Smooth Max-Information as One-Shot Generalization for Mutual Information

Nikola Ciganović, Normand J. Beaudry, and Renato Renner

Abstract

We study formal properties of smooth max-information, a generalization of von Neumann mutual information derived from the max-relative entropy. Recent work suggests that it is a useful quantity in one-shot channel coding, quantum rate distortion theory and the physics of quantum many-body systems.

Max-information can be defined in multiple ways. We demonstrate that different smoothed definitions are essentially equivalent (up to logarithmic terms in the smoothing parameters). These equivalence relations allow us to derive new chain rules for the max-information in terms of min- and max-entropies, thus extending the smooth entropy formalism to mutual information.

Index Terms

Chain rules, mutual information, one-shot information theory, smooth entropy.

I. INTRODUCTION

Mutual information has been an important concept from the beginning of information theory. In classical information theory, the Shannon mutual information,

\[ I(A:B) = H(A) - H(A|B), \]

serves as a measure for the capacity of communication channels [1]. In quantum information theory, its analogue is given by the von Neumann mutual information which is defined in terms of von Neumann entropy in the same way as in (1). It generally presents a measure of correlation between the subsystems \(A\) and \(B\) of a composite quantum system. The operational drawback of these quantities from a practical point of view is that they only characterize processes under the assumption that they can be repeated an arbitrary number of times and that these repetitions are completely uncorrelated. In other words, the assumption states that the available resources are independent and identically distributed or i.i.d.. However, this assumption is not justified in more realistic settings. Channels for instance need not be memoryless and the outputs for consecutive inputs may therefore be correlated. Also, assuming an i.i.d. structure in cryptographic protocols may compromise their security since an adversary may perform an attack that is not i.i.d.. A great amount of research has consequently been devoted to scenarios where the resources are not i.i.d., commonly called the one-shot setting. This scenario is not only closer to realistic communication settings, but can also be regarded as strictly more general. The i.i.d. case is a limiting case and can thus be reproduced from one-shot results. Hence, one-shot information theory also serves as a method for proving i.i.d. statements.

In order to characterize processes in the one-shot scenario, the smooth min- and max-entropies \(H^\varepsilon_{\text{min}}\) and \(H^\varepsilon_{\text{max}}\) have been introduced [2], [3] and studied extensively both operationally and formally (see for example [4]–[9]). They satisfy properties like data processing inequalities [3], [10] and a set of chain rules [8], [11]. Operationally, min- and max-entropies can be used to characterize various information theoretic tasks, including randomness extraction and state merging [4]. When the i.i.d.-limit is taken, i.e. if we evaluate them on average over \(n\) for states of the form \(\rho^\otimes n = \rho \otimes \rho \otimes \cdots \otimes \rho\) with asymptotically large \(n\), they indeed reproduce the von Neumann entropy [5], [10] (this is called the quantum asymptotic equipartition property, or QAEP). Furthermore, smooth entropies have been shown to be asymptotically equivalent to an independent approach to non-asymptotic information theory.

This work was supported by the Swiss National Science Foundation (SNF) through the National Centre of Competence in Research “Quantum Science and Technology” (grant No. 200020-135048) and the European Research Council (grant 258932). N. J. Beaudry, N. Ciganović, and R. Renner are with the Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland. (e-mail: nbeaudry@phys.ethz.ch; nikolac@student.ethz.ch; renner@phys.ethz.ch)
namely the information spectrum method as introduced by Han and Verdú in classical information theory [14], [15] and later generalized to the quantum setting by Nagaoka, Hayashi, Bowen, and Datta [16]–[18]. In light of the success of the smooth entropy formalism, the question arises of how it can be extended to mutual information in a meaningful way.

Recent research has produced a whole variety of expressions that appear to be useful one-shot generalizations for mutual information. Motivated by (1), generalized mutual information quantities can be defined as

\[ I^\varepsilon_{\text{gen}}(A : B) := H^\varepsilon_{\min}(A) - H^\varepsilon_{\min}(A|B) \]

or

\[ H^\varepsilon_{\min}(A) - H^\varepsilon_{\max}(A|B) \]

or

\[ H^\varepsilon_{\max}(A) - H^\varepsilon_{\max}(A|B) \]

or

\[ H^\varepsilon_{\max}(A) - H^\varepsilon_{\min}(A|B) \].

Several of these expressions have been found to have useful applications as bounds on one-shot capacities [19], [20] or in the study of area laws in quantum statistical physics [21].

On the other hand, it is well known that the von Neumann entropy and mutual information can be defined as special cases of the quantum relative entropy

\[ D(\rho||\sigma) = \text{tr} \left( \rho (\log \rho - \log \sigma) \right) \]

where tr denotes the trace and log is the logarithm with base 2 throughout the paper. Therefore, it appears natural to define generalized information theoretic quantities in terms of generalized relative entropies. Min- and max-entropies for example are derived from the max- and min-relative entropy [7],

\[ D_{\max}(\rho||\sigma) = \min \{ \lambda | 2^\lambda \sigma \geq \rho \} \]

and

\[ D_{\min}(\rho||\sigma) = - \log \| \sqrt{\rho} \sqrt{\sigma} \|_1^2, \]

respectively. In this paper, we focus on the mutual information quantity corresponding to the max-relative entropy, called the max-information. Recent work has established the max-information as a relevant quantity in different information theoretic tasks. It has been identified by Berta et al. as a measure for the quantum communication cost of state splitting and state merging protocols [22], [23]. In addition, Datta et al. found the smooth max-information to characterize the minimal one-shot qubit compression size for a quantum rate distortion code [24]. Apart from its information theoretic applications, it also appears to give a good characterization for the amount of correlation in spin systems [25]. However, there is again a priori no unique way in which such a quantity should be defined. Von Neumann mutual information can be defined in multiple ways in terms of the quantum relative entropy \( D(\rho||\sigma) \) [26], since

\[ I(A : B)_\rho = D(\rho_{AB}||\rho_A \otimes \rho_B) \]

\[ = \min_{\sigma_B} D(\rho_{AB}||\rho_A \otimes \sigma_B) \]

\[ = \min_{\sigma_A,\sigma_B} D(\rho_{AB}||\sigma_A \otimes \sigma_B), \]

where the minimizations run over all density operators \( \sigma_A \) and \( \sigma_B \) on \( \mathcal{H}_A \) and \( \mathcal{H}_B \), respectively. For other relative entropy measures these equalities do not hold in general. In fact, if we replace the quantum relative entropy with the max-relative entropy \( D_{\max}(\rho||\sigma) \), the values of the three expressions above can lie arbitrarily far apart [22]: while the expressions of the form (4) and (5) have a general upper bound given by \( 2 \cdot \log \min \{|A|,|B|\} \), the expression of the form (3) is unbounded. Furthermore, the expression of the form (4) is not symmetric in \( A \) and \( B \), unlike the von Neumann mutual information.

In order to consolidate and possibly unify these various approaches, it is of great interest to understand more about the relations among all these different quantities. In this paper, we show that smoothed versions of the max-information can be related to each other and regarded as approximately equivalent up to terms that depend only on the smoothing parameter and not on the specific quantum state or Hilbert space. These results can be employed to obtain chain rules in which we relate the max-information to differences of entropies as in (2). When evaluated for i.i.d.-states, these chain rules reproduce the well known relation (1) and thus imply the QAEP for the
max-information. Since max-information and min-entropy are formally related via their definitions in terms of the max-relative entropy, we can adapt proof techniques from earlier work on min-entropy.

The organization of the paper is as follows. In the next section we present the mathematical terminology and formal definitions necessary for the formulation of our results. Our results concerning the comparability of the different definitions and chain rules are summarized in sections III and IV. Longer proofs, along with useful technical results, can be found in the appendices.

II. MATHEMATICAL PRELIMINARIES

A. Basic Notations and Definitions

In this paper we deal exclusively with finite dimensional Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$ corresponding to physical systems $A, B$. To extend our results to infinite dimensional Hilbert spaces, the techniques of [27] could be used. For tensor products of Hilbert spaces, we use the short notation $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Let $\text{Herm}(\mathcal{H})$ be the space of Hermitian operators that act on $\mathcal{H}$ and $\mathcal{P}(\mathcal{H}) \subseteq \text{Herm}(\mathcal{H})$ the set of positive semi-definite operators on $\mathcal{H}$. For $A, B \in \text{Herm}(\mathcal{H})$ we write $A \geq B$ iff $A - B \in \mathcal{P}(\mathcal{H})$. In this sense, we will sometimes write $A \geq 0$ in order to state that $A \in \mathcal{P}(\mathcal{H})$. The sets of normalized and subnormalized density operators on $\mathcal{H}$ are defined as

$$S_\infty(\mathcal{H}) := \{ \rho \in \mathcal{P}(\mathcal{H}) : \text{tr}\rho = 1 \}$$

and

$$S_\leq(\mathcal{H}) := \{ \rho \in \mathcal{P}(\mathcal{H}) : 0 < \text{tr}\rho \leq 1 \},$$

respectively. Operators are usually written with a subscript that specifies on which system they act, e.g. $\rho_{AB} \in \text{Herm}(\mathcal{H}_{AB})$. Given an operator $O_{AB}$ on a composite Hilbert space $\mathcal{H}_{AB}$, we obtain the reduced operator $O_A$ on $\mathcal{H}_A$ by taking the partial trace over the subsystem $\mathcal{H}_B$: $O_A = \text{tr}_B O_{AB}$. The identity operator on $\mathcal{H}_A$ is denoted by $I_A$.

Quantum operations are represented by completely positive and trace preserving (CPTP) maps, i.e. linear maps $\mathcal{E} : S_{\leq}(\mathcal{H}) \mapsto S_{\leq}(\mathcal{H}')$ with the properties

$$\rho \geq 0 \Rightarrow \mathcal{E}(\rho) \geq 0,$$

and

$$\text{tr}\mathcal{E}(\rho) = \text{tr}\rho,$$

for all $\rho \in S_{\leq}(\mathcal{H})$. Note that the (partial) trace is a CPTP map.

Given any operator $O$, its operator norm $\|O\|_\infty$ is given by its maximal singular value. Its trace norm is defined as $\|O\|_1 := \text{tr}\sqrt{O^*O}$, where $O^*$ is the adjoint of $O$. We will also require a notion of distance between density operators. For this purpose, we make use of the generalized fidelity, which is defined as

$$F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1 + \sqrt{(1 - \text{tr}\rho)(1 - \text{tr}\sigma)},$$

for any $\rho, \sigma \in S_{\leq}(\mathcal{H})$. Note that when at least one of the states $\rho$ and $\sigma$ is normalized,

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1,$$

which corresponds to the standard definition for fidelity. We use the fidelity to define a distance measure on $S_{\leq}(\mathcal{H})$ as

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

which is a metric (Lemma 5 in [6]). $P(\rho, \sigma)$ is called the purified distance between $\rho$ and $\sigma$. We say that two states $\rho$ and $\sigma$ are $\varepsilon$-close and write $\rho \approx_\varepsilon \sigma$ iff $P(\rho, \sigma) \leq \varepsilon$.

Using the purified distance as a distance measure has many technical advantages. We summarize its essential properties, along with important properties of the fidelity in Appendix A.

For any given $\rho \in S_{\leq}(\mathcal{H})$, we can now define the ball of $\varepsilon$-close states around $\rho$ as

$$B^\varepsilon(\rho) := \{ \rho' \in S_{\leq}(\mathcal{H}) : P(\rho, \rho') \leq \varepsilon \},$$

where $\varepsilon$ is called the smoothing parameter and satisfies $0 \leq \varepsilon < \sqrt{\text{tr}\rho}$, since we want to exclude the zero operator from the ball. In all of our statements, we make the implicit assumption that the involved smoothing parameters are small enough in this sense.
B. Generalized Entropy Measures

Let us now give the definitions for two types of generalized relative entropy, the max- and the min-relative entropy [7].

Definition 1: For $\rho, \sigma \in \mathcal{P}(\mathcal{H})$, the max-relative entropy is defined as

$$D_{\max}(\rho\|\sigma) := \min \{ \lambda : 2^{\lambda} \sigma \geq \rho \}. \quad (6)$$

Note that $D_{\max}(\rho\|\sigma)$ to be well defined requires $\text{supp} \rho \subseteq \text{supp} \sigma$, where $\text{supp} O$ denotes the support of the operator $O$, i.e., the space orthogonal to the kernel of $O$. If this is satisfied, there is an alternative way to express the max-relative entropy that we use frequently [22]:

$$D_{\max}(\rho\|\sigma) = \log \left\| \sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \right\|_{\infty}. \quad (7)$$

The inverses here are generalized inverses: given $\sigma \in \mathcal{P}(\mathcal{H})$, its generalized inverse $\sigma^{-1}$ is the unique minimum rank operator such that $\sigma^0 := \sigma \sigma^{-1} = \sigma^{-1} \sigma$ is the projector onto $\text{supp} \sigma$.

Definition 2: For $\rho, \sigma \in \mathcal{P}(\mathcal{H})$, the min-relative entropy of $\rho$ with respect to $\sigma$ is

$$D_{\min}(\rho\|\sigma) := -\log \left\| \sqrt{\rho} \sqrt{\sigma} \right\|_{1}. \quad (8)$$

Given any $\rho_{AB} \in S_\leq(\mathcal{H}_{AB})$, we can now define the (conditional) min- and max-entropies as

$$H_{\min}(A|B)_\rho := -\min_{\sigma_B \in S_\leq(\mathcal{H}_B)} D_{\max}(\rho_{AB}||_A \otimes \sigma_B)$$

and

$$H_{\max}(A|B)_\rho := -\min_{\sigma_B \in S_\leq(\mathcal{H}_B)} D_{\min}(\rho_{AB}||_A \otimes \sigma_B),$$

along with their smoothed versions:

$$H_{\min}^\varepsilon(A|B)_\rho := \max_{\rho' \in B^c(\rho)} H_{\min}(A|B)_{\rho'},$$

and

$$H_{\max}^\varepsilon(A|B)_\rho := \min_{\rho' \in B^c(\rho)} H_{\max}(A|B)_{\rho'},$$

Min- and max-entropy are duals of each other in the sense that for pure $\rho_{ABC}$ [6]

$$H_{\min}^\varepsilon(A|B)_\rho = -H_{\max}^\varepsilon(A|C)_{\rho}. \quad (8)$$

If the system $B$ is trivial, we obtain the definitions for the non-conditional entropies:

$$H_{\min}(A)_\rho = -\log \lambda_{\max}(\rho_A),$$

where $\lambda_{\max}(\rho)$ is the largest eigenvalue of $\rho$, while

$$H_{\max}(A)_\rho = \log \left\| \sqrt{\rho_A} \right\|_1^2$$

presents a measure for the fidelity between $\rho_A$ and the completely mixed state on $\mathcal{H}_A$.

C. (Smooth) Max-Information

As argued before, there is no unique way in which generalized mutual information measures should be obtained from the introduced relative entropies. Based on [5]-[5], we define three different versions of max-information:

$$1 I_{\max}(A : B)_\rho := D_{\max}(\rho_{AB}||_A \otimes \rho_B),$$

$$2 I_{\max}(A : B)_\rho := \min_{\sigma_B \in S_\leq(\mathcal{H}_B)} D_{\max}(\rho_{AB}||_A \otimes \sigma_B),$$

$$3 I_{\max}(A : B)_\rho := \min_{\sigma_A \in S_\leq(\mathcal{H}_A)} \min_{\sigma_B \in S_\leq(\mathcal{H}_B)} D_{\max}(\rho_{AB}||_A \otimes \sigma_B).$$
For $\rho \in S_{\leq} (\mathcal{H}_{AB})$ and $\varepsilon \geq 0$, we obtain smooth max-information from $^i I_{\text{max}} (A : B)_{\rho}$ as
\[
^i I_{\text{max}}^\varepsilon (A : B)_{\rho} := \min_{\rho' \in B^\varepsilon(\rho)} ^i I_{\text{max}} (A : B)_{\rho'}.
\]

It should be pointed out that earlier literature making use of smooth max-information usually refers to $2^i I_{\text{max}}^\varepsilon (A : B)_{\rho}$. In particular, a chain rule, a data processing inequality and the QAEP have been proven for $2^i I_{\text{max}}$ in [22]. The proof of the data processing inequality can straightforwardly be extended to all smooth definitions.

**Lemma 1:** Let $\rho_{AB} \in S_{\leq} (\mathcal{H}_{AB})$, $\varepsilon \geq 0$ and let $E$ be a CPTP map of the form $E = E_A \otimes E_B$. Then
\[
^i I_{\text{max}}^\varepsilon (A : B)_{E(\rho)} \leq ^i I_{\text{max}}^\varepsilon (A : B)_{\rho},
\]
for any $i \in \{1, 2, 3\}$.

**Proof:** We provide the proof for $i = 2$, the other cases being similar. Let $\rho_{AB}' \in B^\varepsilon (\rho_{AB})$ be a state that optimizes $2^i I_{\text{max}}^\varepsilon (A : B)_{\rho}$, i.e. $2^i I_{\text{max}}^\varepsilon (A : B)_{\rho} = 2^i I_{\text{max}} (A : B)_{\rho'}$. Then there exists $\sigma_B \in S_{=}(\mathcal{H}_B)$ such that
\[
2^i I_{\text{max}}^\varepsilon (A : B)_{\rho} = D_{\text{max}} (\rho_{AB}' || \rho_A' \otimes \sigma_B)
\geq D_{\text{max}} (E(\rho_{AB}')) || \rho_A' \otimes E_B(\sigma_B))
\geq \min_{\omega_B \in S_{=}(\mathcal{H}_B)} D_{\text{max}} (\rho_{AB}' || \rho_A' \otimes \omega_B)
\geq \min_{\rho \in B^\varepsilon (E(\rho))} D_{\text{max}} (\rho_{AB}' || \rho_A \otimes \omega_B),
\]
where the first inequality follows from the data processing inequality for the max-relative entropy (cf. Lemma [20]) and the last inequality is a consequence of the monotonicity of the purified distance under trace non-increasing CPMs (cf. Lemma [12]).

**III. APPROXIMATE EQUIVALENCE RELATIONS FOR $^i I_{\text{max}}^\varepsilon$**

Let us now turn to our main problem of relating alternative expressions for smooth max-information to each other. Our key results are given by the following two theorems. For convenience of notation, we introduce the two functions
\[
f(\varepsilon, \varepsilon') := \log \left( \frac{1}{1 - \varepsilon^2} + \frac{1}{1 - \varepsilon'} \right)
\]
and
\[
g(\varepsilon) := \log \left( \frac{2(1 - \varepsilon) + 3}{(1 - \varepsilon)(1 - \sqrt{1 - \varepsilon^2})} \right).
\]

Note that both functions grow logarithmically in $\frac{1}{\varepsilon}$ as $\varepsilon \to 0$.

**Theorem 2:** Let $\rho_{AB} \in S_{=}(\mathcal{H}_{AB})$ and $\varepsilon > 0$, $\varepsilon' \geq 0$. Then
\[
3^i I_{\text{max}}^{\varepsilon+\varepsilon'} (A : B)_{\rho} \leq 2^i I_{\text{max}}^{\varepsilon+\varepsilon'} (A : B)_{\rho}
\leq 3^i I_{\text{max}}^{\varepsilon} (A : B)_{\rho} + f(\varepsilon, \varepsilon').
\]

**Theorem 3:** Let $\rho_{AB} \in S_{=}(\mathcal{H}_{AB})$ and $\varepsilon > 0$, $\varepsilon' \geq 0$. Then,
\[
2^i I_{\text{max}}^{\varepsilon+2\sqrt{\varepsilon}\varepsilon'} (A : B)_{\rho} \leq 1^i I_{\text{max}}^{\varepsilon+2\sqrt{\varepsilon}\varepsilon'} (A : B)_{\rho}
\leq 2^i I_{\text{max}}^{\varepsilon'} (A : B)_{\rho} + g(\varepsilon).
\]

We provide the proofs of these theorems in the appendix and turn immediately to the corollaries. First we complete our set of approximate equivalence relations. In order to compare $^1 I_{\text{max}}^\varepsilon$ and $^3 I_{\text{max}}^\varepsilon$, we only need to combine Theorems [2] and [3].

**Corollary 4:** Let $\rho_{AB} \in S_{=}(\mathcal{H}_{AB})$ and $\varepsilon, \varepsilon' > 0$, $\varepsilon'' \geq 0$. Then
\[
3^i I_{\text{max}}^{\sqrt{\varepsilon}\varepsilon'+\varepsilon''} (A : B)_{\rho} \leq 1^i I_{\text{max}}^{\sqrt{\varepsilon}\varepsilon'+\varepsilon''} (A : B)_{\rho}
\leq 3^i I_{\text{max}}^{\varepsilon} (A : B)_{\rho}
+ f(\varepsilon', \varepsilon'') + g(\varepsilon).
\]
where the maximization runs over all the apparent symmetry of $I_{\text{max}}^\varepsilon$. These relations further imply an estimate on the approximate symmetry of $2I_{\text{max}}^\varepsilon$.

**Corollary 5:** Let $\rho_{AB} \in S_\varepsilon(\mathcal{H}_{AB})$ and $\varepsilon > 0$, $\varepsilon' \geq 0$. Then
\[
2I_{\text{max}}^{\varepsilon+\varepsilon'}(A : B)_{\rho} \leq 2I_{\text{max}}^{\varepsilon'}(B : A)_{\rho} + f(\varepsilon, \varepsilon') \\
\leq 2I_{\text{max}}^{\varepsilon'}(A : B)_{\rho} + f(\varepsilon, \varepsilon') + f(\varepsilon, \varepsilon').
\]

**Proof:** Note that $2I_{\text{max}}^{\varepsilon'}(B : A)_{\rho} \geq 3I_{\text{max}}^{\varepsilon'}(A : B)_{\rho'}$, which follows directly from the definitions. Then, using Theorem 2 and the apparent symmetry of $3I_{\text{max}}^{\varepsilon'}$, we find that
\[
2I_{\text{max}}^{\varepsilon+\varepsilon'}(B : A)_{\rho} \leq 2I_{\text{max}}^{\varepsilon'}(A : B)_{\rho} + f(\varepsilon, \varepsilon'),
\]
and the claim follows.

IV. CHAIN RULES FOR $I_{\text{max}}^\varepsilon$

In this section we prove chain rules for smooth max-information of the form
\[
H_{\text{min}}^\varepsilon(A)_{\rho} - H_{\text{min}}^\varepsilon(A|B)_{\rho} \lesssim I_{\text{max}}^\varepsilon(A : B)_{\rho},
\]
An upper bound chain rule for $2I_{\text{max}}^\varepsilon$ is already known from Lemma B.15 in [22]: for $\rho \in S_\varepsilon(\mathcal{H})$ and $\varepsilon > 0$,
\[
2I_{\text{max}}^\varepsilon(A : B)_{\rho} \leq H_{\text{min}}^{\varepsilon^2/48}(A)_{\rho} - H_{\min}^{\varepsilon^2/48}(A|B)_{\rho} - l(\varepsilon),
\]
where $l(\varepsilon) := 2 \cdot \log \frac{e^2}{24}$. Let us first derive a lower bound chain rule for $2I_{\text{max}}^\varepsilon$. Having both bounds for one of the definitions will allow us to write down chain rules for all $iI_{\text{max}}^\varepsilon$ by exploiting the approximate equivalence relations from the previous section.

**Lemma 6:** Let $\rho \in S_\varepsilon(\mathcal{H}_{AB})$ and $\varepsilon \geq 0$. Then
\[
2I_{\text{max}}^\varepsilon(A : B)_{\rho} \geq H_{\text{min}}^\varepsilon(A)_{\rho} - \Pi_{\varepsilon}^{\varepsilon^2/32}(A : B)_{\rho}.
\]

**Proof:** The proof is similar to the one of (14) in [22]. We rearrange Lemma B.13 from [22] as
\[
H_{\text{min}}(A|B)_{\rho} \geq H_{\text{min}}(A)_{\rho} - 2I_{\text{max}}(A : B)_{\rho}.
\]

Thus,
\[
H_{\text{min}}^\varepsilon(A|B)_{\rho} \geq \max_{\rho' \in \mathcal{B}^\varepsilon(\rho)} \left[ H_{\text{min}}(A)_{\rho'} - 2I_{\text{max}}(A : B)_{\rho'} \right] \\
\geq \max_{\omega \in \mathcal{B}^{\varepsilon^2/32}(\rho)} \max_{\Pi_A} \left[ H_{\text{min}}(A)_{\Pi_{\varepsilon}^{\varepsilon^2/32}} - 2I_{\text{max}}(A : B)_{\Pi_{\varepsilon}^{\varepsilon^2/32}} \right],
\]
where the maximization runs over all $0 \leq \Pi_A \leq \Pi_A$ with $\Pi_A \omega \Pi_A \approx \varepsilon/2\omega$. Next, choose $\omega' \in \mathcal{B}^{\varepsilon^2/32}(\rho)$ such that $2I_{\text{max}}(A : B)_{\omega'} = 2I_{\text{max}}^{\varepsilon^2/32}(A : B)_{\rho}$. This gives us
\[
H_{\text{min}}^\varepsilon(A|B)_{\rho} \geq \max_{\Pi_A} \left[ H_{\text{min}}(A)_{\Pi_{\varepsilon}^{\varepsilon^2/32}} - 2I_{\text{max}}(A : B)_{\Pi_{\varepsilon}^{\varepsilon^2/32}} \right] \\
\geq \max_{\Pi_A} H_{\text{min}}(A)_{\Pi_{\varepsilon}^{\varepsilon^2/32}} - 2I_{\text{max}}(A : B)_{\omega'}.\]
with the maximization running over all \(0 \leq \Pi_A \leq \mathbb{I}_A\) with \(\Pi_A \omega' \Pi_A \approx \varepsilon/2\omega'\). The second inequality is a consequence of Remark 1 (cf. Appendix B). According to Lemma 19 we can choose a \(\Pi_A\) such that \(H_{\min}^{\varepsilon^2/16}(A)_{\omega'} \leq H_{\min}(A)_{\Pi_A \omega' \Pi_A}\). Doing so yields

\[
H_{\min}^\varepsilon(A|B)_\rho \geq H_{\min}^{\varepsilon^2/16}(A)_{\omega'} - 2I_{\max}^{\varepsilon^2/32}(A : B)_\rho \\
\geq H_{\min}^{\varepsilon^2/32}(A)_{\rho} - 2I_{\max}^{\varepsilon^2/32}(A : B)_{\rho}.
\]

Relabelling \(\varepsilon^2/32 \rightarrow \varepsilon\) concludes the proof.

We can now obtain chain rules for alternative definitions of \(I_{\max}^\varepsilon\) as well.

**Corollary 7:** Let \(\rho \in S_\varepsilon(\mathcal{H}_{AB})\) and \(\varepsilon, \varepsilon' > 0\). Then

\[
\frac{1}{2}I_{\max}^{\varepsilon^2 + 2\varepsilon + \varepsilon'}(A : B)_\rho \leq H_{\max}^{\varepsilon^2/48}(A)_{\rho} - H_{\min}^{\varepsilon^2/48}(A|B)_\rho \\
+ g(\varepsilon) - l(\varepsilon').
\]

**Proof:** With (14) and using Theorem 3 to estimate \(I_{\max}^{\varepsilon^2 + 2\varepsilon + \varepsilon'}(A : B)_\rho\) in terms of \(2I_{\max}^{\varepsilon'}(A : B)_{\rho}\), the claim follows immediately.

Similarly, we obtain a lower bound chain rule for \(3I_{\max}^{\varepsilon'}(A : B)_{\rho}\).

**Corollary 8:** For \(\rho \in S_\varepsilon(\mathcal{H}_{AB})\) and \(\varepsilon > 0, \varepsilon' \geq 0\),

\[
3I_{\max}^{\varepsilon'}(A : B)_\rho \geq H_{\min}^{\varepsilon + \varepsilon'}(A)_{\rho} - H_{\min}^{\varepsilon^2 + 2\varepsilon^2 + 2\varepsilon'(\varepsilon')}(A|B)_\rho \\
- f(\varepsilon, \varepsilon').
\]

**Proof:** The claim is a direct consequence of Theorem 2 and Lemma 6.

---

**V. Conclusion**

We have investigated properties of smooth max-information defined as a special case of the max-relative entropy. In earlier work, it has been shown to be an operational quantity in one-shot state splitting and state merging [22], [23]. It is also found to be a useful quantity in quantum rate distortion theory [24] and the statistical physics of many body systems [25]. We have shown that it exhibits some properties that we would expect from previous results on smooth entropies. Alternative definitions of max-information turn out to be essentially equivalent. Furthermore, they satisfy upper and lower bound chain rules in terms of min- and max-entropies. Chain rules are generally an important technical tool in information theory. In this case, they also relate max-information to alternative definitions for one-shot mutual information, made up from differences of entropies as used in [19]–[21].

The primary goal of further research on these quantities is to gain a better understanding of their operational relevance. We hope that the formal tools provided in this paper will be useful for this purpose.

**Acknowledgments**

NC would like to thank Joseph M. Renes and Frédéric Dupuis for helpful discussions.

**Appendix A**

**Properties of the Fidelity and Purified Distance**

Here we summarize the essential properties of the purified distance. For a more extensive discussion, we refer the reader to [8]. The main reasons the purified distance is preferred over the trace distance are the following two lemmas, which state that for given \(\rho, \sigma\), we can always find purifications or extensions \(\tilde{\rho}, \tilde{\sigma}\) such that \(P(\rho, \sigma) = P(\tilde{\rho}, \tilde{\sigma})\).

This is due to Uhlmann’s theorem [23]: for any states \(\rho_A, \sigma_A \in S_\varepsilon(\mathcal{H}_A)\)

\[
\|\sqrt{\rho_A} \sqrt{\sigma_A}\|_1 = \max_{\rho_{AB}, \sigma_{AB}} \|\sqrt{\rho_{AB}} \sqrt{\sigma_{AB}}\|_1,
\]

where the maximization runs over all purifications \(\rho_{AB}\) and \(\sigma_{AB}\) of \(\rho_A\) and \(\sigma_A\).

**Lemma 9 (Lemma 8 in [8]):** Let \(\rho, \sigma \in S_\varepsilon(\mathcal{H})\), \(\mathcal{H}' \cong \mathcal{H}\) and \(\varphi \in \mathcal{H} \otimes \mathcal{H}'\) be a purification of \(\rho\). Then there exists a purification \(\vartheta \in \mathcal{H} \otimes \mathcal{H}'\) of \(\sigma\) with \(P(\rho, \sigma) = P(\varphi, \vartheta)\).

**Lemma 10 (Corollary 9 in [8]):** Let \(\rho, \sigma \in S_\varepsilon(\mathcal{H})\) and \(\tilde{\rho} \in S_\varepsilon(\mathcal{H} \otimes \mathcal{H}')\) be an extension of \(\rho\). Then there exists an extension \(\tilde{\sigma} \in S_\varepsilon(\mathcal{H} \otimes \mathcal{H}')\) of \(\sigma\) with \(P(\rho, \sigma) = P(\tilde{\rho}, \tilde{\sigma})\).
Still, the purified distance is equivalent to the generalized trace distance given by
\[ D(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1 + \frac{1}{2} |\text{tr}\rho - \text{tr}\sigma|. \]
Therefore it retains an operational interpretation as a measure for the maximum guessing probability \[29\]: the maximal probability \( p_{\text{dist}}(\rho, \sigma) \) for correctly distinguishing between two quantum states \( \rho, \sigma \) satisfies
\[ p_{\text{dist}}(\rho, \sigma) \leq \frac{1}{2}(1 + D(\rho, \sigma)). \]

Lemma 11 (Lemma 7 in \[6\]): Let \( \rho, \sigma \in S_\leq(\mathcal{H}) \). Then
\[ D(\rho, \sigma) \leq P(\rho, \sigma) \leq \sqrt{2D(\rho, \sigma)}. \quad (A.20) \]

Another useful property of the purified distance is that it cannot increase under trace non-increasing CPMs.

Lemma 12 (Lemma 7 in \[6\]): Let \( \rho, \sigma \in S_\leq(\mathcal{H}) \) and let \( \mathcal{E} \) be a trace non-increasing CPM. Then \( P(\rho, \sigma) \geq P(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \).

We make use of the following lemmas of the standard fidelity.

Lemma 13 (\[8\]): Let \( \rho, \sigma \in \mathcal{P}(\mathcal{H}) \).
- For any \( \omega \geq \rho \),
  \[ \|\sqrt{\omega}\sqrt{\sigma}\|_1 \geq \|\sqrt{\rho}\sqrt{\sigma}\|_1. \quad (A.21) \]
- For any projector \( \Pi \in \mathcal{P}(\mathcal{H}) \),
  \[ \left\| \sqrt{\Pi\rho\Pi}\sqrt{\sigma}\right\|_1 = \left\| \sqrt{\rho}\sqrt{\Pi\sigma}\right\|_1 \]
  \[ = \left\| \sqrt{\Pi\rho}\sqrt{\Pi\sigma}\right\|_1. \quad (A.22) \]

We conclude this section by stating a few useful technical facts.

Lemma 14 (Lemma 17 in \[9\]): Let \( \rho \in S_\leq(\mathcal{H}) \) and \( \Pi \) a projector on \( \mathcal{H} \), then
\[ P(\rho, \Pi\rho\Pi) \leq \sqrt{2 \cdot \text{tr}(\Pi^\perp \rho) - (\text{tr}(\Pi^\perp \rho))^2}, \quad (A.23) \]
where \( \Pi^\perp = I - \Pi \).

Lemma 15 (Lemma A.7 in \[30\]): Let \( \rho \in S_\leq(\mathcal{H}) \) and \( \Pi \in \mathcal{P}(\mathcal{H}) \) such that \( \Pi \leq I \). Then
\[ P(\rho, \Pi\rho\Pi) \leq \frac{1}{\text{tr}\rho} \sqrt{\left(\text{tr}(\rho)\right)^2 - (\text{tr}(\Pi^2 \rho))^2}. \quad (A.24) \]

Corollary 16: Let \( \rho \in S_\leq(\mathcal{H}) \) and \( 0 < k \leq 1 \). Then
\[ P(\rho, k \cdot \rho) \leq \sqrt{1 - k^2}. \quad (A.25) \]

Proof: Apply Lemma 15 to \( \Pi = \sqrt{k} \cdot I \) and use \( \sqrt{\text{tr}\rho} \leq 1. \)

Appendix B

Technical Lemmas

The following lemma introduces a notion of duality between projectors on subsystems of a multi-partite quantum system with respect to a given pure state. It is essential in the proofs of Theorems 2 and 3.

Lemma 17 (Corollary 16 in \[9\]): Let \( \rho_{AB} = |\varphi\rangle\langle\varphi|_{AB} \in \mathcal{P}(\mathcal{H}_{AB}) \) be pure, \( \rho_A = \text{tr}_B\rho_{AB}, \rho_B = \text{tr}_A\rho_{AB} \) and let \( \Pi_A \in \mathcal{P}(\mathcal{H}_A) \) be a projector in \( \text{supp}\rho_A \). Then, there exists a dual projector \( \Pi_B \) on \( \mathcal{H}_B \) such that
\[ (\Pi_A \otimes \rho_B^{-1/2})|\varphi\rangle_{AB} = (\rho_A^{-1/2} \otimes \Pi_B)|\varphi\rangle_{AB}. \quad (B.26) \]

The proof of Theorem 3 further requires the following inequality for the operator norm.

Lemma 18: Let \( A, B, C \in \mathcal{P}(\mathcal{H}) \) be such that \( \text{supp}\ A \subseteq \text{supp}\ B \) and \( B \leq C \). Then
\[ \left\| C^{-1/2}AC^{-1/2} \right\|_\infty \leq \left\| B^{-1/2}AB^{-1/2} \right\|_\infty. \quad (B.27) \]
Proof: We know from (7) that \( \lambda = \| B^{-1/2} A B^{-1/2} \|_\infty \) is the smallest number such that \( A \leq \lambda B \). Then \( B \leq C \) implies \( A \leq \lambda C \) and the claim follows.

In proving the chain rules for \( I_{\max}^\varepsilon \), we have used the following facts on different entropic quantities.

**Lemma 19** (Lemma 5 in [31]): For any \( \rho \in S_e(\mathcal{H}_A) \) and \( \varepsilon \geq 0 \), there exists an operator \( 0 \leq \Pi \leq \mathbb{I}_A \) such that \( \rho \approx_{\varepsilon/2} \Pi \rho \Pi \) and

\[
H_{\min}^{2^{16}}(A)_\rho \leq H_{\min}(A)_{\Pi \rho \Pi}.
\]  

**Lemma 20** (Lemma 7 in [7]): Let \( \rho, \sigma \in \mathcal{P}(\mathcal{H}) \) and \( \mathcal{E} \) be a CPTP map on \( \mathcal{H} \). Then

\[
D_{\max}(\rho\|\sigma) \geq D_{\max}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)).
\]  

**Remark 1:** This actually holds more generally even if the CPM \( \mathcal{E} \) is not trace preserving. In particular, for any \( \Pi \in \mathcal{P}(\mathcal{H}) \),

\[
D_{\max}(\rho\|\sigma) \geq D_{\max}(\Pi \rho \Pi\|\Pi \sigma \Pi).
\]  

**Appendix C**

**Proofs of Theorems 2 and 3**

### A. Auxiliary Lemmas

Before turning to the main proofs, we want to make a few observations on the normalization of optimal operators for \( I_{\max}^\varepsilon \). Lemma 22 will prove especially useful in the proof of Theorem 3.

The proofs of these lemmas rely on the following fact.

**Lemma 21:** Let \( \rho_{AB} \in S_\varepsilon(\mathcal{H}_{AB}) \), \( \varepsilon \geq 0 \) and \( \rho' \in \mathcal{B}_e(\rho_{AB}) \). Then \( \frac{\rho'}{\text{tr}\rho'} \in \mathcal{B}_e(\rho_{AB}) \) as well.

**Proof:** Remember that the generalized fidelity \( F(\sigma, \tau) \) is equal to \( \| \sqrt{\sigma} \sqrt{\tau} \|_1 \) if at least one of the arguments is normalized. Note also that every subnormalized operator \( \omega' \) can be written as \( \omega' = \text{tr}\omega' \cdot \omega \) with a normalized operator \( \omega \). Let \( \omega_{AB} = \frac{\rho_{AB}}{\text{tr}\rho_{AB}} \). Then

\[
F(\rho_{AB}', \rho_{AB}) = \left\| \sqrt{\rho_{AB}' \rho_{AB}} \right\|_1
\]

\[
= \left\| \sqrt{\text{tr}\rho_{AB}' \cdot \omega_{AB} \rho_{AB}} \right\|_1
\]

\[
\leq \left\| \omega_{AB} \rho_{AB} \right\|_1 = F(\omega_{AB}, \rho_{AB}).
\]

Therefore, the purified distance is

\[
P(\rho_{AB}', \rho_{AB}) = \sqrt{1 - F^2(\rho_{AB}', \rho_{AB})}
\]

\[
\geq \sqrt{1 - F^2(\omega_{AB}, \rho_{AB})}
\]

\[
= P(\omega_{AB}, \rho_{AB}),
\]

which concludes the proof.

With this lemma, we can show that there always exists an optimal operator for \( 2I_{\max}^\varepsilon(A : B)_\rho \) that is normalized.

**Lemma 22:** Let \( \rho \in S_\varepsilon(\mathcal{H}_{AB}) \) and \( \varepsilon \geq 0 \). Then there exists a normalized state \( \rho' \in \mathcal{B}_e(\rho) \) with \( 2I_{\max}(A : B)_{\rho} = 2I_{\max}(A : B)_{\rho'} \).

**Proof:** Let \( \tilde{\rho}_{AB} \in \mathcal{B}_e(\rho) \) be any operator satisfying \( 2I_{\max}(A : B)_{\rho} = 2I_{\max}(A : B)_{\rho} \) and let \( \sigma_B \in S_\varepsilon(\mathcal{H}_B) \) be such that

\[
2^{2I_{\max}(A : B)}_{\rho} \tilde{\rho}_A \otimes \sigma_B \geq \tilde{\rho}_{AB}
\]

\[
\Rightarrow 2^{2I_{\max}(A : B)}_{\rho} \frac{\tilde{\rho}_A}{\text{tr}\tilde{\rho}_{AB}} \otimes \sigma_B \geq \frac{\tilde{\rho}_{AB}}{\text{tr}\tilde{\rho}_{AB}}.
\]

Hence,

\[
2^{2I_{\max}(A : B)}_{\rho} \geq D_{\max} \left( \frac{\tilde{\rho}_{AB}}{\text{tr}\tilde{\rho}_{AB}} \| \frac{\tilde{\rho}_A}{\text{tr}\tilde{\rho}_{AB}} \otimes \sigma_B \right).
\]

but because of Lemma 21, we find that actually equality holds. We thus conclude that if \( \tilde{\rho} \) optimizes \( 2^{2I_{\max}(A : B)}_{\rho} \), then so does \( \rho' = \frac{\tilde{\rho}}{\text{tr}\tilde{\rho}}. \)
We can prove an analogous and in fact stricter statement about $I^e_{\max} (A : B)$. We give it here for the sake of completeness.

**Lemma 23.** Let $\rho_{AB} \in S_\varepsilon (H_{AB})$, $\varepsilon \geq 0$ and let $\rho'_{AB} \in B^c (\rho_{AB})$ optimize $I^e_{\max} (A : B)$. Then $tr \rho'_{AB} = 1$.

**Proof:** Let $\omega_{AB} = \frac{\rho'_{AB}}{tr \rho'_{AB}}$. It holds that for $k = 2 I^e_{\max} (A : B)$

$$k \cdot \rho'_{A} \otimes \rho'_{B} \geq \rho'_{AB}$$

$$\Rightarrow k \cdot tr \rho'_{AB} \cdot \omega_{A} \otimes \omega_{B} \geq \frac{\rho'_{AB}}{tr \rho'_{AB}} = \omega_{AB}$$

$$\Rightarrow I^e_{\max} (A : B) + \log tr \rho'_{AB} \geq I^e_{\max} (A : B) \omega,$$

where $\log tr \rho'_{AB} \leq 0$. If however this inequality is strict, i.e. $\log tr \rho'_{AB} < 0$, this would be a contradiction to the optimality of $\rho'_{AB}$ according to Lemma 21 and therefore $tr \rho'_{AB} = 1$.

**B. Proof of Theorem 2**

The proof is analogous to the reasoning in Lemma 20 in [9]. We divide it into three steps. Claim 1 is a crucial step in the proof of Claim 2 from which in turn the result follows.

**Claim 1.** Let $\rho_{ABC}$ be a purification of $\rho_{AB} \in S_\varepsilon (H_{AB})$ and $\varepsilon > 0$. Then there exists a projector $\Pi_{BC}$ on $H_{BC}$ such that $\rho_{ABC} := \Pi_{BC} \rho_{ABC} \Pi_{BC} \in B^c (\rho_{ABC})$ and

$$\max_{\sigma_b \in S_c (H_b)} D_{\max} (\rho_{AB} \| \sigma_A \otimes \sigma_B) \leq \min_{\sigma_a \in S_c (H_a)} D_{\max} (\rho_{AB} \| \sigma_A \otimes \sigma_B) + \log \frac{1}{1 - \sqrt{1 - \varepsilon^2}}.$$  

(C.31)

**Proof:** The strategy of the proof is to define $\Pi_{BC}$ as the dual projector with respect to $\rho_{ABC}$ (in the sense of Lemma 17) of a conveniently chosen $\Pi_A$ with $\text{supp} \Pi_A \subseteq \text{supp} \rho_A$. Fix $\lambda, \sigma_A$, and $\sigma_B$ such that

$$\min_{\sigma_a \in S_c (H_a)} D_{\max} (\rho_{AB} \| \sigma_A \otimes \sigma_B)$$

$$= \min_{\sigma_a \in S_c (H_a)} D_{\max} (\rho_{AB} \| \sigma_A \otimes \sigma_B) = \log \lambda.$$  

Note that by construction we have $\bar{\rho}_A \leq \rho_A$ and $\text{supp} \bar{\rho}_A \subseteq \text{supp} \rho_A$, so that we find

$$\text{supp} \bar{\rho}_{AB} \subseteq \text{supp} (\bar{\rho}_A \otimes \bar{\rho}_B)$$

$$\subseteq \text{supp} (\rho_A \otimes \rho_B)$$

$$\subseteq \text{supp} (\rho_A \otimes \sigma_B).$$

Therefore $D_{\max} (\bar{\rho}_{AB} \| \rho_A \otimes \sigma_B)$ is well defined and we can write

$$\min_{\sigma_b \in S_c (H_b)} D_{\max} (\bar{\rho}_{AB} \| \rho_A \otimes \sigma_B)$$

$$\leq D_{\max} (\bar{\rho}_{AB} \| \rho_A \otimes \sigma_B)$$

$$= \log \left\| \frac{\rho_A}{\sigma_A} \otimes \frac{\rho_B}{\sigma_B} \bar{\rho}_{AB} \rho_A \otimes \sigma_B \right\|_\infty.$$  

Defining $\Pi_A$ as the dual projector of $\Pi_{BC}$ and using the inequality $\lambda \sigma_A \otimes \sigma_B \geq \bar{\rho}_{AB}$ we obtain

$$\left\| \frac{\rho_A}{\sigma_A} \otimes \frac{\rho_B}{\sigma_B} \bar{\rho}_{AB} \rho_A \otimes \sigma_B \right\|_\infty$$

$$= \left\| \sigma_B \frac{\rho_A}{\sigma_A} \Pi_{BC} \rho_A \rho_A \otimes \sigma_B \right\|_\infty$$

$$= \left\| \sigma_B \frac{\rho_A}{\sigma_A} \Pi_{BC} \rho_A \rho_A \otimes \sigma_B \right\|_\infty$$

$$\leq \lambda \left\| \sigma_B \frac{\rho_A}{\sigma_A} \Pi_{BC} \rho_A \rho_A \otimes \sigma_B \right\|_\infty$$

$$= \lambda \left\| \Pi_{BC} \rho_A \rho_A \otimes \sigma_B \right\|_\infty$$

$$= \lambda \left\| \Pi_{BC} \rho_A \rho_A \otimes \sigma_B \right\|_\infty.$$
where we have introduced \( \Gamma_A := \rho_A^{\frac{1}{2}} \sigma_A \rho_A^{\frac{1}{2}} \). Thus, we find

\[
\min_{\sigma_B \in S_\pi(H_B)} D_{\text{max}}(\tilde{\rho}_{AB} \| \rho_A \otimes \sigma_B) 
\leq \min_{\sigma_B \in S_\pi(H_B)} D_{\text{max}}(\rho_{AB} \| \sigma_A \otimes \sigma_B) + \log \|\Pi_A \Gamma_A \Pi_A\|_\infty .
\]

By Lemma 14 it holds that

\[
P(\rho_{ABC}, \tilde{\rho}_{ABC}) \leq \sqrt{2 \cdot tr(\Pi_{BC}^\perp \rho_{ABC}) - (tr(\Pi_{BC}^\perp \rho_{ABC}))^2} = \sqrt{2 \cdot tr(\Pi_A^\perp \rho_A) - (tr(\Pi_A^\perp \rho_A))^2},
\]

where \( \Pi_{BC}^\perp = \mathbb{I} - \Pi \). Now we define \( \Pi_A \) to be the minimum rank projector on the smallest eigenvalues of \( \Gamma_A \) such that \( tr(\Pi_A^\perp \rho_A) \leq 1 - \sqrt{1 - \varepsilon^2} \). With (C.32) this implies \( P(\rho_{ABC}, \tilde{\rho}_{ABC}) \leq \varepsilon \) since \( t \mapsto \sqrt{2t - t^2} \) is monotonically increasing in the interval \([0, 1]\). It remains to show that with our choice of \( \Pi_A \)

\[
\|\Pi_A \Gamma_A \Pi_A\|_\infty \leq \frac{1}{1 - \sqrt{1 - \varepsilon^2}}
\]

holds. This, however, can be shown in an identical manner as it is done in the proof of Lemma 21 in \([9]\). The only difference is that we have chosen \( \Pi_A \) such that \( tr(\Pi_A^\perp \rho_A) \leq 1 - \sqrt{1 - \varepsilon^2} \), instead of \( tr(\Pi_A^\perp \rho_A) \leq \frac{\varepsilon^2}{2} \), which eventually leads to slightly tighter correction terms.

Claim 2: For any \( \rho_{AB} \in S_\pi(H_{AB}) \) there exists a state \( \tilde{\rho}_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB}) \) that satisfies

\[
\min_{\sigma_B \in S_\pi(H_B)} D_{\text{max}}(\tilde{\rho}_{AB} \| \rho_A \otimes \sigma_B) 
\leq \min_{\sigma_B \in S_\pi(H_B)} D_{\text{max}}(\rho_{AB} \| \sigma_A \otimes \sigma_B) + c(\varepsilon, \rho_{AB}),
\]

where \( c(\varepsilon, \rho_{AB}) := \log \left( \frac{1}{1 - \sqrt{1 - \varepsilon^2}} + \frac{1}{\text{tr} \rho_{AB}} \right) \).

**Proof:** Let \( \lambda, \tilde{\sigma}_A, \tilde{\sigma}_B \) be such that

\[
\min_{\sigma_B \in S_\pi(H_B)} D_{\text{max}}(\rho_{AB} \| \sigma_A \otimes \sigma_B) = D_{\text{max}}(\rho_{AB} \| \tilde{\sigma}_A \otimes \tilde{\sigma}_B) = \log \lambda.
\]

Let us also define the positive semi-definite operator \( \Delta_A := \rho_A - \tilde{\rho}_A \) and set \( \bar{\rho}_{AB} = \tilde{\rho}_{AB} + \Delta_A \otimes \tilde{\sigma}_B \). It holds that \( \tilde{\rho}_A = \rho_A \) and \( \bar{\rho}_{AB} \approx \varepsilon \rho_{AB} \), which can be seen as follows: since \( \bar{\rho}_{AB} \leq \tilde{\rho}_{AB} \), it also holds that \( \| \sqrt{\rho_{AB}} \sqrt{\bar{\rho}_{AB}} \|_1 \leq \| \sqrt{\rho_{AB}} \sqrt{\bar{\rho}_{AB}} \|_1 \). Hence,

\[
F(\bar{\rho}_{AB}, \rho_{AB}) \geq \| \sqrt{\rho_{AB}} \sqrt{\bar{\rho}_{AB}} \|_1 + 1 - \text{tr} \rho_{AB}
\]

\[
\geq \| \sqrt{\rho_{ABC}} \sqrt{\rho_{ABC}} \|_1 + 1 - \text{tr} \rho_{AB}
\]

\[
= 1 - \text{tr} (\Pi_{BC}^\perp \rho_{BC})
\]

\[
\geq \sqrt{1 - \varepsilon^2},
\]

and thus \( P(\bar{\rho}_{AB}, \rho_{AB}) \leq \varepsilon \). The equality in the penultimate line is a consequence of (A.22) in Lemma 13.
Finally, we use $\bar{\rho}_A = \rho_A$ and $\bar{\rho}_{AB} \leq \bar{\rho}_{AB} + \rho_A \otimes \bar{\sigma}_B$ to find
\[
\min_{\sigma_B \in S_\infty(\mathcal{H}_B)} D_{\max}(\bar{\rho}_{AB} \| \rho_A \otimes \sigma_B) \\
\leq \log \left\| \frac{1}{2} \otimes \bar{\sigma}_B - \frac{1}{2} \bar{\rho}_{AB} \rho_A \frac{1}{2} \otimes \bar{\sigma}_B \frac{1}{2} \right\|_{\infty} \\
\leq \log \left\| \frac{1}{2} \otimes \bar{\sigma}_B - \frac{1}{2} (\bar{\rho}_{AB} + \rho_A \otimes \bar{\sigma}_B) \frac{1}{2} \otimes \bar{\sigma}_B \frac{1}{2} \right\|_{\infty} \\
= \log \left\| \frac{1}{2} \otimes \bar{\sigma}_B - \frac{1}{2} \bar{\rho}_{AB} \rho_A \frac{1}{2} \otimes \bar{\sigma}_B \frac{1}{2} + \rho_A \otimes \bar{\sigma}_B \right\|_{\infty} \\
\leq \log \left( \lambda \frac{1}{1 - \sqrt{1 - \varepsilon^2}} + 1 \right).
\]
The first inequality is justified, as
\[
\text{supp } \bar{\rho}_B = \text{supp } (\bar{\rho}_B + \text{tr}(\Delta_A) \cdot \bar{\sigma}_B) \subseteq \text{supp } \bar{\sigma}_B.
\]
Since $\lambda \geq \text{tr} \rho_{AB}$, we conclude
\[
\min_{\sigma_B \in S_\infty(\mathcal{H}_B)} D_{\max}(\bar{\rho}_{AB} \| \rho_A \otimes \sigma_B) \\
\leq \log \lambda + \log \left( \frac{1}{1 - \sqrt{1 - \varepsilon^2}} + \frac{1}{\text{tr} \rho_{AB}} \right),
\]
thus completing the proof of Claim 2.

It is now straightforward to prove the upper bound in the theorem, the lower bound given by
\[
3 I_{\text{max}}^{\varepsilon + \varepsilon'} (A : B)_\rho \\
\leq 2 I_{\text{max}}^{\varepsilon + \varepsilon'} (A : B)_\rho
\]
being clear from the definitions. Let $\rho'_{AB} \in \mathcal{B}^{\varepsilon'}(\rho_{AB})$ be the operator that optimizes $3 I_{\text{max}}^{\varepsilon'} (A : B)_\rho$. Then, by Claim 2, there exists an operator $\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon + \varepsilon'}(\rho_{AB})$ such that
\[
\min_{\sigma_B \in S_\infty(\mathcal{H}_B)} D_{\max}(\bar{\rho}_{AB} \| \rho_A \otimes \sigma_B) \\
\leq 3 I_{\text{max}}^{\varepsilon'} (A : B)_\rho + \log \left( \frac{1}{1 - \sqrt{1 - \varepsilon^2}} + \frac{1}{\text{tr} \rho'_{AB}} \right) \\
\leq 3 I_{\text{max}}^{\varepsilon'} (A : B)_\rho + \log \left( \frac{1}{1 - \sqrt{1 - \varepsilon^2}} + \frac{1}{1 - \varepsilon'} \right).
\]
It remains to notice that by definition of $2 I_{\text{max}}^{\varepsilon}$
\[
2 I_{\text{max}}^{\varepsilon + \varepsilon'} (A : B)_\rho \\
\leq \min_{\sigma_B \in S_\infty(\mathcal{H}_B)} D_{\max}(\bar{\rho}_{AB} \| \rho_A \otimes \sigma_B),
\]
which concludes the proof of Theorem 2.

C. Proof of Theorem 3

The derivation of the equivalence between $2 I_{\text{max}}^{\varepsilon}$ and $I_{\text{max}}^{\varepsilon}$ is very similar to the one of Theorem 2 and is therefore not carried out with all of its details here. Again, we only need to prove the upper bound of the theorem, since
\[
2 I_{\text{max}}^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'} (A : B)_\rho \\
\leq 1 I_{\text{max}}^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'} (A : B)_\rho
\]
follows directly from the definitions of the quantities.

We find the following fact, analogous to Claim 1

Claim 3: Let $\rho_{ABC}$ be a purification of $\rho_{AB} \in S_\leq(\mathcal{H}_{AB})$ and $\varepsilon > 0$. Then there exists a projector $\Pi_{AC}$ on $\mathcal{H}_{AC}$ such that $\bar{\rho}_{ABC} := \Pi_{AC} \rho_{ABC} \Pi_{AC} \in \mathcal{B}^{\varepsilon}(\rho_{ABC})$ and
\[
D_{\max}(\bar{\rho}_{AB} \| \rho_A \otimes \rho_B) \\
\leq \min_{\sigma_B \in S_\infty(\mathcal{H}_B)} D_{\max}(\rho_{AB} \| \rho_A \otimes \sigma_B) + \log \frac{1}{1 - \sqrt{1 - \varepsilon^2}}.
\]
(C.34)
Proof: The proof of this claim can straightforwardly be adapted from the proof of Claim 1. We then end up estimating a term \(\|\Pi_B \Gamma_B \Pi_B\|_\infty\) (with \(\Gamma_B = \rho_B^{-\frac{1}{2}} \sigma_B \rho_B^{-\frac{1}{2}}\)) on system B instead of A. As before, we can choose \(\Pi_B\) such that its dual \(\Pi_{AC}\) satisfies \(\text{tr}(\Pi_{AC} \rho_{ABC}) \leq 1 - \sqrt{1 - \epsilon^2}\). 

In the following it is sufficient for our purposes to assume that \(\rho_{AB}\) is normalized, thanks to Lemma 22. Now define \(\Delta_{ABC} := \rho_{ABC} - \tilde{\rho}_{ABC}\) and
\[
\tilde{\rho}_{AB} := k \cdot (\tilde{\rho}_{AB} + \rho_A \otimes \Delta_B + \Delta_A \otimes \rho_B),
\]
with \(k := \frac{1}{1 + \text{tr}(\Delta_A)}\). Notice that \(\Delta_B \geq 0\), but \(\Delta_A\) and thus \(\tilde{\rho}_{AB}\) is not necessarily positive semi-definite. However, \(\text{tr}(\tilde{\rho}_{AB}) = 1\) and by construction we have that \(\tilde{\rho}_A = \rho_A\) and \(\tilde{\rho}_B = \rho_B\). We now want to construct from it a positive semi-definite and sub-normalized operator \(\hat{\rho}_{AB}\) such that \(1_{\text{max}}(A:B)\hat{\rho}\) is a lower bound to \(D_{\text{max}}(\hat{\rho}_{AB}\|\rho_A \otimes \rho_B)\) and \(P(\rho_{AB}, \hat{\rho}_{AB}) \leq c(\epsilon)\) with \(c(\epsilon)\) a function that vanishes as \(\epsilon \to 0\).

We can write \(\Delta_A\) as \(\Delta_A = \Delta_A^+ - \Delta_A^-\), where \(\Delta_A^+\) and \(\Delta_A^-\) are positive semi-definite operators with mutually orthogonal supports. Now we define
\[
\hat{\rho}_{AB} := n \cdot (\tilde{\rho}_{AB} + k \cdot \Delta_A^- \otimes \rho_B)
= nk \cdot (\tilde{\rho}_{AB} + \rho_A \otimes \Delta_B + \Delta_A^+ \otimes \rho_B),
\]
where \(n := (1 + k \cdot \text{tr}(\Delta_A))^{-1}\) is a normalization constant such that \(\text{tr}(\hat{\rho}_{AB}) = 1\). Clearly now, \(\hat{\rho}_{AB}\) is positive semi-definite and we want to estimate \(D_{\text{max}}(\hat{\rho}_{AB}\|\rho_A \otimes \rho_B)\). Notice that \(\hat{\rho}_A = n \cdot (\rho_A + k \Delta_A^-)\) and \(\hat{\rho}_B = \rho_B\). Hence,
\[
D_{\text{max}}(\hat{\rho}_{AB}\|\rho_A \otimes \rho_B)
\leq \log \left\| \hat{\rho}_{AB}^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \rho_{AB} \rho_A^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \right\|_\infty
= \log \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \tilde{\rho}_{AB} (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \right\|_\infty
+ \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} k \Delta_A^- (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^0 \right\|_\infty,
\]
and, with the triangle inequality,
\[
D_{\text{max}}(\hat{\rho}_{AB}\|\rho_A \otimes \rho_B)
\leq \log \left( \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \tilde{\rho}_{AB} (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \right\|_\infty
+ \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} k \Delta_A^- (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^0 \right\|_\infty \right).
\]
We can decompose the first term in the logarithm even further and with \(k \leq 1\), \(\Delta_B \leq \rho_B\) obtain
\[
D_{\text{max}}(\hat{\rho}_{AB}\|\rho_A \otimes \rho_B)
\leq \log \left( \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \tilde{\rho}_{AB} (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \right\|_\infty
+ 2 \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \tilde{\rho}_A (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \right\|_\infty
+ \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \tilde{\rho}_B (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \right\|_\infty
+ \left\| (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} k \Delta_A^- (\rho_A + k \Delta_A^-)^{-\frac{1}{2}} \otimes \rho_B^0 \right\|_\infty \right).
\]
Now we apply Lemma 18 to all terms inside the logarithm and replace \((\rho_A + k \Delta_A^-)\) with \(\rho_A\) in the first three terms and with \(k \Delta_A^-\) in the last one, obtaining
\[
D_{\text{max}}(\hat{\rho}_{AB}\|\rho_A \otimes \rho_B)
\leq \log \left( 2 \left\| \rho_A^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \tilde{\rho}_{AB} \rho_A^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \right\|_\infty + 3 \right)
\leq D_{\text{max}}(\rho_{AB}\|\rho_A \otimes \rho_B) + \log \left( 2 + \frac{3}{1 - \epsilon} \right).
\]
In the last line, we have used that
\[ \| \rho_A^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \hat{\rho}_{AB} \rho_A^{-\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}} \|_\infty \geq 1 - \varepsilon. \]

Thus,
\[
D_{\text{max}} (\hat{\rho}_{AB} \| \rho_A \otimes \hat{\rho}_B) \\
\leq \min_{\sigma_B \in S_{\infty} (\mathcal{H}_B)} D_{\text{max}} (\rho_{AB} \| \rho_A \otimes \sigma_B) \\
+ \log \left( \frac{2(1 - \varepsilon) + 3}{(1 - \varepsilon)(1 - \sqrt{1 - \varepsilon^2})} \right).
\]

Let us finally find an estimate for $P(\hat{\rho}_{AB}, \rho_{AB})$. Recall that
\[ \text{tr} \Delta_{ABC} = \text{tr} (\Pi_{AC}^+ \rho_{ABC}) \leq 1 - \sqrt{1 - \varepsilon^2} \]
according to our choice of $\Pi_{AC}$ in Claim 3 which implies
\[ k \geq \frac{1}{2 - \sqrt{1 - \varepsilon^2}}. \]

We further have that $\text{tr} \Delta_A \leq 2\varepsilon$ and therefore $n \geq \frac{1}{1 + 2\varepsilon}$. Thus, with Corollary 16
\[ P(\hat{\rho}_{AB}, nk \cdot \hat{\rho}_{AB}) \leq \sqrt{1 - n^2 k^2} \leq 2\varepsilon, \]
and consequently
\[ P(\rho_{AB}, nk \cdot \hat{\rho}_{AB}) \leq P(\rho_{AB}, \hat{\rho}_{AB}) + P(\hat{\rho}_{AB}, nk \cdot \hat{\rho}_{AB}) \leq \varepsilon + 2\sqrt{\varepsilon}. \]

As $nk \cdot \hat{\rho}_{AB} \leq \hat{\rho}_{AB}$ and therefore with Lemma 13
\[ \| \sqrt{n k \cdot \hat{\rho}_{AB} \sqrt{\rho_{AB}}} \|_1 \leq \| \sqrt{\rho_{AB} \sqrt{\rho_{AB}}} \|_1, \]
we conclude that also $P(\hat{\rho}_{AB}, \rho_{AB}) \leq \varepsilon + 2\sqrt{\varepsilon}$.

In summary, we have just proven the following claim.

**Claim 4:** For any $\rho_{AB} \in S_{\infty} (\mathcal{H}_{AB})$ and $\varepsilon > 0$, there exists a state $\hat{\rho}_{AB} \approx_{\varepsilon + 2\sqrt{\varepsilon}} \rho_{AB}$ that satisfies
\[
D_{\text{max}} (\hat{\rho}_{AB} \| \rho_A \otimes \hat{\rho}_B) \leq \min_{\sigma_B \in S_{\infty} (\mathcal{H}_B)} D_{\text{max}} (\rho_{AB} \| \rho_A \otimes \sigma_B) \\
+ \log \left( \frac{2(1 - \varepsilon) + 3}{(1 - \varepsilon)(1 - \sqrt{1 - \varepsilon^2})} \right). \quad (C.35)
\]

To conclude the proof of Theorem 3 let $\rho_{AB} \in S_{\infty} (\mathcal{H}_{AB})$ and let $\rho'_{AB} \in B^{\infty} (\rho_{AB})$ be a normalized operator such that $2I_{\text{max}}^{\varepsilon} (A : B)_\rho = 2I_{\text{max}}^{\varepsilon} (A : B)_{\rho'}. \text{ Applying Claim 4 to } \rho'_{AB} \text{ yields the result.}

**References**

[1] C. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948, available: http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html

[2] R. Renner and S. Wolf, “Smooth Renyi entropy and applications,” in *IEEE International Symposium on Information Theory*, ser. IEEE International Symposium on Information Theory, 2004, p. 233.

[3] R. Renner, “Security of quantum key distribution,” Ph.D. dissertation, ETH Zurich, 2005, available: http://arxiv.org/abs/quant-ph/0512258v2

[4] R. König, R. Renner, and C. Schaffner, “The operational meaning of min- and max-entropy,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 2474–2487, 2009, available: http://arxiv.org/abs/0807.1338v1

[5] M. Tomamichel, R. Colbeck, and R. Renner, “A fully quantum asymptotic equipartition property,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 5840–5847, 2009, available: http://arxiv.org/abs/0811.1221v3

[6] ———, “Duality between smooth min- and max-entropies,” *IEEE Trans. Inf. Theory*, vol. 56, p. 4674–4681, 2010, available: http://arxiv.org/abs/0907.5238v2

[7] N. Datta, “Min- and max- relative entropies and a new entanglement monotone,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 2816–2826, 2009, available: http://arxiv.org/abs/0803.2770v3
[8] M. Tomamichel, “A framework for non-asymptotic quantum information theory,” Ph.D. dissertation, ETH Zurich, 2012, available: http://arxiv.org/abs/1203.2142.

[9] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, “Leftover hashing against quantum side information,” IEEE Trans. Inf. Theory, vol. 57(8), pp. 2703–2707, 2011, available: http://arxiv.org/abs/1002.2436v1.

[10] N. J. Beaudry and R. Renner, “An intuitive proof of the data processing inequality,” Quantum Information and Computation, vol. 482, pp. 432–441, 2012, available: http://arxiv.org/abs/1107.0740.

[11] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner, “Chain rules for smooth min- and max-entropies,” IEEE Trans. Inf. Theory, vol. 59, pp. 2603–2612, 2013, available: http://arxiv.org/abs/1205.5231.

[12] N. Datta and R. Renner, “Smooth Renyi entropies and the quantum information spectrum,” IEEE Trans. Inf. Theory, vol. 55, pp. 2807–2815, 2009, available: http://arxiv.org/abs/0801.0282v1.

[13] M. Tomamichel and M. Hayashi, “A hierarchy of information quantities for finite block length analysis of quantum tasks,” published online, 2012, available: http://arxiv.org/abs/1208.1478.

[14] T. S. Han and S. Verdú, “Approximation theory of output statistics,” IEEE Trans. Inf. Theory, vol. 39, pp. 752–772, 1993.

[15] S. Verdú and T. S. Han, “A general formula for channel capacity,” IEEE Trans. Inf. Theory, vol. 40, pp. 1147–1157, 1994.

[16] M. Hayashi and H. Nagaoka, “General formulas for capacity of classical-quantum channels,” IEEE Trans. Inf. Theory, vol. 49, pp. 1753–1768, 2003, available: http://arxiv.org/abs/quant-ph/0206186.

[17] H. Nagaoka and M. Hayashi, “An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses,” IEEE Trans. Inf. Theory, vol. 53, pp. 534–549, 2007, available: http://arxiv.org/abs/quant-ph/0206185.

[18] N. Datta and M.-H. Hsieh, “The apex of the family tree of protocols: Optimal rates and resource inequalities,” New J. Phys., vol. 13, p. 093042, 2011, available: http://arxiv.org/abs/1103.1135.

[19] N. Datta, M.-H. Hsieh, and M. M. Wilde, “Quantum rate distortion, reverse shannon theorems, and source-channel separation,” IEEE Trans. Inf. Theory, vol. 59, pp. 615–630, 2013, available: http://arxiv.org/abs/1108.4940.

[20] F. G. S. L. Brandao and M. Horodecki, “Exponential decay of correlations implies area law,” published online, 2012, available: http://arxiv.org/abs/1206.2947.

[21] M. Berta, M. Christandl, and R. Renner, “The quantum reverse Shannon theorem based on one-shot information theory,” Commun. Math. Phys., vol. 306, p. 579, 2011, available: http://arxiv.org/abs/0912.3805.

[22] M. Berta, J. M. Renes, and M. M. Wilde, “Identifying the information gain of a quantum measurement,” published online, 2013, available: http://arxiv.org/abs/1301.1594.

[23] N. Datta, J. M. Renes, R. Renner, and M. M. Wilde, “One-shot lossy quantum data compression,” published online, 2013, available: http://arxiv.org/abs/1304.1336.

[24] P. Basler, “In search of operational quantities for characterizing large quantum systems,” Bachelor Thesis, ETH Zurich, 2012.

[25] F. Buscemi and N. Datta, “The quantum capacity of channels with arbitrarily correlated noise,” IEEE Trans. Inf. Theory, vol. 56, pp. 1447–1460, 2010, available: http://arxiv.org/abs/0902.0158v5.

[26] F. Furrer, J. Åberg, and R. Renner, “Min- and max-entropy in infinite dimensions,” Commun. Math. Phys., vol. 306, pp. 165–186, 2011, available: http://arxiv.org/abs/1004.1386.

[27] A. Uhlmann, “The transition probability for states of *-algebras,” Annalen der Physik, vol. 497, p. 524, 1985.

[28] M. Nielsen and I. Chuang, Quantum Computation and Quantum Information. Cambridge University Press, 2000.

[29] M. Berta, M. Christandl, R. Colbeck, J. Renes, and R. Renner, “The uncertainty principle in the presence of quantum memory,” Nature Phys., vol. 6, pp. 659–662, 2010, available: http://arxiv.org/abs/0909.0950.

[30] E. Onorati, “Exploring the quantum min-information,” Semester Thesis, ETH Zurich, 2012.