# Catalytic Decoupling of Quantum Information: Supplemental Material

Christian Majenz,[1,*] Mario Berta,[2] Frédéric Dupuis,[3] Renato Renner,[4] and Matthias Christandl[1]

[1]*QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø.*
[2]*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA 91125, USA.*
[3]*Faculty of Informatics, Masaryk University, Brno, Czech Republic.*
[4]*Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland.*

(Dated: January 29, 2017)

### 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 $\mathrm{End}(\mathcal{H})$. The set of normalized quantum states on a Hilbert space $\mathcal{H}$ is denoted by $\mathcal{S}(\mathcal{H}) = \{\varrho \in \mathrm{End}(\mathcal{H}) \,|\mathrm{Tr}\varrho = 1, \varrho \geq 0\}$, the set of sub-normalized quantum states by $\mathcal{S}_{\leq}(\mathcal{H}) = \{\varrho \in \mathrm{End}(\mathcal{H}) \,|\mathrm{Tr}\varrho \leq 1, \varrho \geq 0\}$. 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 $\mathrm{U}(\mathcal{H})$. We will make use of two matrix norms, the trace norm and the operator norm, defined by

$$\|A\|_1 = \mathrm{Tr}\sqrt{A^\dagger A}$$
$$\|A\|_\infty = \max_{|\phi\rangle\in\mathcal{H}} \|A|\phi\rangle\|_2,$$

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

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

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

*Their purified distance is defined as*

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

*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 $\varrho$.*

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 [1] 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.

---

[1] It is shown in [1] that the generalized trace distance and the generalized purified distance are metrics.

**Supplemental Lemma 3.** *[2, Box 11.2] We have*

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

*where* $\mathrm{spec}(A)$ *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* $\varrho \in \mathbb{S}(\mathcal{H})$ *be a quantum state and* $U \in \mathrm{U}(\mathcal{H})$ *a unitary. Then, we have*

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

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

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

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

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

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

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

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 $\Pi\varrho\Pi/\mathrm{Tr}(\Pi\varrho)$, which implies the claimed inequality. □

*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* $\varrho_{ABC} \in \mathbb{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ *is defined as*

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

*where* $H(A)_\varrho = H(\varrho_A) = -\mathrm{Tr}(\varrho_A \log \varrho_A)$ *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* $\varrho \in \mathbb{S}_{\leq}(\mathcal{H})$ *with respect to a state* $\sigma \in \mathbb{S}(\mathcal{H})$ *is defined as*

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

**Supplemental Definition 8** (Smooth conditional min- and max-entropy, [1, 3])**.** *The conditional min-entropy of a positive semidefinite matrix* $\varrho_{AB} \in \mathrm{End}\left(\mathcal{H}_A \otimes \mathcal{H}_B\right)$ *is defined as*

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

*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,*

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

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

**Supplemental Lemma 9.** *[4, Theorem 3] We have*

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

*where* $|A| = \dim \mathcal{H}_A$.

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

**Supplemental Lemma 10.** *[4] The min and max-entropy are given by*

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

The min-entropy does not decrease under projections.[2]

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

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

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

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

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

The fact that min- and max-entropy are invariant under local isometries [1, Lemma 13], implies that the max-mutual information has the same property.

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

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

*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

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

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

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

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

where $\tilde{\bar{\varrho}}_{A'B'} = V \otimes W \bar{\varrho}_{AB} V^\dagger \otimes W^\dagger$. The other inequality is proven in a way similar to the one in [1, 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 $\lambda = I^\varepsilon_{\max}(A';B')_{\tilde{\varrho}}$. Let $\Pi_V, \Pi_W$ be the projections onto the ranges of $V$ and $W$. It follows that $2^\lambda \bar{\tau}_{A'} \otimes \bar{\eta}_{B'} \geq \bar{\tau}_{A'B'}$, where $\bar{\tau}_{A'B'} = \Pi_V \otimes \Pi_W \tau_{A'B'} \Pi_V \otimes \Pi_W$ and $\bar{\eta}_{B'} = \Pi_W \eta_{B'} \Pi_W$. It follows from the fact that the purified distance contracts under projections that $\bar{\tau}_{A'B'} \in B_\varepsilon(\tilde{\varrho})$ and therefore

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

where $\tilde{\tau}_{AB} = V^\dagger \otimes W^\dagger \bar{\tau}_{A'B'} V \otimes W$ and $\tilde{\eta}_B = W^\dagger \bar{\eta}_B W$. 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. □

---

[2] This is counterintuitive and due to the fact that we only project but do not renormalize the state.

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,*

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

*Proof.* According to [5, 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. □

There are several ways to define the max-mutual information [6], one of the alternative definitions will be useful for catalytic decoupling.

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

$$I_{\max}(A:B)_{\varrho,\varrho} = D_{\max}(\varrho\|\varrho_A\otimes\varrho_B).$$

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

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

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** ([6], Theorem 3)**.** *For a bipartite quantum state $\varrho_{AB}$,*

$$I^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'}_{\max}(A:B)_{\varrho} \lesssim I^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'}_{\max}(A:B)_{\varrho,\varrho} \lesssim I^{\varepsilon'}_{\max}(A:B)_{\varrho}, \tag{1}$$

*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 $\varrho_A \in \mathbb{S}(\mathcal{H}_A)$ the Rényi entropy of order 0 is defined by*

$$H_0(A)_{\varrho} = \log \mathrm{rk}(\varrho_A),$$

*where $\mathrm{rk}(X)$ 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,*

$$H^{\varepsilon}_0(A)_{\varrho} = \min_{\tilde{\varrho}\in B_{\varepsilon}(\varrho)} H_0(A)_{\tilde{\varrho}}.$$

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

**Supplemental Lemma 17.** *[7, Lemma 4.3] We have*

$$H^{2\varepsilon}_{\max}(\varrho) \leq H^{2\varepsilon}_0(A)_{\varrho} \leq H^{\varepsilon}_{\max}(\varrho) + 2\log(1/\varepsilon).$$

### Examples and proofs

Here we give proofs for the theorems given and claims made in the main paper, and explicit examples. We also repeat some of the definitions in the paper in formal mathematical language.

Here we give formal definitions of the minimal remainder system sizes $R^\varepsilon(A:E)_\varrho$ and $R^\varepsilon_c(A:E)_\varrho$ and present two proofs of Theorem 1 in the main text, the achievability of catalytic decoupling.

**Supplemental Definition 18** (Minimal remainder system sizes for standard decoupling)**.** *Let $\varrho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ be a bipartite quantum state and $1 \geq \varepsilon \geq 0$ an error parameter. The minimal remainder system size $R^\varepsilon(A:E)_\varrho$ for decoupling $A$ from $E$ in $\varrho_{AE}$ up to an error $\varepsilon$ is defined as the minimal number $r$ such that there exists a unitary $U_A \in \mathrm{U}(\mathcal{H}_A)$ and a decomposition $\mathcal{H}_A \cong \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2}$ with $\log|A_2| = r$, as well as states $\omega_{A_1} \in \mathcal{S}(\mathcal{H}_{A_1})$ and $\tilde{\omega}_E \in \mathcal{S}(\mathcal{H}_E)$ with the following property:*

$$P\left(\mathrm{Tr}_{A_2}\left(U_A \otimes \mathbb{1}_E\right)\varrho_{AE}\left(U_A^\dagger \otimes \mathbb{1}_E\right), \omega_{A_1} \otimes \tilde{\omega}_E\right) \leq \varepsilon.$$

**Supplemental Definition 19** (Minimal remainder system sizes for catalytic decoupling)**.** *Let $\varrho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ be a bipartite quantum state and $1 \geq \varepsilon \geq 0$ an error parameter. The minimal remainder system size $R^\varepsilon_c(A:E)_\varrho$ for catalytically decoupling $A$ from $E$ in $\varrho_{AE}$ up to an error $\varepsilon$ is defined as the minimum over all finite-dimensional ancilla Hilbert spaces $\mathcal{H}_{A'}$ and all ancilla states $\sigma_{A'} \in \mathcal{S}(\mathcal{H}_{A'})$ of the minimal remainder system size for (standard) decoupling $AA'$ from $E$ in $\varrho_{AE} \otimes \sigma_{A'}$, as a formula*

$$R^\varepsilon_c(A:E)_\varrho = \min_{\mathcal{H}_{A'}, \sigma_{A'}} R^\varepsilon(AA':E)_{\varrho\otimes\sigma}. \tag{2}$$

In the context of the model for erasure of correlations introduced in [8] we define the (catalytic) correlation erasure cost.

**Supplemental Definition 20** ((Catalytic) correlation erasure cost)**.** *Let $\varrho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ be a bipartite quantum state and $1 \geq \varepsilon \geq 0$ an error parameter. The cost $R^\varepsilon_U(A;E)_\varrho$ of erasing the correlations of $A$ with $E$ in the state $\varrho_{AE}$ up to an error $\varepsilon$ is defined as the minimal number $r$ such that there exists a family of unitaries $\{U_A^{(i)}\}_{i\in\{0,\ldots,2^r-1\}}$ with $U_A^{(i)} \in \mathrm{U}(\mathcal{H}_A)$ and a probability distribution $p$ on $\{0,\ldots,2^r-1\}$ as well as states $\omega_A \in \mathcal{S}(\mathcal{H}_A)$ and $\tilde{\omega}_E \in \mathcal{S}(\mathcal{H}_E)$ such that*

$$P\left(\sum_{i=0}^{2^r-1} p_i \left(U_A^{(i)} \otimes \mathbb{1}_E\right)\varrho_{AE}\left(\left(U_A^{(i)}\right)^\dagger \otimes \mathbb{1}_E\right), \omega_A \otimes \tilde{\omega}_E\right) \leq \varepsilon.$$

*As for decoupling, we define the cost $R^\varepsilon_{U,c}(A;E)_\varrho$ of catalytically erasing the correlations of $A$ with $E$ in the state $\varrho_{AE}$ up to an error $\varepsilon$ as the minimum over all finite-dimensional ancilla Hilbert spaces $\mathcal{H}_{A'}$ and all ancilla states $\sigma_{A'} \in \mathcal{S}(\mathcal{H}_{A'})$ of the cost of erasing the correlations of $AA'$ with $E$ in $\varrho_{AE} \otimes \sigma_{A'}$, as a formula*

$$R^\varepsilon_{U,c}(A:E)_\varrho = \min_{\mathcal{H}_{A'}, \sigma_{A'}} R^\varepsilon_U(AA':E)_{\varrho\otimes\sigma}. \tag{3}$$

**Remark 21.** The minimum in Equations (2) and (3) always exists. This is because any remainder system size, or logarithm of a number of unitaries, is the logarithm of an integer greater or equal than one i.e. the minimum is taken over a discrete set that is bounded from below (by zero).

The following is the key lemma of [9] and called convex split lemma by the authors.

**Supplemental Lemma 22.** *[9, Lemma 3.1] Let $\varrho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ and $\sigma \in \mathcal{S}(\mathcal{H}_E)$ be quantum states, $k = D_{max}(\varrho_{AE} \| \varrho_A \otimes \sigma_E)$ and $0 < \delta < \frac{1}{6}$. Define*

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

*For the state*

$$\tau_{AE_1\ldots E_n} = \frac{1}{n}\sum_{j=1}^{n} \varrho_{AE_j} \otimes \left(\sigma^{\otimes(n-1)}\right)_{E_{j^c}} \tag{4}$$

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

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

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

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

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

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 $\varrho \in B_\gamma(\hat{\varrho})$ such that $I_{\max}(E;A)_\varrho = I_{\max}^\gamma(E;A)_{\hat{\varrho}}$. Let $\sigma_A$ be the minimizer in

$$k = I_{\max}(E;A)_\varrho = \min_{\sigma_A \in \mathrm{S}(\mathcal{H}_A)} D_{\max}\left(\varrho_{AE} \middle\| \sigma_A \otimes \varrho_E\right).$$

If $k \leq \frac{\delta^2}{2}$ the state is already decoupled according to Supplemental Lemma 22 and the statement is trivially true, so let us assume $k > \frac{\delta^2}{2}$. We want to use Supplemental Lemma 22 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 $\tilde{\varrho}_{A^{(2)}...A^{(n)}\bar{A}} = \sigma^{\otimes(n-1)} \otimes \tau_{\bar{A}}$, 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 (4) when applied to $\varrho_{AE} \otimes \tilde{\varrho}_{A'}$,

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

where $(1j)_{A^{(1)}...A^{(n)}}$ is the transposition $(1j) \in S_n$ under the representation $S_n \looparrowright \mathcal{H}_A^{\otimes n}$ of the symmetric group that acts by permuting the tensor factors, and $(11) = 1_{S_n}$. Now, we are almost done, as Supplemental Lemma 22 implies that

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

where $\xi = U_{AA'}^{(1)} \varrho_{AE} \otimes \tilde{\varrho}_{A'} \left(U_{AA'}^{(1)}\right)^\dagger$. 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

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

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

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

As $\mathrm{Tr}_{\bar{A}_2}|\psi_{kl}\rangle\langle\psi_{kl}| = \tau_{\bar{A}_1}$ for all $k,l \in \{0,...,m-1\}$, the unitary $V_{AA'\to A_1A_2} = U_{\bar{A}}^{(2)} U_{AA'}^{(1)}$ 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 $P\left(\eta_{A_1E}, \eta_{A_1} \otimes \varrho_E\right) \leq \delta$, with $\eta = V_{AA'\to A_1A_2}\varrho \otimes \tilde{\varrho} V_{AA'\to A_1A_2}^\dagger$. Using the triangle inequality for the purified distance we finally arrive at

$$P\left(\hat{\xi}_{A_1E}, \eta_{A_1} \otimes \varrho_E\right) \leq \gamma + \delta = \varepsilon,$$

for $\hat{\xi} = V_{AA'\to A_1A_2}\hat{\varrho}\otimes\tilde{\varrho}V^\dagger_{AA'\to A_1A_2}$. The size of the remainder system is

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

□

**Remark.** Using the alternative definition of the max-mutual information, Supplemental Definition 14, and the notation from the above theorem and proof, we can prove in the same way that

$$P(\hat{\xi}_{A_1E}, \eta_{A_1}\otimes\varrho_E) \leq \varepsilon$$

with $\hat{\xi} = V_{AA'\to A_1A_2}\hat{\varrho}_{AE}\otimes\varrho_A^{\otimes n}V^\dagger_{AA'\to A_1A_2}$ and $\eta = V_{AA'\to A_1A_2}\varrho_{AE}\otimes\varrho_A^{\otimes n}V^\dagger_{AA'\to A_1A_2}$ in this case, and $n$ defined in the same way as above, just with $k = I^\varepsilon_{\max}(A:E)_{\hat{\varrho},\hat{\varrho}}$. This achieves a stronger notion of decoupling, as catalyst can be approximately handed back in the same state, except the $A_2$ part, which is lost in the partial trace,

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

By Supplemental Lemma 15 this still implies

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

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

The second proof is based on the state splitting protocol in [5]. This uses *embezzling states* [10].

**Supplemental Definition 23** (Embezzling state [10]). *A state $|\mu\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$ is called a* 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 $V_{\psi,X} : \mathcal{H}_X \to \mathcal{H}_X\otimes\mathcal{H}_{X'}$, $X = A, B$ such that*[3]

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

**Supplemental Proposition 24.** *[10] Universal $(d,\delta)$-embezzling states exist for all $d$ and $\varepsilon$.*

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

**Supplemental Lemma 25.** *[5, Theorem 3.1], [11, 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*

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

*such that*

$$P\left(\varrho_{A_1E}, \frac{1_{A_1}}{|A_1|}\otimes\varrho_E\right) \leq \varepsilon.$$

---

[3] The original concept was defined using the trace distance instead of the purified distance [10]. We use the purified distance here as it fits our task, the definitions are equivalent up to a square according to Supplemental Lemma 2.

The difference between the bound given here and the bound from [5, 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 26** (Non-smooth catalytic decoupling from standard decoupling and embezzling states)**.** *Let $\varrho_{AE} \in \mathrm{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ 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 $\mathrm{supp}(\varrho_A)$ 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 $Q = \left\lceil \log|A| + 2\log\left(\frac{1}{\varepsilon}\right) - 1 \right\rceil$ and define the projectors $P_i,\ i = 0, ..., Q+1$ such that $P_{Q+1}$ projects onto the eigenvectors of $\varrho_A$ with eigenvalues in $\left[0, 2^{-(Q+1)}\right]$ and $P_i$ projects onto the eigenvectors of $\varrho_A$ with eigenvalues in $\left[2^{-(i+1)}, 2^{-i}\right]$ for $i = 0, ..., Q$. We can now write the approximate state $|\bar{\varrho}\rangle_{AER} = \frac{1}{\sqrt{\alpha}}(\mathbb{1}_A - P_{Q+1})|\varrho\rangle_{AER}, \alpha = \mathrm{Tr}(\mathbb{1}_A - P_{Q+1})\varrho$ as a superposition of states with almost flat marginal spectra on $A$,

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

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

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

i.e. $P(\varrho, \bar{\varrho}) \leq \varepsilon$. Now, we have a family of states, $\{\varrho^{(i)}_{A^{(i)}E}\}$ to each of which we apply Supplemental Lemma 25. This yields decompositions $\mathcal{H}_{A^{(i)}} \cong \mathcal{H}_{A_1^{(i)}} \otimes \mathcal{H}_{A_2^{(i)}}$ such that

$$P\left(\varrho^{(i)}_{A_1^{(i)}E}, \tau_{A_1^{(i)}} \otimes \varrho^{(i)}_E\right) \leq \varepsilon \tag{5}$$

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 $\left|A^{(i)}\right| \neq \left|A^{(j)}\right|$, 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

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

Now, let $\mathcal{H}_{\tilde{A}_\alpha} \cong \mathbb{C}^{d_\alpha}$ and choose isometries $U^{(\alpha,i)}_{A_\alpha^{(i)} \to \tilde{A}_\alpha}$ for $\alpha = 1,2$, define $U^{(i)}_{A^{(i)} \to \tilde{A}_1 \otimes \tilde{A}_2} = U^{(1,i)}_{A_1^{(i)} \to \tilde{A}_1} \otimes U^{(2,i)}_{A_2^{(i)} \to \tilde{A}_2}$ for $i = 1, ..., Q$. In addition, choose an isometry $U^{Q+1}_{A^{(Q+1)} \to \tilde{A}_1 \otimes \tilde{A}_2}$. Let $|\mu\rangle_{A'B'} \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$ be a $(d_1, \varepsilon)$-embezzling state, and let $\sigma_{A'} = \mathrm{Tr}_{B'} |\mu\rangle\langle\mu|$. Define the isometries $\bar{V}^{(i)}_{A' \to A'\tilde{A}_1}$ that would embezzle a state

$$\tau^{(i)}_{\tilde{A}_i} = U^{(1,i)}_{A_1^{(i)} \to \tilde{A}_1} \tau_{A_1^{(i)}} \left( U^{(1,i)}_{A_1^{(i)} \to \tilde{A}_1} \right)^\dagger$$

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 $V^{(i)}_{A'\tilde{A}_1}$ such that

$$P\left( V^{(i)}_{A'\tilde{A}_1} \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \left( V^{(i)}_{A'\tilde{A}_1} \right)^\dagger, \sigma_{A'} \otimes \tau^{(i)}_{\tilde{A}_i} \right) \leq \varepsilon. \tag{6}$$

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

$$W^I_{A \to \tilde{A}_1 \tilde{A}_2 \tilde{A}_3} = \sum_i U^{(i)}_{A^{(i)} \to \tilde{A}_1 \otimes \tilde{A}_2} P_i \otimes |i\rangle_{\tilde{A}_3}, \quad W^{II}_{A'\tilde{A}_1\tilde{A}_3} = \sum_i \left( V^{(i)}_{A'\tilde{A}_1} \right)^\dagger \otimes |i\rangle\langle i|_{\tilde{A}_3}.$$

The final state of our decoupling protocol is

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

where we omitted the subscripts of the $W$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,

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

The first inequality is the triangle inequality, the second one is Equation (6). It remains to bound the first summand,

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

where the first inequality is the triangle inequality again, and the second one is Equation (5). This shows that we achieved $2\varepsilon$-decoupling, i.e.

$$P\left( \varrho^f_{A_1 E}, \sigma_{A'} \otimes |0\rangle\langle 0|_{\tilde{A}_1} \otimes \varrho_E \right) \leq 2\varepsilon. \tag{7}$$

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

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

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

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

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

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

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

$$I_{\max}^{\eta}(A:E)_\varrho = I_{\max}(A:E)_{\hat{\varrho}}. \tag{8}$$

Define $\varrho' = \Pi\hat{\varrho}\Pi$, where $\Pi$ is the orthogonal projector onto the support of $\varrho$. It follows from Uhlmann's theorem that $P(\varrho,\varrho') \leq P(\varrho,\hat{\varrho})$. As the max-mutual information is non-increasing under projections (cf. [5, Lemma B.19]), it follows that

$$I_{\max}^{\eta}(A:E)_\varrho = I_{\max}(A:E)_{\varrho'} \tag{9}$$

as well. Applying Supplemental Theorem 26 to $\varrho'$ and an application of the triangle inequality yields the claimed bound. □

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 Rényi-0 entropy by the max-entropy.

**Corollary 1.** *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*

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

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

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

$$H_0(A)_{\varrho''_A} = H_0^{\varepsilon',\mathrm{Tr}}(A)_\varrho \leq H_0^{\varepsilon'^2/2}(A)_\varrho \tag{10}$$

and $[\varrho_A, \varrho''_A] = 0$, 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 $\varrho$, that has minimal rank, is just equal to $\varrho$ with the smallest eigenvalues set to zero. This implies that $\varrho''_A$ can be chosen to have the form $\varrho''_A = \Pi_A \varrho_A \Pi_A$. It is easy to see that $P(\varrho_{AE}, \varrho''_{AE}) \leq \varepsilon'$ where $\varrho''_{AE} = \Pi_A \varrho_{AE} \Pi_A$: Pick a purification $|\varrho''\rangle_{AER} = \Pi_A |\varrho\rangle_{AER}$ and observe that

$$F(\varrho_{AE}, \varrho''_{AE}) = \max_{|\sigma\rangle_{AER}} \left|\langle \sigma \mid \varrho'' \rangle\right| = \max_{|\sigma\rangle_{AER}} |\langle \sigma | \Pi | \varrho \rangle| = \mathrm{Tr}\Pi\varrho = F(\varrho_A, \varrho''_A),$$

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

$$P(\hat{\varrho}_{AE}, \varrho_{AE}) = P(\Pi_A \varrho'_{AE} \Pi_A, \varrho_{AE}) = P(\varrho'_{AE}, \Pi_A \varrho_{AE} \Pi_A) = P(\varrho'_{AE}, \varrho''_{AE}) \leq P(\varrho'_{AE}, \varrho_{AE}) + P(\varrho_{AE}, \varrho''_{AE}) \leq 2\varepsilon'. \quad (11)$$

The second equation follows easily by Uhlmann's theorem. According to [5, Lemma B.19] the max-mutual-information decreases under projections, i.e. we have

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

Our choice of $\Pi_A$ gives

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

where the first inequality is Equation (10). Now, we apply Supplemental Theorem 26, to $\hat{\varrho}_{AE}$. Let $\varrho^{(f)}_{A_1 A_2 E}$ be the final state when applying the resulting protocol to $\varrho_{AE}$. Then, we get

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

by using Equations (11), (7) , the triangle inequality and the monotonicity of the purified distance under CPTP maps. □

*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 27.** *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*

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

*Proof.* We can modify the SDP for the smooth min-entropy from [12, Proof of Lemma 5] to work with subnormalized states, by adding an extra dimension. The result is that given a state $\varrho_{AB}$ with $\mathrm{Tr}[\varrho] = p$, the value of the following SDP is $2^{-H^{\varepsilon}_{\min}(A|B)_\varrho}$:

**Primal problem:**

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

**Dual problem:**

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

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 $2^{-H^{\varepsilon}_{\min}(A|B)_{q\Phi}}$ and concludes the proof. □

**Example 2.** Define a probability distribution on $\{0, 1, ..., n\}$ by $p(0) = p_0$ and $p(i) = \frac{1-p_0}{n}$ for $i \neq 0$. Supplemental Lemma 2 shows that $H^{\varepsilon}_{\min}(p) \leq H^{\varepsilon(p),\mathrm{Tr}}_{\min}(p)$, 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

$$H^{\varepsilon,\mathrm{Tr}}_{\min}(p) = -\log(p_0 - \varepsilon),$$

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

Using Supplemental Lemma 27 we get, assuming $\varepsilon^2 \leq 1 - \sqrt{p_0}$, that

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

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

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

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 [11] yields a priori the non-smooth $\log|A_2| \geq \frac{1}{2}\left(\log|A| - H^{\varepsilon}_{\min}(A|E)\right)$ 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 [5, Theorem 3.1] as a basic decoupling theorem.

**Supplemental Theorem 28.** *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*

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

*Then, we have*

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

*Proof.* Let $\tilde{\varrho}_{AE} \in B_\varepsilon(\varrho_{AE})$ such that $H^\varepsilon_{\min}(A|E)_\varrho = H_{\min}(A|E)_{\tilde{\varrho}}$, $\bar{\varrho}_A \in B_\varepsilon(\varrho_A)$ such that $H^\varepsilon_0(A)_\varrho = H_0(A)_{\bar{\varrho}}$ and $\Pi_A$ the projection onto the support of $\bar{\varrho}_A$. Define the state $\hat{\varrho}_{AE} = \Pi_A \tilde{\varrho}_{AE} \Pi_A$. By Supplemental Lemma 5 we can assume that $\bar{\varrho} = \Pi_A \varrho \Pi_A / \mathrm{Tr}\varrho$. $\varrho$ is normalized, so a short calculation shows that

$$F(\varrho, \Pi_A \varrho \Pi_A / \mathrm{Tr}(\Pi_A \varrho)) = \sqrt{\mathrm{Tr}(\Pi_A \varrho)} \quad \Rightarrow \quad \mathrm{Tr}(\Pi_A \varrho) \geq 1 - \varepsilon^2 \tag{12}$$

via the definitions of $H^\varepsilon_0$ and the purified distance. By the triangle inequality we get

$$P(\varrho, \hat{\varrho}) \leq P(\varrho, \bar{\varrho}) + P(\bar{\varrho}, \hat{\varrho}) \leq \varepsilon + P(\Pi\tilde{\varrho}\Pi, \Pi\varrho\Pi / \mathrm{Tr}(\Pi\varrho)). \tag{13}$$

We continue to bound the last term. We have

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

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

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

by the monotonicity of the fidelity under CPTP maps. Let $\phi, \theta \in [0, \pi/2]$ such that $\cos^2\phi = \mathrm{Tr}(\Pi\varrho)$ and $\cos^2\theta = \mathrm{Tr}(\Pi\tilde{\varrho})$. Then, some trigonometric identities yield $\sin(\phi - \theta) \leq \varepsilon$, i.e. in particular $\phi \leq \theta + \arcsin(\varepsilon)$. Using this bound, Equation (12) and some more trigonometry yields

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

This implies now that

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

Together with Equation (13) this yields $P(\varrho, \hat{\varrho}) \leq (1+\sqrt{7})\varepsilon$. Considering $\hat{\varrho} \in \mathcal{S}(\mathrm{supp}\bar{\varrho}_A \otimes \mathcal{H}_E)$, an application of [5, Theorem 3.1] together with Supplemental Lemma 2 results in the following. If $A = A_1 A_2$ and

$$\log|A_2| \geq \frac{1}{2}\left(\log \mathrm{rk}(\bar{\varrho}_A) - H_{\min}(A|E)_{\hat{\varrho}}\right) - \log\frac{1}{\varepsilon}, \tag{14}$$

then we have

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

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

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

Equation (14) together with Supplemental Lemma 11 and 17 implies the claimed bound on the remainder system size. □

In the following we present a correction of the converse for decoupling by CPTP map, Corollary 4.2, from [11], 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 [11], i.e. $\mathcal{T}_{A\to B}\left(\varrho_{AE}\right) \approx \mathcal{T}_{A\to B}\left(\frac{1_A}{|A|}\right) \otimes \varrho_E$.

The following theorem from [11] is already a converse statement for decoupling by CPTP map.

**Supplemental Theorem 29.** *[11, Theorem 4.1] Let $\varrho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ and $\mathcal{T} : \mathrm{End}\left(\mathcal{H}_A\right) \to \mathrm{End}\left(\mathcal{H}_B\right)$ a CPTP map such that*

$$\|\mathcal{T} \otimes \mathrm{id}_E(\varrho_{AE}) - \mathcal{T}(\varrho_A) \otimes \varrho_E\|_1 \leq \varepsilon.$$

*Then, we have*

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

*for all $\varepsilon', \varepsilon'' > 0$, where $\omega_{AB} = \mathrm{id}_A \otimes \mathcal{T}_{A'\to B}(\varrho_{AA'})$ with $\mathcal{H}_{A'} \cong \mathcal{H}_A$ and $\varrho_{AA'}$ a purification of $\varrho_A$.*

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 [11]. The reason for this is that the converse, Corollary 4.2, does not assume decoupling *and* randomizing, while the direct result, [11, 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 [11])**.** Let $\varrho_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ and let the CPTP map $\mathcal{T} : \mathrm{End}\left(\mathcal{H}_A\right) \to \mathrm{End}\left(\mathcal{H}_B\right)$ be such that

$$\int\limits_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}_{A\to B}(U_A \varrho_{AE} U_A^\dagger), \tau_B \otimes \varrho_E\right) \mathrm{d}U_A \leq \varepsilon.$$

Then, we have

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

for all $\varepsilon''', \varepsilon'' > 0$, where $\tau = \mathcal{T}_{A'\to B}(\phi^+_{AA'})$ 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

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

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

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

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

$$\begin{aligned} P\left(\mathcal{T}_{A\to B}\tilde{\varrho}_{AEU}, \mathcal{T}_{A\to B}\tilde{\varrho}_A \otimes \tilde{\varrho}_{EU}\right) &= \frac{1}{|\mathcal{D}|}\sum_{U_A \in \mathcal{D}} P\left(\mathcal{T}^{U_A}_{A\to B}\varrho_{AE}, \mathcal{T}_{A\to B}\tilde{\varrho}_A \otimes \varrho_E\right) \\ &= \int\limits_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}^{D(U_A)}_{A\to B}\varrho_{AE}, \mathcal{T}_{A\to B}\tilde{\varrho}_A \otimes \varrho_E\right) \mathrm{d}U \\ &\leq \int\limits_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}^{U_A}_{A\to B}\varrho_{AE}, \mathcal{T}_{A\to B}\tilde{\varrho}_A \otimes \varrho_E\right) \mathrm{d}U + \sqrt{2\delta}, \end{aligned} \tag{15}$$

where $\mathcal{T}_{A\to B}^{U_A} : X \mapsto \mathcal{T}_{A\to B}(U_A X U_A^\dagger)$ and $D(U_A) = \text{minarg}_{V\in\mathcal{D}}\|U_A - V\|_\infty$. 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

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

as $P(\mathcal{T}_{A\to B}\tau_A, \mathcal{T}_{A\to B}\tilde{\varrho}_A) \leq \sqrt{2\delta}$ by a similar argument as in Equation 15. Using Supplemental Lemma 2 and applying Supplemental Theorem 29 to this situation, i.e. the map $\mathcal{T}_{A\to B}$ that decouples system $A$ from $EU$ applied to the state $\tilde{\varrho}$, we get

$$H_{\min}^{2\sqrt{6\varepsilon''+2\varepsilon+8\delta}+2\sqrt{\varepsilon'}+\varepsilon''}(A|EU)_{\tilde{\varrho}} + H_{\max}^{\varepsilon''}(A|B)_\tau \geq -\log\left(\frac{1}{\varepsilon'}\right), \tag{16}$$

with the Choi-Jamiołkowski state $\tau_{AB} = \mathcal{T}_{A'\to B}\phi_{AA'}^+$. 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 [**?** ],

$$\begin{aligned}
H_{\min}^{\eta}(A|EU)_{\tilde{\varrho}} &\leq H_{\min}^{2\eta+\varepsilon^{(3)}}(AU|E)_{\tilde{\varrho}} - H_{\min}(U|E)_{\tilde{\varrho}} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) \\
&= H_{\min}^{2\eta+\varepsilon^{(3)}}(AU|E)_{\varrho\otimes\tau_U} - H_{\min}(U)_{\tau_U} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) \\
&\leq H_{\min}^{2\eta+\varepsilon^{(3)}+2\varepsilon^{(4)}}(A|E)_\varrho + H_{\max}(U|AE)_{\varrho\otimes\tau_U} - H_{\min}(U)_{\tau_U} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) + 3\log\left(\frac{2}{(\varepsilon^{(4)})^2}\right) \\
&= H_{\min}^{2\eta+\varepsilon^{(3)}+2\varepsilon^{(4)}}(A|E)_\varrho + H_{\max}(U)_{\tau_U} - H_{\min}(U)_{\tau_U} + \log\left(\frac{2}{(\varepsilon^{(3)})^2}\right) + 3\log\left(\frac{2}{(\varepsilon^{(4)})^2}\right) \\
&\leq H_{\min}^{2\eta+\varepsilon^{(5)}}(A|E)_\varrho + 8\log\left(\frac{1}{\varepsilon^{(5)}}\right) + 13,
\end{aligned} \tag{17}$$

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 $V_{UA}\tilde{\varrho}_{AEU}V_{UA}^\dagger = \varrho_{AE}\otimes\tau_U$, 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 (16) and (17) we get

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

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

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

As $\delta$ was arbitrary, we can take the limit[4] $\delta \to 0$, which concludes the proof. □

Proposition 4 follows as an easy corollary.

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

$$\int_{\mathcal{U}(\mathcal{H}_A)} P\left(\mathcal{T}_{A\to B}(U_A\varrho_{AE}U_A^\dagger), \mathcal{T}_{A\to B}(\tau_A)\otimes\varrho_E\right) \mathrm{d}U_A \leq \varepsilon.$$

Then, we have

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

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

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

[4] The limit $\delta \to 0$ exists, as the min-entropy term that depends on $\delta$ is nondecreasing in $\delta$ and bounded from below.

*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 30** (Quantum relative entropy [13] and quantum information variance [14, 15])**.** *For quantum states $\varrho, \sigma \in \mathbb{S}(\mathcal{H})$ the quantum relative entropy is defined as follows:*

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

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

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

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,

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

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 [14, 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 31** (Quantum state redistribution [16, 17])**.**

- *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 here 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.
- *We denote the minimal quantum communication cost $q^\varepsilon(C\rangle B|A)_\psi$ for $\varepsilon$-quantum state redistribution as the the minimal q such that a protocol as described above exists. For trivial system A we write $q^\varepsilon(C\rangle B)_\psi$ for the quantum communication cost of quantum state merging.*

Asymptotically QSR can be achieved with a quantum communication cost of $I(R;C|A)_\psi$ [16], i.e.

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}q^\varepsilon(C\rangle B|A)_\psi = I(R;C|A)_\psi.$$

Anshu *et al.* [9] define the following quantity that that characterizes the quantum communication cost of one-shot QSR:

$$I^{\varepsilon}_{\max}(R;C|A)_{\psi} = \inf I_{\max}(RA;CA')_{U\varrho U^{\dagger}},$$

where the infimum is taken over ancilla systems $A'$, states $\sigma_{A'}$, states $\varrho \in B_{\varepsilon}(\psi_{RAC} \otimes \sigma_{A'})$ and unitaries $U_{ACA'} \in \mathcal{U}(\mathcal{H}_{ACA'})$ such that $\mathrm{Tr}_{CA'} U_{ACA'} \varrho_{RACA'} U^{\dagger}_{RACA'} \in B_{\varepsilon}(\psi_{RA})$. We call this quantity the smooth conditional max-mutual-information. Note that in [9] it is denoted by $Q^{\varepsilon}_{\psi}$. 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 (possibly catalytic) decoupling theorem. The special case of state merging is presented as Proposition 3 in the main text.

**Supplemental Theorem 32** (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*

$$q^{\varepsilon}(R;C|A)_{\psi} = \inf R^{\varepsilon}_{c}(AR;CA'')_{U\varrho U^{\dagger}}, \tag{18}$$

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

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 [9], Theorem 4.2, run backwards.

*Proof.* Let $\mathcal{H}_{A''}, \sigma_{A''} \in \mathrm{S}(\mathcal{H}_{A''})$, $\varrho \in B_{\varepsilon}(\psi_{RAC} \otimes \sigma_{A''})$ and $U_{ACA''} \in \mathcal{U}(\mathcal{H}_{ACA''})$ be a tuple that saturates the infimum in Equation (18). 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 $\psi_{ABCD} \otimes \sigma_{A''B''} \otimes \tilde{\varrho}_{A'B'}$, where $\sigma_{A''B''}$ is a purification of $\sigma_{A''}$, $\tilde{\varrho}_{A'B'}$ is a purification of any state $\tilde{\varrho}_{A'}$ 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 $V_{CA''A' \to C_1C_2}$ that was constructed for decoupling systems $CA''A'$ of the state $U_{ACA''} \varrho_{RACA''} U^{\dagger}_{ACA''}$ from $AR$ and runs it on her state $U_{ACA''} \left(\psi_{RAC} \otimes \sigma_{A''} \otimes \tilde{\varrho}_{A'}\right) U^{\dagger}_{ACA''}$. She then sends the $q^{\varepsilon}(R;C|A)_{\psi}$-qubit 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

$$\xi_{C_1AR} = \mathrm{Tr}_{C_2} VU \left(\psi_{RAC} \otimes \sigma_{A''} \otimes \tilde{\varrho}_{A'}\right) U^{\dagger} V^{\dagger}$$

   the triangle inequality for the purified distance yields $P(\xi_{C_1AR}, \xi_{C_1} \otimes \psi_{AR}) \leq 3\varepsilon$. So according to Uhlmann's theorem Bob can apply an isometry such that the final state of the protocol is $3\varepsilon$-close to $\psi_{ABCR} \otimes \xi'_{C_1C_1'}$ in purified distance, where $\xi'_{C_1C_1'}$ is a purification of $\xi_{C_1}$.

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. □

Together with Theorem 1 this recovers the result from [9] that one-shot quantum state redistribution is achievable with a communication cost of $I^{\varepsilon}_{\max}(R;C|A)_{\psi}$ plus lower order terms.

* majenz@math.ku.dk

[1] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Trans. Inf. Theory **56**, 4674 (2010).
[2] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, 2010).
[3] R. Renner, *Security of quantum key distribution*, Phd thesis (2005).
[4] R. König, R. Renner, and C. Schaffner, IEEE Trans. Inf. Theory **55**, 4337 (2009).
[5] M. Berta, M. Christandl, and R. Renner, Commun. Math. Phys. **306**, 579 (2011).
[6] N. Ciganovic, N. J. Beaudry, and R. Renner, IEEE Trans. Inf. Theory **60**, 1573 (2014).
[7] R. Renner and S. Wolf, in *Proceedings of the IEEE International Symposium on Information Theory* (2004) pp. 233–233.
[8] B. Groisman, S. Popescu, and A. Winter, Phys. Rev. A **72**, 032317 (2005).
[9] A. Anshu, V. K. Devabathini, and R. Jain, preprint arXiv:1410.3031 (2014).
[10] W. van Dam and P. Hayden, Phys. Rev. A **67**, 060302 (2003).
[11] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, Commun. Math. Phys. **328**, 251 (2014).
[12] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner, IEEE Trans. Inf. Theory **59**, 2603 (2013).
[13] H. Umegaki, in *Kodai Mathematical Seminar Reports*, Vol. 14 (Department of Mathematics, Tokyo Institute of Technology, 1962) pp. 59–85.
[14] M. Tomamichel and M. Hayashi, IEEE Trans. Inf. Theory **59**, 7693 (2013).
[15] K. Li, Ann. Stat. **42**, 171 (2014).
[16] J. T. Yard and I. Devetak, IEEE Trans. Inf. Theory **55**, 5339 (2009).
[17] M. Horodecki, J. Oppenheim, and A. Winter, Commun. Math. Phys. **269**, 107 (2007).