The operational meaning of min- and max-entropy

Robert König,1 Renato Renner,2 and Christian Schaffner3

1Institute for Quantum Information, California Institute of Technology, Pasadena, USA
2Institute for Theoretical Physics, ETH Zürich, Switzerland
3Centre for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands

We show that the conditional min-entropy \( H_{\min}(A|B) \) of a bipartite state \( \rho_{AB} \) is directly related to the maximum achievable overlap with a maximally entangled state if only local actions on the \( B \)-part of \( \rho_{AB} \) are allowed. In the special case where \( A \) is classical, this overlap corresponds to the probability of guessing \( A \) given \( B \). In a similar vein, we connect the conditional max-entropy \( H_{\max}(A|B) \) to the maximum fidelity of \( \rho_{AB} \) with a product state that is completely mixed on \( A \). In the case where \( A \) is classical, this corresponds to the security of \( A \) when used as a secret key in the presence of an adversary holding \( B \). Because min- and max-entropies are known to characterize information-processing tasks such as randomness extraction and state merging, our results establish a direct connection between these tasks and basic operational problems. For example, they imply that the (logarithm of the) probability of guessing \( A \) given \( B \) is a lower bound on the number of uniform secret bits that can be extracted from \( A \) relative to an adversary holding \( B \).

I. INTRODUCTION

A central goal of information theory is the (quantitative) analysis of processes involving the acquisition, transmission, and storage of information. For example, given a (noisy) communication channel, one may ask at which rate data can be transmitted reliably (this is the channel capacity). Or, given a source emitting signals, one may be interested in the amount of space needed to store the information in such a way that the signal can be recovered later (this is the compression rate). In the following, we call such quantities operational because they are defined by an actual information-processing task.

Traditionally, most operational quantities are defined asymptotically under the assumption that a certain process is repeated many times independently.1 Consider for example the problem of data compression. For a random variable \( X \) and for \( \epsilon \geq 0 \), let \( \ell_{\text{comp}}(X) \) be the minimum length (measured in terms of bits) of an encoding \( \text{enc}(X) \) such that \( X \) can be recovered from \( \text{enc}(X) \) except with an error probability of at most \( \epsilon \). The compression rate \( \ell_{\text{comp}}(X) \) of a source emitting a sequence of mutually independent pieces of data \( X_1, \ldots, X_n \), each distributed according to \( P_X \), is then defined by

\[
\ell_{\text{comp}}(P_X) := \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{\ell_{\text{comp}}(X_1 \cdots X_n)}{n}.
\]

It is maybe one of the most remarkable features of information theory that a huge variety of operational quantities can be expressed in terms of a few simple entropy measures. In fact, in the asymptotic case where a process is repeated many times independently, the (almost) only relevant entropy measure is the Shannon entropy (or its quantum-mechanical generalization, the von Neumann entropy). For example, the compression rate \( \ell_{\text{comp}}(X) \) of a source emitting data distributed according to \( P_X \) is equal to the Shannon entropy \( S(X) \) of a random variable \( X \) with distribution \( P_X \), i.e.,

\[
r_{\text{comp}}(P_X) = S(X).
\]

This equality is also known as the source-coding theorem 1. Another well-known example is the channel capacity. According to the noisy-channel coding theorem 1, the maximum rate at which information can be transmitted over a noisy communication channel is equal to a difference between two Shannon entropies (see 3 below).

The situation is different in the non-asymptotic case or when the independence assumption is dropped. Here, the Shannon / von Neumann entropies do no longer give a correct characterization of operational quantities.2 One therefore has to replace them by more general entropy measures. In the past few years, several such generalizations have been developed, notably the spectral entropy rates 2, as well as (smooth) min- and max-entropies 2. While both notions completely overcome the need of independence or Markovian assumptions, spectral entropy rates are (as suggested by their name) still restricted to asymptotic considerations. In contrast, smooth min- and max-entropies are fully general.3 In particular, no repetition of random processes is required. That is, one may consider situations where a source only emits one single piece of information or where a channel is only used once.

The aim of the present paper is to propose new operational interpretations of these non-asymptotic entropy measures. Our main findings are motivated and described in the following subsections, which are organized as follows. In Section 1 we review the notion of min-/max-entropies, our central object of interest. These entropy measures are the basis

\footnote{1 The independence assumption is sometimes replaced by the less restrictive requirement that the process is Markovian.}

\footnote{2 For example, the minimum compression length \( \ell_{\text{comp}}(X) \) defined above can deviate arbitrarily from \( S(X) \). This is readily verified by the following example: Let \( X \) be defined as the random variable which takes the value 0 with probability \( \frac{1}{2} \), and with probability \( \frac{1}{2} \) is equal to a uniformly distributed bitstring of length \( n \). Then \( S(X) \approx \frac{n}{2} \) while \( \ell_{\text{comp}}(X) \approx n \) for any sufficiently small \( \epsilon \).}

\footnote{3 The spectral entropy rates can be seen as asymptotic limits of smooth min-/max-entropies 2.}

1 Institute for Quantum Information, California Institute of Technology, Pasadena, USA
2 Institute for Theoretical Physics, ETH Zürich, Switzerland
3 Centre for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands

We show that the conditional min-entropy \( H_{\min}(A|B) \) of a bipartite state \( \rho_{AB} \) is directly related to the maximum achievable overlap with a maximally entangled state if only local actions on the \( B \)-part of \( \rho_{AB} \) are allowed. In the special case where \( A \) is classical, this overlap corresponds to the probability of guessing \( A \) given \( B \). In a similar vein, we connect the conditional max-entropy \( H_{\max}(A|B) \) to the maximum fidelity of \( \rho_{AB} \) with a product state that is completely mixed on \( A \). In the case where \( A \) is classical, this corresponds to the security of \( A \) when used as a secret key in the presence of an adversary holding \( B \). Because min- and max-entropies are known to characterize information-processing tasks such as randomness extraction and state merging, our results establish a direct connection between these tasks and basic operational problems. For example, they imply that the (logarithm of the) probability of guessing \( A \) given \( B \) is a lower bound on the number of uniform secret bits that can be extracted from \( A \) relative to an adversary holding \( B \).
for the definition of smooth min-/max-entropies, which can be seen as generalizations of Shannon / von Neumann entropy, as indicated above. Their properties are discussed later in Section I A. After this preparation, we will turn to connections between (smooth) min/max-entropies and operational quantities, starting with some important examples in Section I B. We then summarize the new operational interpretations derived in this work as well as their implications in Section I C.

A. (Smooth) min-/max-entropy: Basic definitions

Min-/max-entropy

We start with the definition of conditional min-entropy. This quantity (and the closely related conditional max-entropy) is the main object of study of this paper. In what follows, $id_A$ denotes the identity on system $A$.

Definition 1. Let $\rho = \rho_{AB}$ be a bipartite density operator. The min-entropy of $A$ conditioned on $B$ is defined by

$$H_{\min}(A|B)_\rho := -\inf_{\sigma_B} D_{\infty}(\rho_{AB} || id_A \otimes \sigma_B) \tag{3}$$

where the infimum ranges over all normalized density operators $\sigma_B$ on subsystem $B$ and where $^4$

$$D_{\infty}(\tau || \tau') := \inf\{\lambda \in \mathbb{R}: \tau \leq 2^\lambda \tau'\}. \tag{4}$$

It is interesting to note that the Shannon / von Neumann entropy could be defined in a similar way. Namely, if we replace $H_{\min}$ by the von Neumann entropy $S$ and $D_{\infty}$ by the relative entropy $^5$ in (3), we find

$$S(A|B)_\rho = -\inf_{\sigma_B} D(\rho_{AB} || id_A \otimes \sigma_B).$$

This equality is readily verified using the fact that $D(\tau || \tau')$ is nonnegative for any normalized $\tau$, $\tau'$ and equals zero if $\tau = \tau'$.

For a tripartite pure state $\rho = \rho_{ABC}$, the von Neumann entropy satisfies the equality $^6$

$$S(A|B)_\rho = -S(A|C)_\rho. \tag{5}$$

This same is no longer true for the min-entropy. However, it turns out that the entropy obtained by replacing the system $B$ by the “purifying system” $C$ often appears in expressions characterizing operational quantities. This motivates the following definition.

Definition 2. Let $\rho = \rho_{AB}$ be a bipartite density operator. The max-entropy of $A$ conditioned on $B$ is defined by

$$H_{\max}(A|B)_\rho := -H_{\min}(A|C)_\rho \tag{6}$$

where the min-entropy on the rhs. is evaluated for a purification $\rho_{ABC}$ of $\rho_{AB}$.

This is well-defined because all purifications of $\rho_{AB}$ are related by unitaries on $C$, and the quantity $H_{\min}(A|C)_\rho$ is invariant under such unitaries.

We point out that $H_{\min}$ and $H_{\max}$ could have been defined alternatively by starting from an expression for $H_{\max}$ and subsequent definition of $H_{\min}$ by purification (i.e., (6)). In this sense, both quantities are equally fundamental.

If the state $\rho$ is clear from the context, we will omit the subscript in $H_{\min}(A|B)_\rho$ and $H_{\max}(A|B)_\rho$. Also, in the special case where the system $B$ is trivial (i.e., one-dimensional), we omit the conditioning and simply write $H_{\min}(A)_\rho$ and $H_{\max}(A)$. Note that the above definitions also apply to classical probability distributions $P_X$ which can always be written as quantum states $\rho_X = \sum_x P_X(x)\ket{\!x\!}\bra{\!x\!}$ for some orthonormal basis $\{\ket{\!x\!}\}_x$.

To get some more intuition for these definitions, it may help to compute their value for certain special states. One extreme case are product states $\rho_{AB} = \rho_A \otimes \rho_B$, for which one readily verifies that the min-entropy only depends on the maximum eigenvalue $\|\rho_A\|_\infty$ of $\rho_A$, i.e., $H_{\min}(A|B)_\rho = -\log_2 \|\rho_A\|_\infty$. Note that this corresponds to the Rényi entropy of order infinity of the density operator $\rho_A$. Similarly, we get $H_{\max}(A|B)_\rho = 2\log_2 \text{tr}\sqrt{\rho_A}$, which is the Rényi entropy of order $\frac{1}{2}$ of $\rho_A$ (see eq. (24) below). Another extreme case is where $\rho_{AB}$ is a pure state. Here one finds $H_{\min}(A|B)_\rho = -\log_2 (\text{tr} \sqrt{\rho_A})^2$ and $H_{\max}(A|B)_\rho = \log_2 \|\rho_A\|_\infty$.

Smooth min-/max-entropy

The smooth min/max-entropy of a state $\rho$ is defined by the corresponding (non-smooth) min/max-entropy for an “optimal” state $\rho'$ in an $\varepsilon$-neighborhood of $\rho$, where $\varepsilon$ is called smoothness parameter.

Definition 3. Let $\rho = \rho_{AB}$ be a bipartite density operator and let $\varepsilon \geq 0$. The $\varepsilon$-smooth min- and max-entropy of $A$ conditioned on $B$ are given by

$$H_{\min}^{\varepsilon}(A|B)_\rho := \sup_{\rho' \neq \rho} H_{\min}(A|B)_{\rho'},$$

$$H_{\max}^{\varepsilon}(A|B)_\rho := \inf_{\rho' \neq \rho} H_{\max}(A|B)_{\rho'},$$

where the supremum ranges over all density operators $\rho'$ which are $\varepsilon$-close to $\rho$. $^7$

$^4$ For commuting density operators $\tau$ and $\tau'$, the quantity $D_{\infty}(\tau || \tau')$ corresponds to the (classical) relative Rényi entropy of order $\infty$. In general, the relative Rényi entropy of order $\alpha$ of two probability distributions $P$ and $Q$ is defined as $D_{\alpha}(P||Q) := \frac{1}{\alpha - 1} \log_2 \sum_x P(x)^{\alpha}Q(x)^{1-\alpha}$, and $D_{\infty}$ is obtained in the limit $\alpha \to \infty$.

$^5$ Note that the relative entropy (aka Kullback-Leibler divergence) $D(\tau || \tau') := \text{tr}(\tau \log_2 \tau - \log_2 \tau')$ is also defined for unnormalized operators $\tau'$.

$^6$ Note that, by definition, $S(A|B) = S(AB) - S(B)$ and $S(A|C) = S(AC) - S(C)$. The equality then follows from the fact that, by the Schmidt decomposition, $S(AB) = S(C)$ and $S(B) = S(AC)$.

$^7$ In the classical case, smooth entropies are usually defined with respect
Basic properties. It follows directly from the definitions that the same kind of duality between min- and max-entropy holds between the corresponding smooth versions, namely

\[ H_{\text{max}}^\varepsilon(A|B)_\rho = -H_{\text{min}}^\varepsilon(A|C)_\rho \]  

for a purification \( \rho_{ABC} \) of \( \rho_{AB} \).

As already indicated, smooth min-/max-entropies can be seen as generalizations of the Shannon / von Neumann entropy \( S \). More precisely, the latter can be written in terms of the former \( (5) \), i.e.,

\[ S(A|B)_\rho = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{\text{min}}^\varepsilon(A^n|B^n)_\rho \]  

and likewise for \( H_{\text{max}}^\varepsilon \). In fact, inequality \( (10) \) can be seen as a generalization of the strong subadditivity of the von Neumann entropy, \( S(A|B) \geq S(A|BC) \), which can be recovered by virtue of identity \( (8) \), i.e., for any \( \rho_{ABC} \),

\[ S(A|B)_\rho \stackrel{\text{def}}{=} \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{\text{min}}^\varepsilon(A^n|B^n)_\rho \geq \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{\text{max}}^\varepsilon(A^n|B^n^2C^n)_\rho \geq S(A|BC)_\rho . \]

Interestingly, despite its generality, inequality \( (10) \) is easy to prove, as we shall see at the end of Section \[ \text{B} \].

B. Operational quantities in terms of smooth min-/max-entropy

The main reason for considering (smooth) min-/max-entropies is that they are well suited for the characterization of operational quantities in the most general case. Recall that expressions for operational quantities involving the Shannon / von Neumann entropy, e.g., \( (2) \), are typically only valid asymptotically, under the assumption that certain resources can be used many times independently. Interestingly, the structure of such expressions essentially remains the same if one drops these assumptions, except that smooth entropies take the place of Shannon / von Neumann entropy. The purpose of this section is to illustrate this phenomenon with a few examples.

Data compression. We start with the example of data compression, which has already been introduced above. For a random variable \( X \) and \( \varepsilon \geq 0 \), let again \( \ell_{\text{compr}}^\varepsilon(X) \) be the minimum length of an encoding from which the value of \( X \) can be recovered correctly with probability at least \( 1 - \varepsilon \). It can then be shown that \( \ell_{\text{compr}}^\varepsilon(X) \) is essentially equal to the smooth max-entropy of \( X \) \[ (3) \]. More precisely, we have

\[ \ell_{\text{compr}}^\varepsilon(X) = H_{\text{max}}^\varepsilon(X) + O(\log 1/\varepsilon) \]  

for some \( \varepsilon \in [\frac{1}{2}\varepsilon, 2\varepsilon] \). The \( O \)-notation indicates that equality holds up to an additive term of the order \( \log(1/\varepsilon) \).\(^8\) In typical applications, this logarithmic term is much smaller than the other quantities occurring in the expression. In particular, the term is independent of the size of the resource (in our case the random variable \( X \)) and thus becomes irrelevant in the asymptotic limit of large resources.

We stress that \( (11) \) is valid for a single realization of the random variable \( X \) and thus strictly generalizes Shannon’s source coding theorem described at the beginning of this section. Identity \( (2) \) can be recovered as an asymptotic limit of \( (11) \), for \( X \) consisting of many independent and identically distributed pieces \( X_1, \ldots, X_n \), i.e.,

\[ r_{\text{compr}}(P_X) \stackrel{\text{def}}{=} \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\ell_{\text{compr}}(X_1 \cdots X_n)}{n} = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{\text{max}}^\varepsilon(X_1 \cdots X_n) = S(X) . \]

Channel coding. As a second example, we consider the noisy-channel coding problem. For any \( \varepsilon \geq 0 \), let \( \ell_{\text{trans}}^\varepsilon(X \to Y) \) be the maximum number of bits that can be transmitted in one use of a classical noisy channel \( X \to Y \) (specified by a conditional probability distribution \( P_{Y|X} \)) with maximum error probability \( \varepsilon \). As shown in \[ (6) \], this quantity is given by

\[ \ell_{\text{trans}}^\varepsilon(X \to Y) = \max_{P_X} (H_{\text{min}}^\varepsilon(X) - H_{\text{max}}^\varepsilon(X|Y)) + O(\log 1/\varepsilon) \]  

for some \( \varepsilon \in [\frac{1}{2}\varepsilon, 2\varepsilon] \).

---

\(^8\) Note that the smooth entropies are monotonic functions of \( \varepsilon \). Equality \( (11) \) is thus just a way to state that the operational quantity \( \ell_{\text{compr}}^\varepsilon(X) \) lies in the interval \( [H_{\text{max}}^\varepsilon(X), H_{\text{max}}^\varepsilon(X)] \), up to some additive constant of the order \( \log 1/\varepsilon \).
where
\[ \varepsilon \]

that information bitstring of applications, particularly in the context of cryptography, where it is also called length of a bitstring can be extracted from and let information. The goal of transmission is the maximum rate at which information can be transmitted by \( n \) uses of the channel \( X \rightarrow Y \), in the limit of large \( n \). Using the non-asymptotic statement \((8)\) together with \((9)\), we find
\[
\rho_{\text{trans}}(P_Y | X) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\ell_{\text{trans}}(X^n \rightarrow Y^n)}{n},
\]
that is, the maximum rate at which information can be transmitted by \( n \) uses of the channel \( X \rightarrow Y \), in the limit of large \( n \). Using the non-asymptotic statement \((8)\) together with \((9)\), we find
\[
\rho_{\text{trans}}(P_Y | X) = \max_{P_X} \left( S(X) - S(X|Y) \right) = \max_{P_X} I(X : Y).
\]
(13)
This is Shannon’s well-known noisy-channel coding theorem.

**Privacy amplification.** Let \( X \) be a classical random variable and let \( B \) be (possibly quantum-mechanical) side information. The goal of randomness extraction is to compute a bitstring \( f(X) \) which is uniform and independent of the side information \( B \). Randomness extraction is crucial for a number of applications, particularly in the context of cryptography, where it is also called privacy amplification [2]. For example, in a key-agreement scheme, one may want to turn a (only partially secure) raw key \( X \) into a fully secure key \( f(X) \). Security of \( f(X) \) is then akin to uniformity relative to side information \( B \) held by a potential adversary.

The maximum number of uniform and independent bits that can be extracted from \( X \) is directly given by the smooth min-entropy of \( X \). More precisely, let \( \ell_{\text{extr}}(X|B) \) be the maximum length of a bitstring \( f(X) \) that can be computed from \( X \) such that \( f(X) \) is \( \varepsilon \)-close to a string \( Z \) which is perfectly uniform and independent of the side information \( B \). \(^9\) One can show that \( \ell_{\text{extr}}(X|B) = H_{\text{min}}^\varepsilon(X|B) + O(\log 1/\varepsilon) \)
\[
\ell_{\text{extr}}(X|B) = H_{\text{min}}^\varepsilon(X|B) + O(\log 1/\varepsilon)
\]
where \( \varepsilon \in \left[ \frac{1}{2} \varepsilon, 2\varepsilon \right] \). In the special case where \( B \) is independent of \( X \), this corresponds to the leftover hash lemma \([7, 9]\). For later reference, we also note that
\[
\ell_{\text{extr}}(X|B) \geq H_{\text{min}}(X|B) + O(\log 1/\varepsilon),
\]
(14)
which holds because \( H_{\text{min}}^\varepsilon(X|B) \) is monotonically increasing in \( \varepsilon \) and equals \( H_{\text{min}}(X|B) \) for \( \varepsilon = 0 \).

**Decoupling.** The previous result can be extended to a fully quantum-mechanical setting as follows. Let \( A \) and \( B \) be two quantum systems. The goal is to find a maximum subsystem \( A' \) of \( A \) such that the state on \( A' \) is completely mixed and decoupled from \( B \) conditioned on a suitable measurement on the remaining part of \( A \). Let \( \ell_{\text{deci}}(A|B) \) be the maximum size of \( A' \) (measured in qubits) such that this is possible up to a distance \( \varepsilon \). \(^10\) One then finds \( [10, 11, 12] \)
\[
\ell_{\text{deci}}(A|B) = H_{\text{min}}^\varepsilon(A|B) + O(\log 1/\varepsilon)
\]
(15)
with \( \varepsilon \in \left[ \frac{1}{2} \varepsilon, 2\varepsilon \right] \).

**State merging.** In the same manner as privacy amplification generalizes to decoupling in the fully quantum case, data compression (and its relatives such as coding with side information) extends to a fully quantum setting; this is referred to as state merging. The setting is described by a tripartite pure state \( |\Psi_{ABC}\rangle \). The aim is to redistribute the A-part to the system \( B \) by local operations and classical communications (LOCC) between \( A \) and \( B \). Depending on the (reduced) state \( \rho = \rho_{AB} \), this either consumes or generates bipartite entanglement. Let \( (\cdot)_{\varepsilon} \) be minimal (maximal) number of ebits of entanglement required (generated) by this process (the distinction between consumed/generated entanglement is reflected by the sign of the quantity \( \ell_{\text{merg}}(A|B) \), such that the outcome is \( \varepsilon \)-close to the desired output. \(^12\) One then finds
\[
\ell_{\text{merg}}(A|B)_\rho = H_{\text{max}}^\varepsilon(A|B)_\rho + O(\log 1/\varepsilon)
\]
(16)
where again \( \varepsilon \in \left[ \frac{1}{2} \varepsilon, 2\varepsilon \right] \) (see \([12]\) for details). In fact, the \( \geq \)-part of this statement is a direct consequence of decoupling result above \([11]\), the arguments in \([10]\) (cf. also Section \([11B]\) and the definition of \( H_{\text{max}}(A|B)_\rho \).

**C. Contribution: Min-/max-entropies as operational quantities**

In this paper, we show that min-/max-entropies have direct\(^13\) operational interpretations. We begin by presenting the corresponding results for the special case where we condition classical information \( X \) on a (possibly) quantum system \( B \). The fully general case is discussed later in Section \([1C2]\).

1. **Uncertainty about classical information**

Consider an agent with access to a (classical or quantum) system \( B \) whose state \( \rho_B^\varepsilon \) depends on a classical random

\(^9\) See paragraph on max-entropy of classical information in Section \([1C]\) below for more details.

\(^10\) This distance is quantitatively expressed by the decoupling accuracy, see below.

\(^11\) This is based on a tightened version \([11]\) of a bound obtained in \([10]\), which shows that projecting onto a random subspace of dimension \( \dim A' \) achieves decoupling. More precisely, it can be shown \([11]\) that the decoupling accuracy of the residual state is, on average over the measurement outcome, exponentially small in the difference \( H_{\text{min}}(A|B) - \log \dim A' \).

\(^12\) Closeness is measured in terms of the distance of the output state \( \rho_{AB'B_B'C_{B_C}} \) of the protocol to the state \( |\Phi_{AB'}^\varepsilon \rangle \otimes |\Psi_{B_B'C_{B_C}}\rangle \), where \( |\Phi_{AB'}\rangle \) is an ebit between \( A \) and \( B \), \( \varepsilon \) is the number of ebits generated, and \( |\Psi_{B_B'C_{B_C}}\rangle \) is identical to \( |\Psi_{ABC}\rangle \) when identifying the subsystems \( B_B \) and \( A \) as well as \( B_C \) and \( B \).

\(^13\) The term direct refers to the fact that no smoothing is required, in contrast to the examples of Section \([1B]\)
variable $X$. This situation can be described by a classical-quantum state

$$\rho = \rho_{XB} := \sum_x P_X(x)|x\rangle\langle x| \otimes \rho_B^x, \tag{17}$$

with $\{|x\rangle\}_x$ a family of mutually orthogonal vectors representing the (classical) values of $X$.

**Min-entropy of classical information is guessing probability.** Let $p_{\text{guess}}(X|B)$ be the probability that the agent correctly guesses $X$ when using an optimal strategy; that is, $p_{\text{guess}}(X|B) = \sum_x P_X(x)\text{tr}(E_x \rho_B^x)$, where the optimal measurement strategy is described by the POVM $\{E_x\}_x$ on $B$ that maximizes this expression. We then show that (cf. Theorem 1)

$$p_{\text{guess}}(X|B) = 2^{-H_{\min}(X|B)\rho}. \tag{18}$$

where the entropy is evaluated for the state $\rho_{XB}$ given by (17).

If no side information $B$ is available or, more generally, if the state of $B$ is independent of $X$, we have $2^{-H_{\min}(X|B)} = |\rho_X|_{\infty} = \max_x P_X(x)$ as noted in Section I A. Identity (18) then reduces to the trivial fact that the maximum probability of correctly guessing $X$ without prior information is equal to $\max_x P_X(x)$.

**Max-entropy of classical information is security of key.** The secrecy of $X$ when used as a key in the presence of an adversary with access to system $B$ is conventionally measured in terms of the distance of the state $\rho_{XB}$ (cf. (17)) to a product state of the form $\tau_X \otimes \rho_B$, where $\tau_X$ is the completely mixed state (corresponding to the uniform distribution on $X$) and where $\rho_B$ is the reduced state on subsystem $B$. This models an ideal situation where the key is perfectly uniform and independent of the adversary’s system. If the trace distance is used, then this distance is directly related to the *distinguishing advantage* between the real and the ideal system.

One may relax the above and only require that the desired state is of the form $\tau_X \otimes \sigma_B$, for some arbitrary density operator $\sigma_B$. When using the trace distance, this relaxed definition is equivalent to the above up to a factor of 2. Also, since the trace distance and the fidelity are essentially equivalent, we can use the fidelity. We then get the following measure for the secrecy of $X$ relative to $B$,

$$p_{\text{sec}}(X|B)_\rho := |X| \max_{\sigma_B} F(\rho_{XB}, \tau_X \otimes \sigma_B)^2 = \max_{\sigma_B} \left( \sum_x \sqrt{P_X(x) F(\rho_B^x, \sigma_B)^2} \right)^2. \tag{19}$$

where $|X|$ is the alphabet size of $X$ (we include this factor here for convenience). We show that (cf. Theorem 3)

$$p_{\text{sec}}(X|B)_\rho = 2^{H_{\max}(X|B)_\rho}. \tag{19}$$

If no side information $B$ is available or, more generally, if $B$ is independent of $X$, we obtain $2^{H_{\max}(X)_\rho} = \left( \sum_x \sqrt{P_X(x)} \right)^2$ (cf. Section I A). Identity (19) then simply expresses the fact that the secrecy of $X$ in this case is quantified by the distance of $P_X$ to the uniform distribution (where distance is measured in terms of the fidelity).

2. **Uncertainty about quantum information**

We now discuss the fully general case, where we have an arbitrary bipartite state $\rho = \rho_{AB}$. The min-/max-entropies carry the following operational interpretations.

**Min-entropy is maximum achievable singlet fraction.** Define the maximally entangled state

$$|\Phi_{AB}\rangle := \frac{1}{\sqrt{d}} \sum_x |x_A\rangle|x_B\rangle$$

where $\{|x_A\rangle\}^d_{x=1}$ is an orthonormal basis of subsystem $A$ (of dimension $d$) and $\{|x_B\rangle\}^d_{x=1}$ is a family of mutually orthogonal vectors on subsystem $B$ (we assume that $\text{dim } A \leq \text{dim } B$). We define the “quantum correlation” $q_{\text{cor}}(A|B)_\rho$ as the maximum overlap with the singlet state $|\Phi_{AB}\rangle$ that can be achieved by local quantum operations $\mathcal{E}$ (trace-preserving completely positive maps) on subsystem $B$, that is

$$q_{\text{cor}}(A|B)_\rho := d \max_{\mathcal{E}} F((\text{id}_A \otimes \mathcal{E})(\rho_{AB}), |\Phi_{AB}\rangle \langle \Phi_{AB}|) \tag{20}.$$
Implications. A main implication of our results is that they establish a connection between seemingly different operational quantities. For example, because the number $f_{\text{extr}}(X|B)$ of uniform bits that can be extracted from $X$ with respect to side information $B$ is lower bounded by $H_{\min}(X|B)$ (see (14)), we find that

$$f_{\text{extr}}(X|B) \geq - \log_2 p_{\text{guess}}(X|B) + O(\log 1/\varepsilon) ,$$

In other words, the negative logarithm of the guessing probability of $X$ tells us how many uniform bits we can extract from $X$ (relative to some system $B$). This connection between randomness extraction and guessing entropy may be useful for applications, e.g., in cryptography. Here, the derivation of lower bounds on the amount of extractable randomness is usually a central part of the security analysis (see [8, 14]).

Our results can also be used to prove additivity properties of the min-max-entropies. One of them is additivity of the min-max-entropies for independent systems. Let $\rho_{A'\oplus'B'} = \rho_{AB} \otimes \rho_{A'B'}$. Then, by the definition of $q_{\text{corr}}$,

$$q_{\text{corr}}(AA'|BB') \geq q_{\text{corr}}(A|B) \cdot q_{\text{corr}}(A'|B') .$$

By virtue of (21), this is equivalent to

$$H_{\min}(AA'|BB') \leq H_{\min}(A|B) + H_{\min}(A'|B') .$$

Note that the opposite inequality follows immediately from the definition of $H_{\min}$ and the additivity of $D_\infty$. We thus have

$$H_{\min}(AA'|BB') = H_{\min}(A|B) + H_{\min}(A'|B')$$

and, equivalently (by the definition (6))

$$H_{\max}(AA'|BB') = H_{\max}(A|B) + H_{\max}(A'|B')$$

A second example is the strong subadditivity of conditional min-entropy (10). Here, it suffices to notice that every trace-preserving completely positive map $E$ acting on $B$ can also be understood as acting on registers $B$ and $C$, hence,

$$q_{\text{corr}}(A|B)_{\rho} \leq q_{\text{corr}}(ABC)_{\rho}$$

for every quantum state $\rho_{ABC}$. By (21), this is equivalent to

$$H_{\min}(A|B)_{\rho} \geq H_{\min}(A|BC)_{\rho} .$$

The extension to smooth min-entropy (10) is straightforward (see Lemma 3.2.7 of [5]).

Our results also simplify the calculation of the min-max-entropies. As an example, let us calculate the entropy $H_{\max}(A|B)_{\rho}$ for a state of the form $\rho_{AB} = \rho_{A} \otimes \rho_{B}$. By (23), it suffices to determine the quantity $q_{\text{deep}}(A|B)_{\rho}$, which is given by

$$q_{\text{deep}}(A|B)_{\rho} = \max_{\sigma_{B}} d_{A} F(\rho_{A} \otimes \rho_{B}, \tau_{A} \otimes \sigma_{B})$$

where $\tau_{A}$ is the completely mixed state on the $d_{A}$-dimensional Hilbert space $A$. Using the multiplicativity of the fidelity, we find

$$q_{\text{deep}}(A|B)_{\rho} = d_{A} F(\rho_{A}, \tau_{A}) \max_{\sigma_{B}} F(\rho_{B}, \sigma_{B})$$

$$= d_{A} F(\rho_{A}, \tau_{A}) \sqrt{\text{tr} \rho_{B}} \|_{1}^{2} .$$

We thus obtain

$$H_{\max}(A|B)_{\rho} = 2 \log \text{tr} \rho_{A} ,$$

for any $\rho = \rho_{AB}$ of the form $\rho_{A} \otimes \rho_{B}$. This corresponds to the Rényi entropy of order $\frac{1}{2}$, which is hence the natural counterpart to the min-entropy (Rényi entropy of order $\infty$). As noted in [3], the Rényi entropy of order $\alpha$, for any $\alpha < 1$, is up to small additive terms of the order $\frac{1}{\alpha}$ determined by a smoothed version of $H_{0}(\rho_{A}) = - \log \text{rank}(\rho_{A})$. The max-entropy $H_{\max}$ of a density operator can thus be interpreted as a measure for its rank.

Outline of the remainder of this paper

In Section II, we discuss some mathematical preliminaries, in particular semidefinite programming, which plays a crucial role in our arguments. Our main results are then stated and proved in Section III.

II. SOME TECHNICAL PRELIMINARIES

A. Semidefinite programming

Our central tool will be the duality between certain pairs of semidefinite programs. It will be convenient to use a fairly general formulation of this duality; a derivation of the results summarized in this section can be found, e.g., in [15, Section 6]. The presentation here follows this reference, but specializes certain statements to the situation of interest for simplicity. We start by introducing a few definitions.

A subset $K \subseteq \mathcal{V}$ of a vector space $\mathcal{V}$ is called a convex cone if $0 \in K$ and $\mu v + \nu w \in K$ for all nonnegative $\mu, \nu \geq 0$ and $v, w \in K$. A convex cone $K$ gives rise to a partial order relation $\leq_{K}$ on $\mathcal{V}$, defined by $v \leq_{K} w$ if and only if $w - v \in K$. If $\mathcal{V}$ is an Euclidean space with inner product $\langle \cdot, \cdot \rangle$, then the dual cone $K^{\ast} \subseteq \mathcal{V}$ of $K$ is defined by $K^{\ast} = \{ v \in \mathcal{V} \mid \langle v, w \rangle \geq 0 \text{ for all } w \in K \}$. The interior $\text{int} K \subseteq K$ is the subset of points $w \in K$ for which there exists an open ball centered around $w$ and contained in $K$.

Let $\mathcal{V}_{1}$ and $\mathcal{V}_{2}$ be Euclidean spaces with inner products $\langle \cdot, \cdot \rangle_{1}$ and $\langle \cdot, \cdot \rangle_{2}$, respectively. A linear map $E^{\ast} : \mathcal{V}_{2} \rightarrow \mathcal{V}_{1}$ is called dual of or adjoint to a linear map $E : \mathcal{V}_{1} \rightarrow \mathcal{V}_{2}$ if

$$\langle Ev_{1}, v_{2} \rangle_{2} = \langle v_{1}, E^{\ast} v_{2} \rangle_{1} \quad \text{for all } v_{1} \in \mathcal{V}_{1}, v_{2} \in \mathcal{V}_{2} .$$

For a given map $E$, the dual map $E^{\ast}$ is necessarily unique if it exists. The two linear programming problems we are interested in are defined in terms of a pair of such maps. They
are referred to as the primal and dual problem, and are specified by parameters $c \in \mathcal{V}_1$ and $b \in \mathcal{V}_2$. The programs are expressed by the following optimizations:

$$
\gamma_{\text{primal}} = \inf_{v_1 \geq K_1, b} (v_1, c),
$$

$$
\gamma_{\text{dual}} = \sup_{v_2 \geq K_2, b} \langle b, v_2 \rangle.
$$

(25)

We will usually assume that the sets we optimize over are non-empty. (In the language of linear programming, there exists a feasible plan and a dual feasible plan.) The weak duality theorem states that $\gamma_{\text{primal}} \geq \gamma_{\text{dual}}$. We are particularly interested in conditions for equality. (This is referred to as a zero duality gap.) A simple criterion is Slater’s interiority condition, which states the following

**Lemma 1.** Suppose that there is an element $v \in \text{int} K_1$ such that $Ev - b \in \text{int} K_2$. Suppose further that the infimum in (25) is attained. Then $\gamma_{\text{primal}} = \gamma_{\text{dual}}$.

**B. Quantum operations**

Let $\mathcal{H}_A$ be a Hilbert space and let $\mathcal{L}(\mathcal{H}_A)$ be the set of linear maps $E : \mathcal{H}_A \rightarrow \mathcal{H}_A$. An element $E \in \mathcal{L}(\mathcal{H}_A)$ is called nonnegative (written $E \succeq 0$) if $\langle \psi | E | \psi \rangle \geq 0$ for all $| \psi \rangle \in \mathcal{H}_A$. A positive element $E$ (written $E > 0$) is defined in the same way with a strict inequality.

An operation is a linear map $E : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$. It is called trace-preserving if $\text{tr}(E(E)) = \text{tr}(E)$ for all $E \in \mathcal{L}(\mathcal{H}_A)$. It is unital if it maps the identity on $\mathcal{H}_A$ to the identity on $\mathcal{H}_B$, i.e., if $E(\text{id}_A) = \text{id}_B$. The map is called positive if $E(E) \succeq 0$ for all $E \succeq 0$. It is completely positive (CP) if $E \otimes \rho : \mathcal{L}(\mathcal{H}_R \otimes \mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_R \otimes \mathcal{H}_B)$ is positive for any auxiliary space $\mathcal{H}_R$, where $\text{id}_R$ is the identity operation. A quantum operation is a completely positive trace-preserving map (CPTP). The adjoint map of an operation $E : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ is the unique map $E^\dagger : \mathcal{L}(\mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_A)$ satisfying

$$
\text{tr}(F_B E(A)) = \text{tr}(E^\dagger(F_B) E_A)
$$

for all $E_A \in \mathcal{L}(\mathcal{H}_A)$ and $F_B \in \mathcal{L}(\mathcal{H}_B)$. Note that $(E^\dagger)^\dagger = E$, $\text{id}_A^\dagger = \text{id}_A$, and $(E \otimes F)^\dagger = E^\dagger \otimes F^\dagger$ for two maps $E$ and $F$. Two easily verified properties which follow directly from this definition are

$E$ is unital if and only if $E^\dagger$ is trace-preserving,

and

$E$ is positive if and only if $E^\dagger$ is positive.

In particular, the last statement implies that $E$ is completely positive (CP) if and only if $E^\dagger$ is CP.

(27)

**Lemma 2.** The adjoint map

$$
\dagger : \text{CPTPM}(\mathcal{H}_A, \mathcal{H}_B) \rightarrow \text{CPUM}(\mathcal{H}_B, \mathcal{H}_A)
$$

is a bijection with inverse

$$
\dagger : \text{CPUM}(\mathcal{H}_B, \mathcal{H}_A) \rightarrow \text{CPTPM}(\mathcal{H}_A, \mathcal{H}_B).
$$

Let $d_A$ be the dimension of $\mathcal{H}_A$ and let $\{|x\rangle_A\}_{x \in [d_A]}$ be an orthonormal basis of $\mathcal{H}_A$. (We will restrict our attention to finite-dimensional Hilbert spaces.) Let $\mathcal{H}_A' \cong \mathcal{H}_A$ be a Hilbert space with orthonormal basis $\{|x\rangle_A'\}_{x \in [d_A]}$. The maximally entangled state on $\mathcal{H}_A \otimes \mathcal{H}_A'$ is defined as

$$
\Phi_{AA'} = \dfrac{1}{\sqrt{d_A}} \sum_{x \in [d_A]} |x\rangle_A \otimes |x\rangle_A'.
$$

(28)

The **Choi-Jamiołkowski map** $J$ takes operations $E : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ to operators $J(E) \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. It is defined as

$$
J(E) = d_A (\text{id}_A \otimes E (|\Phi_{AA'}\rangle \langle \Phi_{AA'}|)).
$$

It has the following well-known properties. The equivalence of statements (i) and (ii) in the following lemma is an immediate consequence of Lemma 2.

**Lemma 3** (Choi-Jamiołkowski isomorphism, [16]). Let $\mathcal{H}_A \cong \mathcal{H}_A'$ and $\mathcal{H}_B$ be arbitrary Hilbert spaces. The map $J$ bijectively maps

(i) the set $\text{CPTPM}(\mathcal{H}_A, \mathcal{H}_B)$ to the set of operators $F_{AB} \geq 0$ with $\text{tr}_B F_{AB} = \text{id}_A$.

(ii) the set $\text{CPUM}(\mathcal{H}_A', \mathcal{H}_B)$ to the set of operators $E_{AB} \geq 0$ with $\text{tr}_A E_{AB} = \text{id}_B$.

Another concept we will need is the notion of classicality, which allows us to treat ensembles as quantum states. We will say that a Hermitian operator $E_{AB}$ on a bipartite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ is classical relative to an orthonormal basis $\{|x\rangle_A\}_{x \in [d_A]}$ of $\mathcal{H}_A$ if it is a linear combination of operators of the form $|x\rangle \langle x| \otimes E_B$, where $x \in [d_A]$ and $E_B \in \text{Herm}(\mathcal{H}_B)$.

**III. MAIN RESULTS AND THEIR DERIVATION**

We are now ready to prove our main statements. We first focus on the min-entropy in Section III A. The interpretation of max-entropy will then follow, as discussed in Section III B.

**A. Proof of the operational characterization of $H_{\text{min}}$**

With Lemma 1 from Section III A it is straightforward to prove the following statement. Note that we restrict our attention to finite-dimensional Hilbert spaces. Since the optimizations are now taken over compact sets, we can replace inf and sup by min and max, respectively.
Lemma 4. Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be finite-dimensional Hilbert spaces, and let $\rho_{AB}$ and $\sigma_B$ be nonnegative operators on $\mathcal{H}_A \otimes \mathcal{H}_B$, respectively. Then
\[
\min_{\sigma_B \geq 0} \text{tr}(\sigma_B) = \max_{E_{AB} \geq 0} \text{tr}(\rho_{AB} E_{AB}) \tag{29}
\]
In addition, if $\rho_{AB}$ is classical on $\mathcal{H}_A$ relative to an orthonormal basis $\{|x\rangle\}_{x}$, then the maximization rhs. of (29) can be further restricted to those operators $E_{AB}$ which are classical on $\mathcal{H}_A$ relative to $\{|x\rangle\}_{x}$.

Proof. For a nonnegative operator $E_{AB}$ with $\text{tr}A E_{AB} \leq \text{id}_B$, we can define the operator
\[
E'_{AB} = E_{AB} + \kappa_A \otimes (\text{id}_B - \text{tr}_A(E_{AB}))
\]
where $\kappa_A$ is an arbitrary normalized density operator on $\mathcal{H}_A$. We then have
\[
\text{tr}(\rho_{AB} E'_{AB}) \geq \text{tr}(\rho_{AB} E_{AB})
\]
with $E'_{AB} \geq 0$ and $\text{tr}_A(E'_{AB}) = \text{id}_B$. This shows that we can extend the maximization on the rhs. of (29) to all operators $E_{AB}$ whose partial trace $\text{tr}_A(E_{AB})$ is bounded by $\text{id}_B$ (instead of being equal to $\text{id}_B$). The claim is therefore equivalent to
\[
\min_{\sigma_B \geq 0} \text{tr}(\sigma_B) = \max_{E_{AB} \geq 0} \text{tr}(\rho_{AB} E_{AB}) \tag{30}
\]
To relate this to the general linear programming problem (25), we define $\mathcal{V}_1 = \text{Herm}(\mathcal{H}_B)$ and $\mathcal{V}_2 = \text{Herm}(\mathcal{H}_A \otimes \mathcal{H}_B)$ as the (real) vector spaces of Hermitian operators on $\mathcal{H}_B$ and $\mathcal{H}_A \otimes \mathcal{H}_B$, respectively, with standard Hilbert-Schmidt inner product. Furthermore, we define the convex cones $K_1$ and $K_2$ as the set of nonnegative operators in $\text{Herm}(\mathcal{H}_B)$ and $\text{Herm}(\mathcal{H}_A \otimes \mathcal{H}_B)$, respectively. We claim that these cones are self-dual, i.e., $K_1^* = K_1$ and $K_2^* = K_2$. This is easily seen from the spectral decomposition of a Hermitian operator. Finally, we define $E : \mathcal{V}_1 \rightarrow \mathcal{V}_2$ as the linear map $E(\theta_B) = E\theta_B := \text{id}_A \otimes \theta_B$. It is easy to check that the adjoint $E^* : \mathcal{V}_2 \rightarrow \mathcal{V}_1$ is equal to the partial trace $\text{tr}_A : \text{Herm}(\mathcal{H}_A \otimes \mathcal{H}_B) \rightarrow \text{Herm}(\mathcal{H}_B)$; indeed, for all $\theta_B \in \text{Herm}(\mathcal{H}_B)$ and $F_{AB} \in \text{Herm}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we have
\[
\langle E\theta_B, F_{AB} \rangle = \text{tr}((\text{id}_A \otimes \theta_B) F_{AB}) = \text{tr}(\theta_B \text{tr}_A(F_{AB})).
\]
We also set $b = \rho_{AB}$ and $c = \text{id}_B$. With these definitions, we conclude that the two optimization problems defined by (30) are a special instance of (25); the claim is equivalent to the statement that the duality gap vanishes. According to Lemma 4, it suffices to check Slater’s interiority condition. For this purpose, we set $v = 2\lambda_{\text{max}}(\rho_{AB}) \cdot \text{id}_B$, where $\lambda_{\text{max}}$ denotes the maximal eigenvalue. Clearly, $v$ is in the interior of $K_1$. We also have
\[
E v - b = 2\lambda_{\text{max}}(\rho_{AB}) \text{id}_B - \rho_{AB} > 0,
\]
hence $Ev - b \in \text{int} K_2$; this proves the claim (30).

To prove the claim about the case where $\rho_{AB}$ is classical relative to an orthonormal basis $\{|x\rangle\}_{x}$ of $\mathcal{H}_A$, we simply set $\mathcal{V}_2 = \text{span}\{|x\rangle\}_{x} \otimes \text{Herm}(\mathcal{H}_B)$ equal to the set of Hermitian operators that are classical on $\mathcal{H}_A$. The remainder of the proof is identical to the general case. \hfill $\square$

Observe that the lhs. of (29) is equivalent to a minimization of the distance measure $D_\infty$ from $\mathcal{H}_A$, i.e., we have
\[
\log \min_{\sigma_B \geq 0} \text{tr}(\sigma_B) = \min_{\sigma_B \geq 0} D_\infty(\rho_{AB}||\text{id}_A \otimes \sigma_B) = -H_{\text{min}}(A|B)_\rho. \tag{31}
\]
Let us discuss the case where $\rho_{XB}$ is classical on $X$. Lemma 4 allows us to show that the min-entropy $H_{\text{min}}(X|B)_\rho$ is equivalent to the “guessing-entropy” of $X$ given $B$.

Theorem 1. Let $\rho_{XB} = \sum_x p_x |x\rangle \otimes \rho_{B|x}$ be classical on $\mathcal{H}_X$. Then
\[
H_{\text{min}}(X|B)_\rho = -\log p_{\text{guess}}(X|B)_\rho,
\]
where $p_{\text{guess}}(X|B)_\rho$ is the maximal probability of decoding $X$ from $B$ with a POVM $\{E_{B|x}\}_x$ on $\mathcal{H}_B$, i.e.,
\[
p_{\text{guess}}(X|B)_\rho := \max_{\{E_{B|x}\}_x} \sum_x p_x \text{tr}(E_{B|x} \rho_{B|x}).
\]
Proof. According to (31), it suffices to show that the rhs. of (29) is equal to $p_{\text{guess}}(X|B)_\rho$. But this is a direct consequence of the fact that every nonnegative operator $E_{XB}$ with $\text{tr}_X(E_{XB}) = \text{id}_B$ which is classical on $\mathcal{H}_X$ has the form
\[
E_{XB} = \sum_x |x\rangle \langle x| \otimes E_{B|x}^x,
\]
where the family $\{E_{B|x}^x\}_x$ is a POVM on $\mathcal{H}_B$. \hfill $\square$

The Choi-Jamiołkowski isomorphism yields an operational interpretation of the min-entropy in the general case. We can express the min-entropy as the maximal achievable singlet fraction as follows.

Theorem 2. The min-entropy of a state $\rho_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ can be expressed as
\[
H_{\text{min}}(A|B)_\rho = -\log q_{\text{corr}}(A|B)_\rho, \tag{32}
\]
where $q_{\text{corr}}(A|B)_\rho$ is the maximal achievable singlet fraction, i.e.,
\[
q_{\text{corr}}(A|B)_\rho := d_A \max_F \langle (\text{id}_A \otimes F)(\rho_{AB}), |\Phi_{AA'}\rangle |\Phi_{AA'}\rangle \rangle^2,
\]
with maximum taken over all quantum operations $F : \mathcal{L}(\mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_A')$, $\mathcal{H}_A' \cong \mathcal{H}_A$ and $|\Phi_{AA'}\rangle$ defined by (28).
Proof. Let us rewrite statement (32) as
\[
\min_{\sigma_B \geq 0} D_\infty(\rho_{AB} || \id_A \otimes \sigma_B) = \log \left( d_A \cdot \max_{F} F((\id_A \otimes F)(\rho_{AB}), |\Phi_{AA'}\rangle \langle \Phi_{AA'}|)^2 \right),
\]
(33)
where $|\Phi_{AA'}\rangle$ is the maximally entangled state. Let $E_{AB}$ be a nonnegative operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\Tr E_{AB} = \id_B$, and let $\mathcal{E} = J^{-1}(E_{AB}) \in \text{CPUM}(\mathcal{H}_A, \mathcal{H}_B)$ be the unit map corresponding to $E_{AB}$ under the Choi-Jamiolkowski isomorphism (cf. Lemma 3 (iii)). Let $\mathcal{F} = \mathcal{E}^\dagger : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_A)$ be the adjoint quantum operation (cf. Lemma 2). By definition of $\mathcal{E}$ and the adjoint (id$_A$ $\otimes$ $\mathcal{E}$)$^\dagger$ = id$_A$ $\otimes$ $\mathcal{E}$ $\otimes$ $\mathcal{F}$, we have
\[
\Tr(\rho_{AB} E_{AB}) = d_A \Tr(\rho_{AB}(\id_A \otimes \mathcal{E})(|\Phi_{AA'}\rangle \langle \Phi_{AA'}|)) = d_A \Tr((\id_A \otimes \mathcal{E})^\dagger(\rho_{AB})|\Phi_{AA'}\rangle \langle \Phi_{AA'}|) = d_A \Tr((\id_A \otimes \mathcal{F})(\rho_{AB})|\Phi_{AA'}\rangle \langle \Phi_{AA'}|).
\]
Observe that the operators $E_{AB} \succeq 0$ with $\Tr E_{AB} = \id_B$ are in one-to-one correspondence with quantum operations $\mathcal{F} \in \text{CPUM}(\mathcal{H}_B, \mathcal{H}_A)$ constructed in this fashion. The claim (33) therefore follows from Lemma 4 and (31). \hfill $\blacksquare$

Remark 1. The result of Theorem 2 can be extended to give an alternative expression for the maximal achievable fidelity with a non-maximally entangled state $|\Psi_{AA'}\rangle = \sum_{\lambda} \sqrt{\lambda} |\lambda\rangle_A |\lambda\rangle_A \in \mathcal{H}_A \otimes \mathcal{H}_A$. We assume that $\mathcal{H}_A' \cong \mathcal{H}_A$ and that $|\Psi_{AA'}\rangle$ has maximal Schmidt rank. Let $\tau_A = \Tr_A |\Psi_{AA'}\rangle \langle \Psi_{AA'}|$ be its reduced density operator. Then
\[
\min_{\sigma_B \geq 0} D_\infty(\rho_{AB} || \tau_A^{-1} \otimes \sigma_B) = \log \max_{\mathcal{F}} F((\id_A \otimes F)(\rho_{AB}), |\Psi_{AA'}\rangle \langle \Psi_{AA'}|)^2
\]
for any bipartite state $\rho_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$. Statement (34) follows by substituting $(\tau_A^{1/2} \otimes \id_B) \rho_{AB} (\tau_A^{1/2} \otimes \id_B)$ for $\rho_{AB}$ in (33), using the fact that conjugating with an invertible matrix does not change operator inequalities, and $\sqrt{d_A} \cdot (\tau_A^{1/2} \otimes \id_B)(|\Phi_{AA'}\rangle = |\Psi_{AA'}\rangle$.

B. Proof of the operational characterization of $H_{\text{max}}$

To obtain the operational characterization of $H_{\text{max}}$, we use Theorem 2. Recall the definition of the decoupling accuracy of a bipartite state $\rho = \rho_{AB}$, that is,
\[
q_{\text{deep}}(A|B)_\rho := d_A \max_{\sigma_B} F(\rho_{AB}, \tau_A \otimes \sigma_B)^2,
\]
where $d_A$ is the dimension of $\mathcal{H}_A$ and $\tau_A$ is the completely mixed state on $\mathcal{H}_A$. We begin by showing the following lower bound on the decoupling accuracy.

Lemma 5. For all bipartite states $\rho_{AB}$, we have
\[
2^{H_{\text{max}}(A|B)_\rho} \leq q_{\text{deep}}(A|B)_\rho.
\]
Proof. Let $\rho_{ABC} = |\varphi_{ABC}\rangle \langle \varphi_{ABC}|$ be a purification of $\rho_{AB}$, and let $\mathcal{F} : \mathcal{L}(\mathcal{H}_C) \rightarrow \mathcal{L}(\mathcal{H}_A)$ be a quantum operation that satisfies (cf. Theorem 2)
\[
-2 \cdot H_{\min}(A|C)_\rho = d_A F((\id_A \otimes \mathcal{F})(\rho_{AC}), |\Phi_{AA'}\rangle \langle \Phi_{AA'}|)^2.
\]
Let $\rho'_{AA'BR} = |\varphi'_{AA'BR}\rangle \langle \varphi'_{AA'BR}|$ be a purification of $\rho'_{AB} = (\id_A \otimes \mathcal{F})(\rho_{AB})$. We then have
\[
2^{H_{\max}(A|B)_\rho} = 2^{-H_{\min}(A|C)_\rho} = d_A F(\rho'_{AA'}, |\Phi_{AA'}\rangle \langle \Phi_{AA'}|)^2
\]
(35)
However,
\[
F(\rho'_{AA'}, |\Phi_{AA'}\rangle \langle \Phi_{AA'}|) = F(|\varphi'_{AA'BR}\rangle \langle \varphi'_{AA'BR}|, |\Phi_{AA'}\rangle \langle \Phi_{AA'}| \otimes |\theta_{BR}\rangle \langle \theta_{BR}|)
\]
for some state $|\theta_{BR}\rangle$ on $\mathcal{H}_B \otimes \mathcal{H}_R$. By the monotonicity of the fidelity, we therefore get
\[
F(\rho'_{AA'}, |\Phi_{AA'}\rangle \langle \Phi_{AA'}|) \leq F(\rho_{AB}, \tau_A \otimes \Tr_{\mathcal{B}} |\theta_{BR}\rangle \langle \theta_{BR}|) \leq \max_{\sigma_B} F(\rho_{AB}, \tau_A \otimes \sigma_B),
\]
where we used the fact that $\rho'_{AB} = \rho_{AB}$. Inserting this into (35) gives the claim. \hfill $\blacksquare$

The proof of the converse inequality closely follows a derivation in [13]. We include it here for completeness.

Lemma 6. For all bipartite states $\rho_{AB}$, we have
\[
2^{H_{\max}(A|B)_\rho} \geq q_{\text{deep}}(A|B)_\rho.
\]
Proof. We use the following fact, which is a consequence of the fact that all purifications of a fixed state are related by a unitary transformation on a (possibly extended) ancilla. If $|\phi_{ABCC'}\rangle$ has a reduced state of the form $\Tr_C |\phi_{ABCC'}\rangle \langle \phi_{ABCC'}| = \tau_A \otimes \sigma_B$, where $\tau_A$ is the completely mixed state on $\mathcal{H}_A$, then there exists a unitary $U_{CC'}$ such that
\[
|\id_A \otimes \id_B \otimes U_{CC'}\rangle |\phi_{ABCC'}\rangle \langle \phi_{ABCC'}| = |\Psi_{AC}\rangle \langle \Psi_{BC}|.
\]
(36)
for some state $|\theta_{BC}\rangle$ on $\mathcal{H}_B \otimes \mathcal{H}_C$, where $|\Psi_{AC}\rangle$ denotes the completely entangled state on $\mathcal{H}_A \otimes \mathcal{H}_C$ (without loss of generality, we can assume that $d_A \leq d_C$).

Let $\sigma_B$ be an arbitrary state on $\mathcal{H}_B$. Let $\rho_{ABC} = |\psi_{ABC}\rangle \langle \psi_{ABC}|$ be a purification of $\rho_{AB}$, where we assume the dimension of $\mathcal{H}_C$ to be sufficiently large.

According to the definition of the fidelity, there exists a purification $|\phi_{ABCC'}\rangle$ of $\tau_A \otimes \sigma_B$ such that
\[
F(\rho_{AB}, \tau_A \otimes \sigma_B) = F(|\psi_{ABC}\rangle \langle 0_{C'}|, |\phi_{ABCC'}\rangle \langle \phi_{ABCC'}|).
\]
Applying the unitary $U_{CC'}$ from (36) gives
\[
F(\rho_{AB}, \tau_A \otimes \sigma_B) = F(|\psi'_{ABCC'}\rangle \langle \psi'_{ABCC'}|, |\Psi_{AC}\rangle \langle \Psi_{BC}|),
\]
where $|\psi'_{ABCC'}\rangle = (\id_A \otimes \id_B \otimes U_{CC'} |\psi_{ABC}\rangle \langle 0_{C'}|$ because of the invariance of the fidelity under unitary operations. Using the monotonicity of the fidelity, we conclude that
\[
F(\rho_{AB}, \tau_A \otimes \sigma_B) \leq F(\tau_{BC} |\psi'_{ABCC'}\rangle \langle \psi'_{ABCC'}|, |\Psi_{AC}\rangle \langle \Psi_{BC}|)
\]
\[
= F(|\id_A \otimes \mathcal{F}(\rho_{AC}), |\Phi_{AC}\rangle \langle \Phi_{AC}|).
\]
where $F : \mathcal{L}(\mathcal{H}_C) \to \mathcal{L}(\mathcal{H}_C)$ is the quantum operation $F(\rho) = \text{tr}_{C'}(U_{CC'}(\rho \otimes |0\rangle \langle 0|_{C'}) U_{CC'}^T)$. Squaring both sides of the previous inequality, multiplying by $d_A$, taking the maximum over all quantum operations and using Theorem 2 therefore gives

$$d_A F(\rho_{AB}, \tau_A \otimes \sigma_B)^2 \leq 2^{-H_{\min}(A|C)_\rho}.$$ 

Since $\sigma_B$ was arbitrary, we can maximize the lhs. over all $\sigma_B$. The claim then follows from the definitions of $q_{\text{decpl}}(A|B)_\rho$ and $H_{\max}(A|B)_\rho$.

In summary, we have shown the following result.

**Theorem 3.** Let $\rho_{AB}$ be a state on $\mathcal{H}_A \otimes \mathcal{H}_B$, and let $\tau_A$ be the completely mixed state on $\mathcal{H}_A$. Then

$$H_{\max}(A|B)_\rho = \log q_{\text{decpl}}(A|B)_\rho,$$

where $q_{\text{decpl}}(A|B)_\rho$ is the decoupling accuracy, defined by

$$q_{\text{decpl}}(A|B)_\rho := d_A \max_{\sigma_B} F(\rho_{AB}, \tau_A \otimes \sigma_B)^2,$$

with the maximum taken over all normalized states $\sigma_B$ on $\mathcal{H}_B$.

**IV. CONCLUSIONS**

In information theory, entropies are generally interpreted as measures of uncertainty. One method to make this interpretation more precise is to establish relations between entropy measures and operational quantities, that is, quantities that characterize actual information-theoretic tasks.

Here we consider a general scenario consisting of a (possibly quantum-mechanical) system $A$ as well as an observer with (quantum or classical) side information $B$. The uncertainty of the observer about the state of system $A$ then depends on the distribution of these states as well as the correlation between $A$ and $B$.

There are two extreme situations, namely when $A$ is completely undetermined and when $A$ is determined. Taking into account the side information $B$, these two situations are described as follows.

1. The state of $A$ is fully correlated with (parts of) $B$.\footnote{In the general case where $A$ and $B$ are quantum-mechanical systems, full correlation is akin to maximal entanglement.}

2. The state of $A$ is uniformly distributed and independent of the side information $B$.

Note that in the first case, the requirement is merely that $A$ is correlated with parts of $B$. This is because the side-information $B$ may consist of additional information that is unrelated to $A$.

For any given state $\rho_{AB}$, we may characterize the uncertainty of $A$ given $B$ by the distance to these extreme situations.

If we take as a distance measure the overlap (i.e., the square of the fidelity), we retrieve the definitions of $q_{\text{corr}}(A|B)$ and $q_{\text{decpl}}(A|B)$ (see (20) and (22), respectively). Our main results imply that these correspond to $H_{\min}(A|B)$ and $H_{\max}(A|B)$, respectively. We thus conclude that $H_{\min}(A|B)$ quantifies the closeness to a situation where $A$ is determined by $B$, and, likewise, $H_{\max}(A|B)$ corresponds to the closeness to a situation where $A$ is independent of $B$ (see second column of Table I).

Given a bipartite state $\rho_{AB}$, we may also ask for the number of maximally entangled or completely independent qubits one can extract from $A$. Very roughly speaking, this is the idea underlying the definitions of $\ell_{\text{merg}}(A|B)$ and $\ell_{\text{decpl}}(A|B)$, respectively (see Section I B for more details, in particular the interpretation of negative quantities). Remarkably, these quantities are (approximately) given by the smooth entropies $H_{\max}(A|B)$ and $H_{\min}(A|B)$ (see last column of Table I).\footnote{Note that compared to the discussion of the distance, the role of max and min is interchanged.}

Despite these similarities between the (previously known) operational interpretations summarized in the last column of Table I and those given in the second column (the ones derived here), there are at least two fundamental differences. The first is that the new interpretations are exact and, in particular, valid without a smoothness parameter. In contrast, all previously established interpretations only hold up to additive terms of the order $\log \frac{1}{\epsilon}$, where $\epsilon$ is a smoothness parameter (whose meaning is that of an error or failure probability). A second difference is that there does not seem to exist an obvious asymptotic counterpart for our identities. In particular, there are no analogous operational interpretations of the von Neumann entropy.

The results of this paper suggest that studying operationally defined quantities may be a viable approach to identifying relevant single-shot information measures in a multipartite setting. Of particular interest is the conditional mutual information, which has only recently been given an asymptotic interpretation [17].

**Acknowledgments**

RK acknowledges support by NSA under ARO contract no. W911NF-05-1-0294 and by NSF under contract no. PHY-0456720. CS is supported by EU fifth framework project QAP IST 015848 and the NWO VICI project 2004-2009.
| goal (extremal state) | quality of a state $\rho_{AB}$ (measured in terms of overlap) | amount of extremal states contained in a state $\rho_{AB}$ (measured in # of qubits on $A$) |
|----------------------|-----------------------------------------------------|--------------------------------------------------|
| $A$ fully entangled with $B$ | $-\log q_{\text{corr}}(A|B)$ \hspace{1cm} $-\log p_{\text{corr}}(A|B)$ $= H_{\min}(A|B)$ | $\varepsilon_{\text{merg}}(A|B) \approx -H_{\varepsilon'}(A|B)$ |
| classical $A$ fully determined by $B$ | $-\log q_{\text{decpl}}(A|B)$ \hspace{1cm} $-\log p_{\text{secr}}(A|B)$ $= H_{\max}(A|B)$ | $\varepsilon_{\text{decpl}}(A|B) \approx H_{\varepsilon'}(A|B)$ |

TABLE I: Operational interpretations of (smooth) min- and max-entropies. The approximation ($\approx$) indicates that equality holds up to an additive term of order $\log \frac{1}{\varepsilon}$ and for an appropriate choice of the smoothness parameter $\varepsilon'$. 

[1] C. E. Shannon, Bell System Technical Journal **27**, 379 (1948). 
[2] S. Verdú and T. S. Han, IEEE Transactions on Information Theory **40**, 1147 (1994). 
[3] R. Renner and S. Wolf, in *Proc. International Symposium on Information Theory* (IEEE, 2004), p. 233. 
[4] N. Datta and R. Renner (2008), arXiv:0801.0282. 
[5] R. Renner, Ph.D. thesis, ETH Zurich (2005), arXiv.org:quant-ph/0512258. 
[6] R. Renner, S. Wolf, and J. Wullschleger, in *Proceedings of 2006 IEEE International Symposium on Information Theory* (2006), pp. 1424–1427. 
[7] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer, IEEE Transaction on Information Theory **41**, 1915 (1995). 
[8] R. König and R. Renner, *Sampling of min-entropy relative to quantum knowledge* (2007), arXiv.org:0712.4291. 
[9] R. Impagliazzo, L. A. Levin, and M. Luby, in *STOC ’89: Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, USA, 1989), pp. 12–24, ISBN 0-89791-307-8. 
[10] M. Horodecki, J. Oppenheim, and A. Winter, *Quantum state merging and negative information* (2005), arXiv.org:quant-ph/0512247. 
[11] A. Winter and R. Renner, *Single-shot state merging* (2007), unpublished note. 
[12] M. Berta, Master’s thesis, ETH Zurich (2008). 
[13] M. Horodecki, S. Lloyd, and A. Winter, Open Systems and Information Dynamics **15**, 47 (2008), arXiv.org:quant-ph/0702006. 
[14] I. B. Damgård, S. Fehr, R. Renner, L. Salvail, and C. Schaffner, in *Advances in Cryptology—CRYPTO ’07* (Springer-Verlag, 2007), vol. 4622 of *Lecture Notes in Computer Science*, pp. 360–378. 
[15] A. Barvinok, *A Course in Convexity*, no. 54 in Graduate Studies in Mathematics (American Mathematical Society, Providence, 2002). 
[16] A. Jamiołkowski, Rev. of Mod. Phys. **3**, 275 (1972). 
[17] I. Devetak and J. Yard, *The operational meaning of quantum conditional information* (2006), arXiv:quant-ph/0612050.