}_{A_2} \left( U_A \varrho_{AE} U_A^\dagger \right), \frac{1_{A_1}}{|A_1|} \otimes \varrho_E \right) dU_A \leq (2 + \sqrt{7}) \varepsilon \leq 5\varepsilon.$$

*Proof.* Let  $\tilde{\rho}_{AE} \in B_\varepsilon(\rho_{AE})$  such that  $H_{\min}^\varepsilon(A|E)_\rho = H_{\min}(A|E)_{\tilde{\rho}}$ ,  $\tilde{\rho}_A \in B_\varepsilon(\rho_A)$  such that  $H_0^\varepsilon(A)_\rho = H_0(A)_{\tilde{\rho}}$  and  $\Pi_A$  the projection onto the support of  $\tilde{\rho}_A$ . Define the state  $\hat{\rho}_{AE} = \Pi_A \tilde{\rho}_{AE} \Pi_A$ . By Supplemental Lemma 5 we can assume that  $\tilde{\rho} = \Pi_A \rho \Pi_A / \text{Tr}(\Pi_A \rho)$ .  $\rho$  is normalized, so a short calculation shows that

$$F(\rho, \Pi_A \rho \Pi_A / \text{Tr}(\Pi_A \rho)) = \sqrt{\text{Tr}(\Pi_A \rho)} \Rightarrow \text{Tr}(\Pi_A \rho) \geq 1 - \varepsilon^2 \quad (38)$$

via the definitions of  $H_0^\varepsilon$  and the purified distance. By the triangle inequality we get

$$P(\rho, \hat{\rho}) \leq P(\rho, \tilde{\rho}) + P(\tilde{\rho}, \hat{\rho}) \leq \varepsilon + P(\Pi \tilde{\rho} \Pi, \Pi \rho \Pi / \text{Tr}(\Pi \rho)). \quad (39)$$

We continue to bound the last term. We have

$$\begin{aligned} F(\Pi \tilde{\rho} \Pi, \Pi \rho \Pi / \text{Tr}(\Pi \rho)) &= \frac{1}{\sqrt{\text{tr}(\Pi \rho)}} \left\| \sqrt{\Pi \tilde{\rho} \Pi} \sqrt{\Pi \rho \Pi} \right\|_1 \\ &= \frac{1}{\sqrt{\text{tr}(\Pi \rho)}} \left( F(\Pi \tilde{\rho} \Pi, \Pi \rho \Pi) - \sqrt{(1 - \text{Tr}(\Pi \rho))(1 - \text{Tr}(\Pi \tilde{\rho}))} \right) \\ &\geq \left( F(\tilde{\rho}, \rho) - \sqrt{(1 - \text{Tr}(\Pi \rho))(1 - \text{Tr}(\Pi \tilde{\rho}))} \right). \end{aligned}$$

The last step, i.e. that the generalized fidelity does not decrease under projections, follows easily from the fact that the regular fidelity does not decrease under CPTP maps. To bound the remaining term, note that

$$\sqrt{(1 - \text{Tr}(\Pi \rho))(1 - \text{Tr}(\Pi \tilde{\rho}))} + \sqrt{\text{Tr}(\Pi \rho) \text{Tr}(\Pi \tilde{\rho})} \geq F(\rho, \tilde{\rho}) \geq \sqrt{1 - \varepsilon^2}$$

by the monotonicity of the fidelity under CPTP maps. Let  $\phi, \theta \in [0, \pi/2]$  such that  $\cos^2 \phi = \text{Tr}(\Pi \rho)$  and  $\cos^2 \theta = \text{Tr}(\Pi \tilde{\rho})$ . Then, some trigonometric identities yield  $\sin(\phi - \theta) \leq \varepsilon$ , i.e. in particular  $\phi \leq \theta + \arcsin(\varepsilon)$ . Using this bound, Equation (38) and some more trigonometry yields

$$\sqrt{(1 - \text{Tr}(\Pi \rho))(1 - \text{Tr}(\Pi \tilde{\rho}))} = \sin \phi \sin \theta \leq 3\varepsilon^2 \sqrt{1 - \varepsilon^2}.$$

This implies now that

$$F(\Pi \tilde{\rho} \Pi, \Pi \rho \Pi / \text{Tr}(\Pi \rho)) \geq \sqrt{1 - \varepsilon^2} (1 - 3\varepsilon^2) \Rightarrow P(\Pi \tilde{\rho} \Pi, \Pi \rho \Pi / \text{Tr}(\Pi \rho)) \leq \sqrt{7\varepsilon^2 - 15\varepsilon^4 + 9\varepsilon^6} \leq \sqrt{7}\varepsilon.$$

Together with Equation (39) this yields  $P(\rho, \hat{\rho}) \leq (1 + \sqrt{7})\varepsilon$ . Considering  $\hat{\rho} \in \mathcal{S}(\text{supp} \tilde{\rho}_A \otimes \mathcal{H}_E)$ , an application of [11, Theorem 3.1] together with Supplemental Lemma 2 results in the following. If  $A = A_1 A_2$  and

$$\log |A_2| \geq \frac{1}{2} (\log \text{rk}(\tilde{\rho}_A) - H_{\min}(A|E)_{\hat{\rho}}) - \log \frac{1}{\varepsilon}, \quad (40)$$

then we have

$$\int_{\mathcal{U}(\mathcal{H}_A)} P \left( \text{Tr}_{A_2} \left( U_A \hat{\rho}_{AE} U_A^\dagger \right), \frac{1_{A_1}}{|A_1|} \otimes \hat{\rho}_E \right) dU_A \leq \varepsilon.$$

The last equation implies, together with the triangle inequality, that

$$\int_{\mathcal{U}(\mathcal{H}_A)} P \left( \text{Tr}_{A_2} \left( U_A \rho_{AE} U_A^\dagger \right), \frac{1_{A_1}}{|A_1|} \otimes \hat{\rho}_E \right) dU_A \leq (2 + \sqrt{7}) \varepsilon.$$

Equation (40) together with Supplemental Lemma 11 and 17 implies the claimed bound on the remainder system size.  $\square$

In the following we present a correction of the converse for decoupling by CPTP map, Corollary 4.2, from [13], a slightly tighter version of Proposition 4.

In the context of decoupling by partial trace we observed that it makes a big difference whether we demand that the decoupled system is randomized as well, i.e. that it is left in the maximally mixed state. This stops making sense in the context of decoupling by a general CPTP map  $\mathcal{T}$ , as the maximally mixed state might not even be in the *range* of  $\mathcal{T}$ . Instead one can demand randomizing in the sense that is achieved in the direct result in [13], i.e.  $\mathcal{T}_{A \rightarrow B}(\rho_{AE}) \approx \mathcal{T}_{A \rightarrow B}\left(\frac{1_A}{|A|}\right) \otimes \rho_E$ .

The following theorem from [13] is already a converse statement for decoupling by CPTP map.

**Supplemental Theorem 25.** [13, Theorem 4.1] Let  $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  and  $\mathcal{T} : \text{End}(\mathcal{H}_A) \rightarrow \text{End}(\mathcal{H}_B)$  a CPTP map such that

$$\|\mathcal{T} \otimes \text{id}_E(\rho_{AE}) - \mathcal{T}(\rho_A) \otimes \rho_E\|_1 \leq \varepsilon.$$

Then, we have

$$H_{\min}^{2\sqrt{6\varepsilon''+2\varepsilon}+2\sqrt{\varepsilon'+\varepsilon''}}(A|E)_\rho + H_{\max}^{\varepsilon''}(A|B)_\omega \geq \log \varepsilon'$$

for all  $\varepsilon', \varepsilon'' > 0$ , where  $\omega_{AB} = \text{id}_A \otimes \mathcal{T}_{A' \rightarrow B}(\rho_{AA'})$  with  $\mathcal{H}_{A'} \cong \mathcal{H}_A$  and  $\rho_{AA'}$  a purification of  $\rho_A$ .

It involves, however, the term  $H_{\max}^{\varepsilon''}(A|B)_\omega$  that depends on both the state and the CPTP-map. Unfortunately the proof of the Corollary following this theorem, Corollary 4.2, contains a mistake and the statement is incorrect as it is stated in [13]. The reason for this is that the converse, Corollary 4.2, does not assume decoupling *and* randomizing, while the direct result, [13, Theorem 3.1], provides a condition for exactly that. Adding this condition to the statement of Corollary 4.2 renders it true and we give a proof of it in the following.

**Proposition 4'** (Corrected version of Corollary 4.2 in [13]). Let  $\rho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  and let the CPTP map  $\mathcal{T} : \text{End}(\mathcal{H}_A) \rightarrow \text{End}(\mathcal{H}_B)$  be such that

$$\int_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}_{A \rightarrow B}(U_A \rho_{AE} U_A^\dagger), \tau_B \otimes \rho_E\right) dU_A \leq \varepsilon.$$

Then, we have

$$H_{\min}^{4\sqrt{6\varepsilon''+2\varepsilon}+2\varepsilon'+\varepsilon'''}(A|E)_\rho + H_{\max}^{\varepsilon''}(A|B)_\tau \geq -10 \log\left(\frac{1}{\varepsilon'''}\right) - 7$$

for all  $\varepsilon''', \varepsilon'' > 0$ , where  $\tau = \mathcal{T}_{A' \rightarrow B}(\phi_{AA'}^+)$  is the Choi-Jamiołkowski state of  $\mathcal{T}$ .

*Proof.* For  $\delta \geq 0$  arbitrary, let  $\mathcal{D} \subset \mathcal{U}_A$ ,  $|\mathcal{D}| < \infty$  be a  $\delta$ -net in  $\mathcal{U}_A$  in the operator norm, i.e. a finite subset such that for all  $U \in \mathcal{U}_A$  there exists  $V \in \mathcal{D}$  such that  $\|U - V\|_\infty \leq \delta$ . Now, define the state

$$\tilde{\rho}_{AEU} = \sum_{U_A \in \mathcal{D}} q_{U_A} U_A \rho_{AE} U_A^\dagger \otimes |U_A\rangle\langle U_A|_U$$

where  $\mathcal{H}_U = \mathbb{C}^{|\mathcal{D}|}$ ,

$$q_{U_A} = \mu\left(\left\{U \in \mathcal{U}_A \mid \|U - U_A\|_\infty \leq \|U - V\|_\infty \forall V \in \mathcal{D}\right\}\right),$$

and  $\mu$  denotes the Haar measure. The assumption implies that  $\mathcal{T}_{A \rightarrow B}$  decouples  $A$  from  $EU$ . To see this, note that

$$\begin{aligned} P(\mathcal{T}_{A \rightarrow B} \tilde{\rho}_{AEU}, \mathcal{T}_{A \rightarrow B} \tilde{\rho}_A \otimes \tilde{\rho}_{EU}) &= \frac{1}{|\mathcal{D}|} \sum_{U_A \in \mathcal{D}} P\left(\mathcal{T}_{A \rightarrow B}^{U_A} \rho_{AE}, \mathcal{T}_{A \rightarrow B} \tilde{\rho}_A \otimes \rho_E\right) \\ &= \int_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}_{A \rightarrow B}^{D(U_A)} \rho_{AE}, \mathcal{T}_{A \rightarrow B} \tilde{\rho}_A \otimes \rho_E\right) dU \\ &\leq \int_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}_{A \rightarrow B}^{U_A} \rho_{AE}, \mathcal{T}_{A \rightarrow B} \tilde{\rho}_A \otimes \rho_E\right) dU + \sqrt{2\delta}, \end{aligned} \quad (41)$$

where  $\mathcal{T}_{A \rightarrow B}^{U_A} : X \mapsto \mathcal{T}_{A \rightarrow B}(U_A X U_A^\dagger)$  and  $D(U_A) = \min_{V \in \mathcal{D}} \|U_A - V\|_\infty$ . The last inequality follows easily using the  $\delta$ -net-property, the triangle inequality, the fact that the purified distance decreases under CPTP maps and Supplemental Lemma 4. By assumption we then have

$$P(\mathcal{T}_{A \rightarrow B} \tilde{\rho}_{AEU}, \mathcal{T}_{A \rightarrow B} \tau_A \otimes \tilde{\rho}_{EU}) \leq \varepsilon + 2\sqrt{2\delta},$$

as  $P(\mathcal{T}_{A \rightarrow B} \tau_A, \mathcal{T}_{A \rightarrow B} \tilde{\rho}_A) \leq \sqrt{2\delta}$  by a similar argument as in Equation 41. Using Supplemental Lemma 2 and applying Supplemental Theorem 25 to this situation, i.e. the map  $\mathcal{T}_{A \rightarrow B}$  that decouples system  $A$  from  $EU$  applied to the state  $\tilde{\rho}$ , we get

$$H_{\min}^{2\sqrt{6\varepsilon''+2\varepsilon+8\delta}+2\sqrt{\varepsilon'+\varepsilon''}}(A|EU)_{\tilde{\rho}} + H_{\max}^{\varepsilon''}(A|B)_\tau \geq -\log\left(\frac{1}{\varepsilon'}\right), \quad (42)$$

whit the Choi-Jamiołkowski state  $\tau_{AB} = \mathcal{T}_{A' \rightarrow B} \phi_{AA'}^+$ . Let  $\eta = 2\sqrt{6\varepsilon''+4\varepsilon} + 2\sqrt{\varepsilon'+\varepsilon''}$ . The min-entropy term can be transformed using the chain rules for smooth entropies [50],

$$\begin{aligned} H_{\min}^\eta(A|EU)_{\tilde{\rho}} &\leq H_{\min}^{2\eta+\varepsilon^{(3)}}(AU|E)_{\tilde{\rho}} - H_{\min}(U|E)_{\tilde{\rho}} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) \\ &= H_{\min}^{2\eta+\varepsilon^{(3)}}(AU|E)_{\rho \otimes \tau_U} - H_{\min}(U)_{\tau_U} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) \\ &\leq H_{\min}^{2\eta+\varepsilon^{(3)}+2\varepsilon^{(4)}}(A|E)_\rho + H_{\max}(U|AE)_{\rho \otimes \tau_U} - H_{\min}(U)_{\tau_U} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) + 3\log\left(\frac{2}{(\varepsilon^{(4)})^2}\right) \\ &= H_{\min}^{2\eta+\varepsilon^{(3)}+2\varepsilon^{(4)}}(A|E)_\rho + H_{\max}(U)_{\tau_U} - H_{\min}(U)_{\tau_U} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) + 3\log\left(\frac{2}{(\varepsilon^{(4)})^2}\right) \\ &\leq H_{\min}^{2\eta+\varepsilon^{(5)}}(A|E)_\rho + 8\log\left(\frac{1}{\varepsilon^{(5)}}\right) + 13, \end{aligned} \quad (43)$$

where we used a chain rule in the first inequality, in the second line that  $U$  is independent from  $E$ , the invariance of the smooth entropies under isometries and that there exists a controlled unitary  $V_{UA}$  such that  $V_{UA} \tilde{\rho}_{AEU} V_{UA}^\dagger = \rho_{AE} \otimes \tau_U$ , and another chain rule in the fourth line. In the last line we set  $\varepsilon^{(5)} = 2\varepsilon^{(4)} + \varepsilon^{(3)}$  and  $\varepsilon^{(4)} = 2\varepsilon^{(3)}/3$  to get an optimal error term. Combining Equations (42) and (43) we get

$$H_{\min}^{4\sqrt{6\varepsilon''+2\varepsilon+4\delta}+4\sqrt{\varepsilon'+2\varepsilon''+\varepsilon^{(5)}}}(A|E)_\rho + H_{\max}^{\varepsilon''}(A|B)_\tau \geq -\log\left(\frac{1}{\varepsilon'}\right) - 8\log\left(\frac{1}{\varepsilon^{(5)}}\right) - 13.$$

Fixing  $\varepsilon''' = 4\sqrt{\varepsilon'} + \varepsilon^{(5)}$  and optimizing the logarithmic error term yields

$$H_{\min}^{4\sqrt{6\varepsilon''+2\varepsilon+4\delta}+2\varepsilon''+\varepsilon'''}(A|E)_\rho + H_{\max}^{\varepsilon''}(A|B)_\tau \geq -10\log\left(\frac{1}{\varepsilon'''}\right) - 7.$$

As  $\delta$  was arbitrary, we can take the limit [51]  $\delta \rightarrow 0$ , which concludes the proof.  $\square$

Theorem 3 follows as an easy corollary.

**Proposition 4.** Let  $\rho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$  and let the CPTP map  $\mathcal{T} : \text{End}(\mathcal{H}_A) \rightarrow \text{End}(\mathcal{H}_B)$  be such that

$$\int_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}_{A \rightarrow B}(U_A \rho_{AE} U_A^\dagger), \mathcal{T}_{A \rightarrow B}(\tau_A) \otimes \rho_E\right) dU_A \leq \varepsilon.$$

Then, we have

$$H_{\min}^{15\sqrt{\varepsilon}}(A|E)_\rho + H_{\max}^\varepsilon(A|B)_\tau \geq -10\log\left(\frac{1}{\varepsilon}\right) - 7,$$

where  $\tau = \mathcal{T}_{A' \rightarrow B}(\phi_{AA'}^+)$  is the Choi-Jamiołkowski state of  $\mathcal{T}$ .

*Proof.* Setting  $\varepsilon = \varepsilon' = \varepsilon'' = \varepsilon'''$  in Theorem 4' and bounding  $\varepsilon \leq \sqrt{\varepsilon}$  yields the result.  $\square$

*Asymptotic expansion*

Here we give the necessary definitions and point to the relevant references to derive the asymptotic expansion given in Equation (15) in the main paper.

**Supplemental Definition 26** (Quantum relative entropy [52] and quantum information variance [34]). *For quantum states  $\varrho, \sigma \in \mathcal{S}(\mathcal{H})$  the quantum relative entropy is defined as follows:*

$$D(\varrho\|\sigma) = \begin{cases} \text{Tr} \varrho (\log \varrho - \log \sigma) & \text{supp} \varrho \subset \text{supp} \sigma \\ \infty & \text{else} \end{cases},$$

*i.e. as the expectation of  $\log \varrho - \log \sigma$  with respect to  $\varrho$ . The quantum information variance is the corresponding variance,*

$$V(\varrho\|\sigma) = \begin{cases} \text{Tr} \varrho (\log \varrho - \log \sigma)^2 & \text{supp} \varrho \subset \text{supp} \sigma \\ \infty & \text{else} \end{cases}.$$

The von Neumann entropy and derived quantities can be expressed in terms of the quantum relative entropy and thereby given a corresponding variance. In particular we have that

$$I(A; B)_\varrho = D(\varrho_{AB}\|\varrho_A \otimes \varrho_B).$$

Consequently we define  $V(A; B)_\varrho = V(\varrho\|\varrho_A \otimes \varrho_B)$ .

Equation (15) in the main paper makes use of the cumulative normal distribution,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-y^2/2) dy.$$

Note that this function is invertible.

The derivation of the asymptotic expansion is not detailed here, as it is completely analogous to the derivation in [34, Section VI].

*Quantum state redistribution from catalytic decoupling*

In this section we show how to apply any decoupling with ancilla protocol to quantum state redistribution. Let us first define the task of quantum state redistribution (QSR).

**Supplemental Definition 27** (Quantum state redistribution [3, 6]).

- Let  $|\psi\rangle_{ABCR}$  be a four party quantum state where Alice holds systems  $A$  and  $C$ , Bob holds System  $B$  and a referee holds system  $R$ . An  $\varepsilon$ -quantum state redistribution protocol with communication cost  $q$  is a protocol in which Alice performs some encoding operation on her shares of  $|\psi\rangle_{ABCR}$  and some resource state  $|\phi\rangle_{A'B'}$  shared between Alice and Bob, then she sends a quantum register  $C_2$  of size  $\log |C_2| = q$  to Bob who performs some decoding operation such that the final state is  $|\tilde{\psi}\rangle_{ABCR} \otimes |\tilde{\phi}\rangle_{A''B''}$  with  $P(|\psi\rangle_{ABCR}, |\tilde{\psi}\rangle_{ABCR}) \leq \varepsilon$  and Alice holds  $A$ , Bob holds  $B$  and  $C$  and the referee still holds  $R$ .
- For trivial system  $A$ , i.e.  $\mathcal{H}_A = \mathbb{C}$ , the task is called quantum state merging, for  $\mathcal{H}_B = \mathbb{C}$  quantum state splitting.

Asymptotically QSR can be achieved with a quantum communication cost of  $I(R; C|A)_\psi$  [6], i.e. there exists a sequence of QSR protocols for  $|\psi\rangle_{ABCR}^{\otimes n}$  with quantum communication cost  $q_n$  such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} q_n = I(R; C|A)_\psi.$$

Anshu *et al.* [32] define the following quantity that characterizes the quantum communication cost of one-shot QSR:

$$I_{\max}^\varepsilon(R; C|A)_\psi = \inf I_{\max}(RA; CA')_{U_{ACA'} \varrho_{RACA'} U_{ACA'}^\dagger}$$

where the infimum is taken over ancilla systems  $A'$ , states  $\sigma_{A'}$ , states  $\rho \in B_\varepsilon(\psi_{RAC} \otimes \sigma_{A'})$  and unitaries  $U_{ACA'} \in \mathcal{U}(\mathcal{H}_{ACA'})$  such that  $\text{Tr}_{CA'} U_{ACA'} \rho_{RACA'} U_{RACA'}^\dagger \in B_\varepsilon(\psi_{RA})$ . We call this quantity the smooth conditional max-mutual-information. Note that in [32] it is denoted by  $Q_\psi^\varepsilon$ . Using the same minimization idea we can get a QSR protocol, that improves over the naive use of a state splitting or state merging protocol to achieve QSR, from any decoupling theorem. The special case of state merging is presented as Proposition 3 in the main text.

**Supplemental Theorem 28** (Quantum state redistribution from decoupling). *Quantum state redistribution for a state  $|\psi\rangle_{ABCR}$  can be achieved up to a purified distance error of  $3\varepsilon$  with a quantum communication cost of*

$$q^\varepsilon(R; C|A)_\psi = \inf R_c^\varepsilon(AR; CA'')_{U_{ACA''} \rho_{ARCA''} U_{ACA''}^\dagger} \quad (44)$$

where the infimum is taken over ancilla systems  $A''$ , states  $\sigma_{A''}$ , states  $\rho_{ARCA''} \in B_\varepsilon(\psi_{RAC} \otimes \sigma_{A''})$  and unitaries  $U = U_{ACA''} \in \mathcal{U}(\mathcal{H}_{ACA''})$  such that  $\text{Tr}_{CA''} U_{ACA''} \rho_{ARCA''} U_{ACA''}^\dagger \in B_\varepsilon(\psi_{RA})$ . For state merging the quantity  $q^\varepsilon(R; C|A)_\psi$  reduces to  $R_c^\varepsilon(R; C)$ .

The problem that the infimum is taken over unbounded Hilbert space dimensions, and therefore might not be achievable using a finite-dimensional Hilbert space, is artificial in view of the fact that any one-shot protocol has a communication cost  $q \in \log \mathbb{N}$ , which is discrete, i.e. the infimum is actually a minimum.

The proof is an adaptation of the protocol used in [32], Theorem 4.2, run backwards.

*Proof.* Let  $\mathcal{H}_{A''}$ ,  $\sigma_{A''} \in \mathcal{S}(\mathcal{H}_{A''})$ ,  $\rho \in B_\varepsilon(\psi_{RAC} \otimes \sigma_{A''})$  and  $U_{ACA''} \in \mathcal{U}(\mathcal{H}_{ACA''})$  be a tuple that saturates the infimum in Equation (44). If such does not exist because the infimum is taken over unbounded finite Hilbert space dimensions, take a tuple that saturates the infimum up to  $\varepsilon$ . Now, consider the following protocol:

1. Starting point of the protocol is that Alice, Bob and the Referee share a state  $\psi_{ABCD} \otimes \sigma_{A''B''} \otimes \tilde{\rho}_{A'B'}$ , where  $\sigma_{A''B''}$  is a purification of  $\sigma_{A''}$ ,  $\tilde{\rho}_{A'B'}$  is a purification of any state  $\tilde{\rho}_{A'}$  that Alice will need for decoupling, and Alice holds systems  $ACA''A'$ , Bob holds systems  $BB''B'$  and the Referee holds  $R$ .
2. Alice applies the unitary  $U_{ACA''}$
3. Alice takes the  $\varepsilon$ -decoupling isometry  $V_{CA''A' \rightarrow C_1 C_2}$  that was constructed for decoupling systems  $CA''A'$  of the state  $U_{ACA''} \rho_{RACA''} U_{ACA''}^\dagger$  from  $AR$  and runs it on her state  $U_{ACA''} (\psi_{RAC} \otimes \sigma_{A''} \otimes \tilde{\rho}_{A'}) U_{ACA''}^\dagger$ . She then sends the  $q^\varepsilon(R; C|A)_\psi$ -qbit remainder system  $C_2$  to Bob. The decoupling isometry with these properties exists by assumption.
4. For Alice's and the Referee's joint state

$$\tilde{\xi}_{C_1 AR} = \text{Tr}_{C_2} V U (\psi_{RAC} \otimes \sigma_{A''} \otimes \tilde{\rho}_{A'}) U^\dagger V^\dagger$$

the triangle inequality for the purified distance yields  $P(\tilde{\xi}_{C_1 AR}, \tilde{\xi}_{C_1} \otimes \psi_{AR}) \leq 3\varepsilon$ . So according to Uhlmann's theorem Bob can apply an isometry such that the final state of the protocol is  $3\varepsilon$ -close to  $\psi_{ABCR} \otimes \tilde{\xi}'_{C_1 C_1}$  in purified distance, where  $\tilde{\xi}'_{C_1 C_1}$  is a purification of  $\tilde{\xi}_{C_1}$ .

For state merging, i.e. the case of trivial  $A$ , the ancilla  $A''$  becomes unnecessary and the unitary  $U_{ACA''}$  can be taken to be equal to the identity. This yields the claimed improvement.  $\square$

Together with Theorem 1 this recovers the result from [32] that one-shot quantum state redistribution is achievable with a communication cost of  $I_{\max}^\varepsilon(R; C|A)_\psi$  plus lower order terms.