The Single-Serving Channel Capacity

Renato Renner*  Stefan Wolf†  Jürg Wullschleger‡
*Centre for Quantum Computation, University of Cambridge, United Kingdom
E-mail: r.renner@damtp.cam.ac.uk
†Computer Science Department, ETH Zürich, Switzerland
E-mail: {wolf,juerg}@inf.ethz.ch

Abstract—In this paper we provide the answer to the following question: Given a noisy channel \( P_{Y|X} \) and \( \varepsilon > 0 \), how many bits can be transmitted with an error of at most \( \varepsilon \) by a single use of the channel?

I. INTRODUCTION

Shannon entropy and information [14] have been shown very significant in the scenario of i.i.d. distributions and asymptotic rates. Unfortunately, however, these two assumptions fail to be realistic in many real-world scenarios. First of all, a given primitive or random experiment is actually available only a limited number of times, and an asymptotic analysis has, therefore, a limited significance. Second, the assumption that a certain primitive is repeated \( \text{independently} \) many times is not always realistic. An important example is cryptography, where this assumption leads to a strong restriction on the adversary’s behavior and possibilities.

In [6], the assumption of independence has been dropped, but the analysis still remains asymptotic. In the present paper, we drop both assumptions at once and consider the case where a certain information-theoretic primitive, such as a communication channel, or random experiment is available only \( \text{once} \). This single-serving case has also been called ”single shot” in the literature.

Let us first consider an example from cryptography or, more precisely, information-theoretic key agreement from correlated pieces of information. Let two parties, Alice and Bob, as well as an adversary, Eve, have access to \( n \) independent realizations of random variables \( X, Y, \) and \( Z \), respectively, with joint probability distribution \( P_{X,Y,Z} \). Moreover, authenticated but public communication from Alice to Bob (but not in the other direction) is possible. Their goal is to generate a common secret key of length \( \ell(n) \), i.e., a uniform string about which the adversary is virtually ignorant. Asymptotically, for large \( n \), the rate at which such a key can be generated is given by

\[
\lim_{n \to \infty} \frac{\ell(n)}{n} = \max_{Y \sim X \sim U \sim V} (H(U|ZV) - H(U|YV))
\]

(see, for instance, [17], [3], [1], [10]).

Let us now consider the non-asymptotic case where \( n = 1 \), i.e., the random experiment defined by \( P_{X,Y,Z} \) is only run \( \text{once} \). How many virtually secret bits can then be extracted? First of all, note that (1) fails to provide the correct answer in this case. To see this, assume, e.g., that \( X \) is uniformly distributed and that \( Y = X \), whereas \( Z = X \) holds with probability 1/2 (and \( Z = \Delta \) otherwise). Then the right-hand side of (1) is non-zero, but no secret can be extracted at all by Alice and Bob since, with probability 1/2, Eve knows everything. We conclude that Shannon entropy fails to be the right measure in this setting. But what does it have to be replaced by?

Results on randomness extraction, also known as privacy amplification [2], [9], [8], indicate that the right answer might be given by so-called min-entropies rather than Shannon entropies. Indeed, it is shown in [13] that the so-called conditional smooth min- and max-entropies [12], [13] \( H^\varepsilon_{\text{min}} \) and \( H^\varepsilon_{\text{max}} \) (for the precise definitions see below) replace Shannon entropy in this case; the achievable secret-key length \( \ell \) is approximated (up to a term \( \log(1/\varepsilon) \), where \( \varepsilon \) is the security of the final key) by

\[
\ell \approx \max_{Y \sim X \sim U \sim V} (H^\varepsilon_{\text{min}}(U|ZV) - H^\varepsilon_{\text{max}}(U|YV))
\]

It is the goal of this paper to show that smooth min- and max-entropy has a similar significance in communication theory, i.e., it can be used for the characterization of communication tasks in a single-serving setting. Among others, we consider the following question: Given a noisy communication channel \( W = P_{Y|X} \) and \( \varepsilon > 0 \), what is the maximum number \( C^\varepsilon_{\text{comm}}(W) \) of bits that can be transmitted with error at most \( \varepsilon \) by a single use of the channel. Recall that, in the i.i.d. case, i.e., if the channel can be used many times independently, an asymptotic answer to this question is given by the channel capacity \( C_{\text{comm}} \), which can be expressed by the well-known formula [14]

\[
C_{\text{comm}}(W) = \max_{P_X} (H(X) - H(X|Y))
\]

As we shall see, the answer for the single-serving case looks very similar, but the (conditional) Shannon entropies are replaced by smooth min- and max-entropies:

\[
C^\varepsilon_{\text{comm}}(W) \approx \max_{P_X} (H^\varepsilon_{\text{min}}(X) - H^\varepsilon_{\text{max}}(X|Y))
\]

II. NOTATION AND PREVIOUS WORK

A. Smooth Min- and Max-Entropies

Let \( X \) be a random variable with probability distribution \( P_X \). The max-entropy of \( X \) is defined as the binary logarithm of the size of the support of \( P_X \), i.e.,

\[
H_{\text{max}}(X) = \log |\{x \in \mathcal{X} : P_X(x) > 0\}|
\]

Non-zero, but no secret can be extracted at all by Alice and Bob since, with probability 1/2, Eve knows everything. We conclude that Shannon entropy fails to be the right measure in this setting. But what does it have to be replaced by?

Results on randomness extraction, also known as privacy amplification [2], [9], [8], indicate that the right answer might be given by so-called min-entropies rather than Shannon entropies. Indeed, it is shown in [13] that the so-called conditional smooth min- and max-entropies [12], [13] \( H^\varepsilon_{\text{min}} \) and \( H^\varepsilon_{\text{max}} \) (for the precise definitions see below) replace Shannon entropy in this case; the achievable secret-key length \( \ell \) is approximated (up to a term \( \log(1/\varepsilon) \), where \( \varepsilon \) is the security of the final key) by

\[
\ell \approx \max_{Y \sim X \sim U \sim V} (H^\varepsilon_{\text{min}}(U|ZV) - H^\varepsilon_{\text{max}}(U|YV))
\]

It is the goal of this paper to show that smooth min- and max-entropy has a similar significance in communication theory, i.e., it can be used for the characterization of communication tasks in a single-serving setting. Among others, we consider the following question: Given a noisy communication channel \( W = P_{Y|X} \) and \( \varepsilon > 0 \), what is the maximum number \( C^\varepsilon_{\text{comm}}(W) \) of bits that can be transmitted with error at most \( \varepsilon \) by a single use of the channel. Recall that, in the i.i.d. case, i.e., if the channel can be used many times independently, an asymptotic answer to this question is given by the channel capacity \( C_{\text{comm}} \), which can be expressed by the well-known formula [14]

\[
C_{\text{comm}}(W) = \max_{P_X} (H(X) - H(X|Y))
\]

As we shall see, the answer for the single-serving case looks very similar, but the (conditional) Shannon entropies are replaced by smooth min- and max-entropies:

\[
C^\varepsilon_{\text{comm}}(W) \approx \max_{P_X} (H^\varepsilon_{\text{min}}(X) - H^\varepsilon_{\text{max}}(X|Y))
\]
Similarly, the min-entropy of $X$ is given by the negative logarithm of the maximum probability of $P_X$, i.e.,
\[ H_{\min}(X) = -\log(\max_x P_X(x)) \,.
\]
Note that $H_{\min}(X) \leq H(X) \leq H_{\max}(X)$, i.e., the min- and max-entropies are lower and upper bounds for the Shannon entropy (and also for any Rényi entropy of order $\alpha \in [0, \infty]$), respectively.

For random variables $X$ and $Y$ with joint distribution $P_{XY}$, the “conditional” versions of these entropic quantities are defined as
\[ H_{\max}(X|Y) = \max_y H_{\max}(X|Y = y) \,,
\]
\[ H_{\min}(X|Y) = \min_y H_{\min}(X|Y = y) \,.
\]

In [13], max- and min-entropies have been generalized to so-called smooth max- and min-entropies. For any $\varepsilon \geq 0$, they are defined by optimizing the "non-smooth" quantities over all random variables $\tilde{X}$ and $\tilde{Y}$ which are equal to $X$ and $Y$ except with probability $\varepsilon$, i.e.,
\[ H_{\max}^{\varepsilon}(X|Y) = \min_{\tilde{X}Y : Pr[\tilde{X}\tilde{Y} = XY] \leq \varepsilon} H_{\max}(\tilde{X}|\tilde{Y}) \]
\[ H_{\min}^{\varepsilon}(X|Y) = \max_{\tilde{X}Y : Pr[\tilde{X}\tilde{Y} = XY] \leq \varepsilon} H_{\min}(\tilde{X}|\tilde{Y}) \,.
\]

Equivalently, smooth max- and min-entropies can be expressed in terms of a optimization over events $\mathcal{E}$ that have probability at least $1 - \varepsilon$. Let $P_{X\mathcal{E}|Y=y}(x)$ be the probability that $X = x$ and the event $\mathcal{E}$ occurs, conditioned on $Y = y$. We then have
\[ H_{\max}^{\varepsilon}(X|Y) = \min_{\mathcal{E} : Pr[\mathcal{E}] \geq 1 - \varepsilon} \max_y \log |\{x : P_{X\mathcal{E}|Y=y}(x) > 0\}| \]
\[ H_{\min}^{\varepsilon}(X|Y) = \max_{\mathcal{E} : Pr[\mathcal{E}] \geq 1 - \varepsilon} \min_x (\log P_{X\mathcal{E}|Y=y}(x)) \,.
\]

These smooth entropies have properties similar to Shannon entropy—this is in contrast to the usual, non-smooth min- and max-entropies which have many counterintuitive properties that make them less useful in many contexts. For example, the chain rule $H(X|Y) = H(X) - H(Y)$ translates to [13]
\[ H_{\max}^{\varepsilon}(X|Y) - H_{\min}^{\varepsilon}(Y) \leq H_{\max}^{\varepsilon}(X|Y) \,.
\]
\[ \leq H_{\max}^{\varepsilon}(XY) - H_{\min}^{\varepsilon}(Y) + \log(1/(\varepsilon - \varepsilon_1 - \varepsilon_2)) \]
and
\[ H_{\min}^{\varepsilon}(X|Y) - H_{\max}^{\varepsilon}(Y) - \log(1/(\varepsilon - \varepsilon_1 - \varepsilon_2)) \]
\[ \leq H_{\min}^{\varepsilon}(X|Y) - H_{\min}^{\varepsilon}(XY) - H_{\min}^{\varepsilon}(Y) \,.
\]

B. Operational Interpretation of Smooth Max- and Min-Entropies

In [15] it was shown that the rate at which many independent realizations of $X$ can be compressed is asymptotically $H(X|Y)$ if the decoder is provided with side information $Y$. It is easy to see that $H(X|Y)$ also is the rate at which uniform randomness can be extracted from $X$, in such a way that it is independent of $Y$. In [13], it was shown that the smooth entropies $H_{\max}^{\varepsilon}$ and $H_{\min}^{\varepsilon}$ quantify compression and randomness extraction, respectively, in the single-serving case. More precisely, let $H_{\text{comp}}^{\varepsilon}(X|Y)$ be the length of a bit string needed to store one instance of $X$ such that $X$ can later be recovered with an error of at most $\varepsilon$ using this string and $Y$. This quantity is then roughly equal to $H_{\max}^{\varepsilon}$, i.e.,
\[ H_{\max}^{\varepsilon}(X|Y) \leq H_{\text{comp}}^{\varepsilon}(X|Y) \]
\[ \leq H_{\max}^{\varepsilon}(X|Y) + \log(1/(\varepsilon - \varepsilon)) \,.
\]

Similarly, let $H_{\text{ext}}^{\varepsilon}(X|Y)$ be the maximum length of a string that can be computed from $X$, such that this string is uniformly distributed and independent of $Y$, with an error of at most $\varepsilon$. We then have
\[ H_{\text{ext}}^{\varepsilon}(X|Y) - 2 \log(1/(\varepsilon - \varepsilon')) \leq H_{\min}^{\varepsilon}(X|Y). \]

C. Common Information

The common information is the rate at which uniform random bits can be extracted both from $X^n$ and $Y^n$, which come from independent repeated realizations of the random experiment $P_{XY}$ without communicating. It has been shown in [5] that the common information is equal to the maximum entropy of a common random variable that both players can compute. As in [4], [16], we will denote this random variable by $X \land Y$, i.e., the common information of $X$ and $Y$ is given by $H(X \land Y)$.

It is shown in [16] that the common information can be used to characterize the zero-error capacity $C_{\text{sym}}^{\text{comm}}(W)$ of a channel $W$ as follows:
\[ C_{\text{sym}}^{\text{comm}}(W) = \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} H(X^n \land Y^n) \,.
\]

Note that the usual (asymptotic) channel capacity $C_{\text{sym}}^{\text{comm}}(W)$ of $W$ is given by a similar expression, where the common information is replaced by the mutual information, i.e.,
\[ C_{\text{sym}}^{\text{comm}}(W) = \max_{P_{X^n}} I(X^n; Y^n) = \lim_{n \to \infty} \frac{1}{n} I(X^n; Y^n) \,.
\]

III. Extractable Common Randomness

We denote by $C_{\text{ext}}^{\varepsilon}(X, Y)$ the maximum amount of uniform randomness that can be extracted from $X$ and $Y$, without any communication, with an error of at most $\varepsilon$. Asymptotically, it follows from [5] that
\[ \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{C_{\text{ext}}^{\varepsilon}(X^n, Y^n)}{n} = H(X \land Y). \]

In the following, we analyze the quantity $C_{\text{ext}}^{\varepsilon}(X, Y)$ for the single-serving case. First, we will show that $C_{\text{ext}}^{\varepsilon}(X, Y)$ is characterized by the following quantity.

**Definition 1:**
\[ C_{\text{min}}^{\varepsilon}(X; Y) = \max_{\tilde{X} \tilde{Y} : Pr[\tilde{X}\tilde{Y} \neq XY] \leq \varepsilon} H_{\min}(\tilde{X} \land \tilde{Y}) \,.
\]

**Theorem 1:** For all random variables $X$ and $Y$, and for all $\varepsilon' > \varepsilon$, we have
\[ C_{\text{ext}}^{\varepsilon}(X; Y) \geq C_{\text{min}}^{\varepsilon}(X; Y) - 2 \log(1/(\varepsilon - \varepsilon')) \,.
\]

**Proof:** Let Alice and Bob have $\tilde{X}$ and $\tilde{Y}$, respectively. They both can calculate $\tilde{X} \land \tilde{Y}$ and extract at least $H_{\min}(\tilde{X} \land$
Since $\Pr[XY \neq XY] \leq \varepsilon'$, we get at most an additional error of $\varepsilon'$ if they use $X$ and $Y$ instead of $\bar{X}$ and $\bar{Y}$. The total error is, therefore, at most $\varepsilon$.

**Theorem 2:** For all random variables $X$ and $Y$, and for all $\varepsilon$, we have

$$C^\varepsilon_{\text{ext}}(X;Y) \leq C^\varepsilon_{\text{min}}(X;Y).$$

**Proof:** Let us assume that Alice and Bob can extract more than $C^\varepsilon_{\text{min}}(X;Y)$ bits with an error at most $\varepsilon$. Therefore there exist functions $f$ and $g$ such that with probability $1 - \varepsilon$ both functions output the same uniform random string $R$ of length bigger than $C^\varepsilon_{\text{min}}(X;Y)$, which means that there exist $X, Y$ such that $\Pr[(X, Y) \neq (X, Y)] \leq \varepsilon$ and $f(\bar{X}) = g(\bar{Y}) = R$. As shown in Lemma 1 of [16], this implies that $R$ can be computed from $X \wedge Y$, that is, there exists a function $h$ such that $R = h(X \wedge Y)$. The function $h$ could thus be used to extract more than $H_{\min}$ bit from $X \wedge Y$, which is impossible.

In the following, we derive an upper bound on $C^\varepsilon_{\text{min}}(X;Y)$ in terms of smooth min- and max-entropies.

**Lemma 1:** For all random variables $X$ and $Y$, and for all $\varepsilon$, $\varepsilon_1$, and $\varepsilon_2$, we have

$$C^\varepsilon_{\min}(X;Y) \leq H^\varepsilon_{\max}(X) - H^{\varepsilon_1 + \varepsilon_2 + 2\varepsilon}(X|Y) + \log(1/\varepsilon_1).$$

**Proof:** Let $\bar{X}$ and $\bar{Y}$ be the random variables that maximize $C^\varepsilon_{\text{min}}(X;Y)$, and let $\bar{C} = \bar{X} \wedge \bar{Y}$. We have

$$H_{\min}(\bar{C}) \leq H^\varepsilon_{\max}(X) - H^{\varepsilon_1}(X) + \log(1/\varepsilon_1).$$

$C$ is a function of $X$ and of $Y$ with probability at least $1 - \varepsilon$. Therefore, we can bound

$$H^{\varepsilon_2}(X) \leq H^\varepsilon_{\max}(X)$$

and

$$H^{\varepsilon_1 + \varepsilon_2 + \varepsilon}(X|Y).$$

We get

$$H_{\min}(\bar{C}) \leq H^{\varepsilon_2 - \varepsilon}(X) - H^{\varepsilon_1 + \varepsilon_2 + \varepsilon}(X|Y) + \log(1/\varepsilon_1).$$

The statement follows when $\varepsilon$ is added to $\varepsilon_2$.

No non-trivial lower bound is known so far for $C^\varepsilon_{\text{ext}}(X;Y)$. However, one can bound $\max_{P_X} C^\varepsilon_{\text{min}}(X;Y)$. This will turn out to be useful for the considerations in the next section.

**Lemma 2:** For all conditional distributions $P_{Y|X}$ and for all $\varepsilon_1, \varepsilon_2$, and $\varepsilon_3$, we have

$$\max_{P_X} C^\varepsilon_{\text{min}} + \varepsilon_2 - \varepsilon_3 \geq \max_{P_X} \left(H_{\min}^\varepsilon(X) - H_{\max}^\varepsilon(X|Y)\right) + \log(1/\varepsilon_3).$$

**Proof:** Let $P_X$ be the distribution that maximizes $H_{\min}^\varepsilon(X) - H_{\max}^\varepsilon(X|Y)$. There exist random variables $\bar{X}$ and $\bar{Y}$ with $\Pr[XY \neq \bar{X} \bar{Y}] \leq \varepsilon_1 + \varepsilon_2$ such that $H_{\min}(\bar{X}) - H_{\max}(\bar{X}|\bar{Y}) = H_{\min}^\varepsilon(X) - H_{\max}^\varepsilon(X|Y)$. We choose, independently and according to the distribution $P_{\bar{X}}, 2H_{\min}(\bar{X}) - H_{\max}(\bar{X}|\bar{Y}) + \log(1/\varepsilon_3)$ values. Let $\bar{S}$ be the set of these values and let $\bar{X}$ be a random variable that takes on a value in $\bar{S}$ with equal probability. Since $P_{\bar{X}}(x) \cdot 2^{2H_{\min}(\bar{X})} \leq 1$, the probability that a value $x$ chosen according to $P_{\bar{X}}$ is in $S$ is at most

$$P_{\bar{X}}(x) \cdot 2H_{\min}(\bar{X}) - H_{\max}(\bar{X}|\bar{Y}) + \log(1/\varepsilon_3) \leq 2H_{\max}(\bar{X}|\bar{Y}) + \varepsilon_3.$$

Let $\bar{x}$ and $\bar{y}$ be chosen according to the distribution $P_{\bar{X}}P_{\bar{Y}|\bar{X}}$. The probability that there exists a value $\bar{x}' \in S$ such that $\bar{x}' \neq \bar{x}$ and $P_{\bar{Y}|\bar{X}}(\bar{y}, \bar{x}') > 0$ is at most $2H_{\max}(\bar{X}|\bar{Y}) + \log(1/\varepsilon_3) = \varepsilon_3$. Therefore, there exists a function $f$ such that $\Pr[\bar{X} \neq f(\bar{Y})] \leq \varepsilon_3$ holds, and we have

$$C^\varepsilon_{\min}(\bar{X};\bar{Y}) = \max_{P_X} H_{\min}(\bar{X})$$

and

$$H^\varepsilon_{\max}(X) - H^\varepsilon_{\min}(X|Y) + \log(1/\varepsilon_3).$$

The statement now follows from the fact that

$$\max_{P_X} C^\varepsilon_{\text{min}}(X;Y) \geq C^\varepsilon_{\text{min}}(\bar{X};\bar{Y}).$$

**IV. COMMUNICATION**

Let us now come back to the question posed in the abstract. We define the $\varepsilon$ single-serving channel capacity of a channel $W = P_{Y|X}$, denoted $C^\varepsilon_{\text{comm}}(W)$, as the maximum number of bits (i.e., the logarithm of the number of symbols) that can be transmitted in a single use of $W$, such that every symbol can be decoded by an error of at most $\varepsilon$. Theorem III shows the connection between the the extractable common randomness and single-serving channel capacity, similar to the connection between the common information and the zero-error capacity shown in [16].

**Theorem 3:** For all channels $W = P_{Y|X}$ and for $\varepsilon' < \varepsilon$, we have

$$\max_{P_X} C^\varepsilon_{\text{min}}(X;Y) - \log(\varepsilon/(\varepsilon - \varepsilon')) \leq C^\varepsilon_{\text{comm}}(W) \leq \max_{P_X} C^\varepsilon_{\text{min}}(X;Y).$$

**Proof:** Let $C \subset X'$ be a code that can be decoded with an error of at most $\varepsilon$ and let $X$ be uniformly distributed over $C$. Then there exists a $\bar{Y}$ with $\Pr[\bar{Y} = \bar{Y}] \geq 1 - \varepsilon$, such that $X = X' \wedge \bar{Y}$. It follows that

$$\max_{P_X} C^\varepsilon_{\text{min}}(X;Y) \geq C^\varepsilon_{\text{comm}}(W).$$

Let $P_X$ be a distribution that maximizes $\max_{P_X} C^\varepsilon_{\text{min}}(X;Y)$, and let $X, \bar{Y}$ be random variables for which $H(X \wedge \bar{Y}) = C^\varepsilon_{\text{min}}(X;Y)$ holds as well as $\Pr[XY = XY] \geq 1 - \varepsilon'$. Let $\bar{C} := \bar{X} \wedge \bar{Y}$. We can write $C$ as a combination of uniform random variables $C_i$ with $H_{\min}(C_i) = H_{\min}(C)$. More precisely, we have $P_{C} = \sum_i \lambda_i P_{C_i}$, where $P_{C_i}(x) \in \{0, 2^{-H_{\min}(C)}\}$ for all $x$. The support of the random variable $C_i$ which minimizes the error probability defines a code $C_i \subset X'$ that can be decoded with an error of at most $\varepsilon$, if the input is uniformly distributed. Since we need a code that works for any input distribution, we delete all symbols which get decoded with an error bigger than $\varepsilon > \varepsilon'$. From the Markov inequality follows that the reduced code still contains at least $\varepsilon - \varepsilon' 2^{-H_{\min}(C_i)}$ symbols. It follows that

$$C^\varepsilon_{\text{comm}}(W) \geq \max_{P_X} C^\varepsilon_{\text{min}}(X;Y) - \log(\varepsilon/(\varepsilon - \varepsilon')).$$
From Lemma 1 we have
\[
\max_P \epsilon \left( C_{\text{sem}}^e(X; Y) \right) \leq \max_P \left( H_{\text{max}}^e(X) - H_{\text{max}}^{e_1 + 2\epsilon}(X|Y) \right) + \log \frac{1}{\epsilon_1}.
\]
From the same argument as in the proof of Theorem 3 follows that \( \max_P C_{\text{sem}}^e(X; Y) \) is maximized by a distribution where all \( x \) with positive probability have equal probabilities. Therefore, we have \( H_{\text{max}}^e(x) = H_{\text{min}}^e(x) \) and get
\[
\max_P \left( H_{\text{min}}^e(X) - H_{\text{max}}^{e_1 + 2\epsilon}(X|Y) \right) - \log \frac{1}{\epsilon - \epsilon' - \epsilon''} \leq \max_P C_{\text{sem}}^e(X; Y)
\]
\[
\leq \max_P \left( H_{\text{min}}^{e_1 + 2\epsilon}(X) - H_{\text{max}}^{e_1 + 2\epsilon}(X|Y) \right) + \log \frac{1}{\epsilon_1}.
\]
Together with Theorem 4 this implies the following bound on the single-serving channel capacity \( C_{\text{comm}}^e(W) \).

**Theorem 4:** For all channels \( W = P_{Y|X} \) and all \( \epsilon', \epsilon'', \epsilon > \epsilon' + \epsilon'' \), \( \epsilon_1 \), and \( \epsilon_2 \), we have
\[
\max_P \left( H_{\text{min}}^{e_1 + 2\epsilon}(X) - H_{\text{max}}^{e_1 + 2\epsilon}(X|Y) \right) - \log \frac{4\epsilon}{(\epsilon - \epsilon' - \epsilon'')^2} \leq C_{\text{comm}}^e(W)
\]
\[
\leq \max_P \left( H_{\text{min}}^{e_1 + 2\epsilon}(X) - H_{\text{max}}^{e_1 + 2\epsilon}(X|Y) \right) + \log \frac{1}{\epsilon_1}.
\]

**V. CONCLUSIONS**

Shannon entropy can be used to characterize a variety of information-processing tasks such as communication over noisy channels in the scenario where the primitive can be used independently many times. We have shown that smooth min- and max-entropies play a similar role in the more general single-serving case. In particular, we have given an explicit expression for the “single-serving channel capacity.” We suggest as an open problem to find other such examples and contexts.

The notion of conditional smooth entropies has recently been generalized to quantum information theory [11] (see also [7] for the non-conditional case). It is likely (but still unproven) that, similarly to our classical Theorem 4, these quantities can be used to characterize single-serving capacities of quantum channels.

**ACKNOWLEDGMENT**

This work was supported by the Swiss National Science Foundation (SNF) and Hewlett Packard Research Labs.

**REFERENCES**

[1] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography – part I: Secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, 1993.

[2] C. H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, 1988.

[3] I. Csiszar and J. Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, 1978.

[4] M. Fitzi, S. Wolf, and J. Wullschleger. Pseudo-signatures, broadcast, and multi-party computation from correlated randomness. In *Advances in Cryptology—CRYPTO ’04*, Springer-Verlag, 2004.

[5] P. Gács and J. Körner. Common information is far less than mutual information. *Probl. Contr. Inform. Theory*, 2:149–162, 1973.

[6] T. S. Han, S. Verdú. Approximation Theory of Output Statistics. *IEEE Trans. on Information Theory*, vol. IT-39, no. 3, pp. 752–772, 1993.

[7] P. Hayden and A. Winter. On the communication cost of entanglement transformations. *Phys. Rev. A*, 67:012326, 2003.

[8] J. Hastad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[9] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC ’89)*, pages 12–24. ACM Press, 1989.

[10] U. Maurer. Secret key agreement by public discussion. *IEEE Transaction on Information Theory*, 39(3):733–742, May 1993.

[11] R. Renner. Security of QKD. PhD thesis, ETH, 2005. [quant-ph/0512258](quant-ph/0512258)

[12] R. Renner and S. Wolf. Smooth Rényi entropy and applications. In *Proceedings of 2004 IEEE International Symposium on Information Theory*, page 233. IEEE, June 2004.

[13] R. Renner and S. Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *Advances in Cryptology—ASIACRYPT 2005*, volume 3788, pages 199–216. Springer-Verlag, December 2005.

[14] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. Journal*, 27:379–423, 623–656, 1948.

[15] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[16] S. Wolf and J. Wullschleger. Zero-error information and applications in cryptography. In *Proceedings of 2004 IEEE Information Theory Workshop (ITW 2004)*, 2004.

[17] A. D. Wyner. The wiretap channel. *Bell System Tech. Journal*, 54:1355–1387, 1975.