Information-Theoretic Key Agreement: From Weak to Strong Secrecy for Free

Ueli Maurer* Stefan Wolf*

Abstract. One of the basic problems in cryptography is the generation of a common secret key between two parties, e.g., in order to communicate privately. In this paper we consider information-theoretically secure key agreement. Wyner and subsequently Csiszár and Körner described and analyzed settings for secret-key agreement based on noisy communication channels. Maurer as well as Ahlswede and Csiszár generalized these models to a scenario based on correlated randomness and public discussion. In all these settings, the secrecy capacity and the secret-key rate, respectively, have been defined as the maximal achievable rates at which a highly-secret key can be generated by the legitimate partners. However, the privacy requirements were too weak in all these definitions, requiring only the adversary's ratio of information to be negligible, but hence tolerating her to obtain a possibly substantial amount of information about the resulting key. It has been unknown previously how to generate keys about which the adversary has virtually no information. We give natural new definitions of secrecy capacity and secret-key rate satisfying this stronger requirement and show that not only secret-key agreement is possible with respect to the strong secrecy condition, but even that the achievable key-generation rates are equal to the previous weak notions of secrecy capacity and secret-key rate. Hence the unsatisfactory old definitions can be completely replaced by the new ones. The proofs require novel privacy-amplification techniques based on extractor functions.

Keywords. Unconditional security, key agreement, wire-tap channel, secrecy capacity, universal hashing, extractors, privacy amplification, typical sequences.

1 Introduction and Preliminaries

1.1 Models of Information-Theoretic Secret-Key Agreement

This paper is concerned with *unconditional security* in cryptography. Unlike *computationally-secure cryptosystems*, the security of which is based on the assumed yet unproven hardness of a certain problem such as integer factoring, a proof without any computational assumption, based on information theory rather than complexity theory, can be given for the security of an unconditionally-secure system.

* Computer Science Department, Swiss Federal Institute of Technology (ETH Zürich), CH-8092 Zürich, Switzerland. E-mail addresses: {maurer,wolf}@inf.ethz.ch.

A fundamental problem is the generation of a mutual key about which an adversary has virtually no information. Wyner [18] and later Csiszár and Körner [9] considered the natural message-transmission scenarios in which the legitimate partners Alice and Bob, as well as the adversary Eve, are connected by noisy channels. In Csiszár and Körner's setting, Alice sends information (given by the random variable X) to Bob (receiving Y) (and to the opponent Eve who obtains Z) over a noisy broadcast channel characterized by the conditional distribution $P_{YZ|X}$. Wyner's model corresponds to the special case where $X \rightarrow Y \rightarrow Z$ is a Markov chain.

The *secrecy capacity* $C_S(P_{YZ|X})$ of the channel $P_{YZ|X}$ has been defined as the maximal rate at which Alice can transmit a secret string to Bob by using only the given noisy (one-way) broadcast channel such that the rate at which the eavesdropper receives information about the string can be made arbitrarily small. More precisely, the secrecy capacity is the maximal asymptotically-achievable ratio between the number of generated key bits and the number of applications of the noisy broadcast channel.

As a natural generalization of these settings, Maurer [12] and subsequently Ahlswede and Csiszár [1] have considered the model of secret-key agreement by public discussion from correlated randomness. Here, two parties Alice and Bob, having access to specific dependent information, use authentic public communication to agree on a secret key about which an adversary, who also knows some related side information, obtains only a small fraction of the total information. More precisely, it is assumed in this model that Alice and Bob and the adversary Eve have access to repeated independent realizations of random variables X, Y, and Z, respectively. A special example is the situation where all the parties receive noisy versions of the outcomes of some random source, e.g., random bits broadcast by a satellite at low signal power.

The *secret-key rate* $S(X;Y||Z)$ has, in analogy to the secrecy capacity, been defined in [12] as the maximal rate at which Alice and Bob can generate a secret key by communication over the noiseless and authentic but otherwise insecure channel in such a way that the opponent obtains information about this key only at an arbitrarily small rate.

Note that Maurer's model is a generalization of the earlier settings in the sense that only the correlated information, but not the insecure communication is regarded as a resource. In particular, the communication can be interactive instead of only one-way, and the required *amount* of communication has no influence on the resulting secret-key rate. These apparently innocent modifications have dramatic consequences for the possibility of secret-key agreement.

1.2 The Secrecy Capacity and the Secret-Key Rate

The precise definitions of $C_S(P_{YZ|X})$ and of $S(X;Y||Z)$ will be given later, but we discuss here some of the most important bounds on these quantities. Roughly speaking, the possibility of secret-key agreement in the broadcast-channel model is restricted to situations for which Alice and Bob have an initial advantage in terms of $P_{YZ|X}$, whereas interactive secret-key generation can be possible in settings that are initially much less favorable for the legitimate partners.

In [9] it was shown that $C_S(P_{YZ|X}) \geq \max_{P_X}(I(X;Y) - I(X;Z))$, where the maximum is taken over all possible distributions P_X on the range $\mathcal{X}$ of X, and that equality holds whenever $I(X;Y) - I(X;Z)$ is non-negative for all distributions P_X. On the other hand, it is clear from the above bound that if $U \to X \to YZ$ is a Markov chain, then $C_S(P_{YZ|X}) \geq I(U;Y) - I(U;Z)$ is also true. If the maximization is extended this way, then equality always holds:

$$C_S(P_{YZ|X}) = \max_{P_{UX}\,:\,U \to X \to YZ} (I(U;Y) - I(U;Z)) \tag{1}$$

is one of the main results of [9]. It is a consequence of equality (1) that Alice and Bob can generate a secret key by noisy one-way communication exactly in scenarios that provide an advantage of the legitimate partners over the opponent in terms of the broadcast channel's conditional distribution $P_{YZ|X}$.

The secret-key rate $S(X;Y||Z)$, as a function of P_{XYZ}, has even been studied more intensively. Lower and upper bounds on this quantity were derived, as well as necessary and sufficient criteria for the possibility of secret-key agreement [12], [14]. The lower bound

$$S(X;Y||Z) \geq \max\left[I(X;Y) - I(X;Z)\,,\, I(Y;X) - I(Y;Z)\right] \tag{2}$$

follows from equality (1) [12]. It should be pointed out that secret-key agreement can also be possible when the right-hand side of inequality (2) is zero or negative. However, a special protocol phase, called *advantage distillation*, requiring feedback instead of only one-way communication, must be used.

On the other hand however, it was shown in [14] that

$$S(X;Y||Z) \leq I(X;Y{\downarrow}Z) := \min_{P_{\overline{Z}|Z}} \left[I(X;Y|\overline{Z})\right] ,$$

where $I(X;Y{\downarrow}Z)$ is called the *intrinsic conditional information between X and Y, given Z*. It has been conjectured in [14], and is supported by some evidence, that

$$S(X;Y||Z) = I(X;Y{\downarrow}Z)$$

holds for all P_{XYZ}, or at least that $S(X;Y||Z) > 0$ holds whenever $I(X;Y{\downarrow}Z) > 0$. However, these statements have been proven only for special cases.

1.3 Contributions of this Paper

In all the mentioned scenarios, the conditions on the resulting secret key were too weak. As it is often done in information theory, all the involved quantities, including the information about the key the adversary is tolerated to obtain, were measured in terms of an *information rate*, which is defined as the ratio between the information quantity of interest and the number of independent repetitions of the underlying random experiment. Unfortunately, the total information the adversary gains about the resulting secret key is then, although arbitrarily small in terms of the rate, not necessarily bounded, let alone negligibly small, because for a given (small) ratio $\varepsilon > 0$, key agreement with respect to the security parameter ε is required to work only for strings of length N exceeding some bound $N_0(\varepsilon)$ which can depend on ε (in particular, $N_0(\varepsilon) \cdot \varepsilon \to \infty$ for $\varepsilon \to 0$ is possible). Clearly, this is typically unacceptable in a cryptographic scenario. For instance, the generated key cannot be used for a one-time-pad encryption because the entire message must be protected. However, it has been previously unknown how a message can be sent entirely secretly in the described key-agreement settings.

Motivated by these considerations, stronger definitions of the rates at which a secret key can be generated are given for the different scenarios. More specifically, it is required that the information the adversary obtains about the entire key be negligibly small in an *absolute* sense, not only in terms of a rate. In the setting of secret-key agreement by noiseless public discussion from common information it is additionally required that the resulting secret key, which must be equal for Alice and Bob with overwhelming probability, is (perfectly-) uniformly distributed.

The main results of this paper are Theorems 1 and 2, stating the somewhat surprising facts that both for the secrecy capacity and for the secret-key rate, strengthening the security requirements does not at all reduce the achievable key-generation rates. This is particularly interesting for the case of the secrecy capacity because in this model, all the communication must be carried out over the noisy channel. Recent advances in the theory of extractors are necessary for closing the gap between weak and strong security in this case.

An important conclusion of our results is that all the previously-known results on $C_S(P_{YZ|X})$ and on $S(X;Y||Z)$, briefly described in Section 1.2, immediately carry over to the strong notions although they were only proved for the weaker definitions (which is an apparently much easier task). All the previous definitions were hence unnecessarily weak and can from now on be entirely replaced by the new notions.

A basic technique used for proving the main results is privacy amplification, introduced in [3], where we use both universal hashing and, as a new method in this context, extractors. A particular problem to be dealt with is to switch between (conditional) Shannon-, Rényi-, and min-entropy of random variables or, more precisely, of blocks of independent repetitions of random variables, and the corresponding probability distributions. A powerful tool for doing this are typical-sequences techniques.

Both Theorems 1 and 2 solve open problems stated in the literature [13], [10]. In the public-discussion setting, a slightly weaker variant of the notion of strong secret-key agreement was already defined by Maurer in [13], where a lower bound on a strengthened secret-key rate was derived. (We give a much simpler proof of the same bound below.) It was not proved however that the rates are equal in any case.

1.4 Shannon-, Rényi-, and Min-Entropy, and Variational Distance

We recall the definitions of some entropy measures needed in this paper. For a good introduction to information theory, we refer to [8]. Let R be a discrete random variable with range $\mathcal{R}$. Then the *(Shannon) entropy* $H(R)$ is defined as[2] $H(R) := -\sum_{r\in\mathcal{R}} P_R(r) \cdot \log(P_R(r))$. The *Rényi entropy* $H_2(R)$ is defined as $H_2(R) := -\log(\sum_{r\in\mathcal{R}} P_R^2(r))$. Finally, the *min-entropy* $H_\infty(R)$ is $H_\infty(R) := -\log \max_{r\in\mathcal{R}}(P_R(r))$. For two probability distributions P_X and P_Y on a set $\mathcal{X}$, the *variational distance* between P_X and P_Y is defined as $d(P_X, P_Y) := \frac{1}{2}\sum_{x\in\mathcal{X}} |P_X(x) - P_Y(x)|$.

2 Secret-Key Agreement from Correlated Randomness

In this section we define a stronger variant of the secret-key rate of a distribution P_{XYZ} and show that this new quantity is equal to the previous, weak secret-key rate as defined in [12]. The protocol for strong key agreement consists of the following steps. First, weak key agreement is repeated many times. Then, so-called information reconciliation (error correction) and privacy amplification are carried out. These steps are described in Section 2.2. Of central importance for all the arguments made are typical-sequences techniques (Section 2.3). The main result of this section, the equality of the secret-key rates, is then proven in Section 2.4.

2.1 Definition of Weak and Strong Secret-Key Rates

Definition 1 [12] The *(weak) secret-key rate of X and Y with respect to Z*, denoted by $S(X;Y||Z)$, is the maximal $R \geq 0$ such that for every $\varepsilon > 0$ and for all sufficiently large $N \geq N_0(\varepsilon)$ there exists a protocol, using public communication over an insecure but authenticated channel, such that Alice and Bob, who receive $X^N = [X_1, \ldots, X_N]$ and $Y^N = [Y_1, \ldots, Y_N]$, can compute keys S and S', respectively, with the following properties. First, they are equal with probability at least $1 - \varepsilon$, and second,

$$\frac{1}{N} I(S; CZ^N) \leq \varepsilon \qquad \text{and} \qquad \frac{1}{N} H(S) \geq R - \varepsilon$$

hold. Here, C denotes the collection of messages sent over the insecure channel by Alice and Bob, and Z^N stands for $[Z_1, \ldots, Z_N]$.

As pointed out in Section 1.3, the given definition of the secret-key rate is unsatisfactorily (and, as shown later, unnecessarily) weak. We give a strong definition which limits the information leaked to the adversary in an absolute sense (and additionally requires that the resulting key be perfectly-uniformly distributed).

[2] All the logarithms in this paper are to the base 2, unless otherwise stated.

Definition 2 The *strong secret-key rate of X and Y with respect to Z*, denoted by $\overline{S}(X;Y||Z)$, is defined in the same way as $S(X;Y||Z)$ with the modifications that Alice and Bob compute strings S_A and S_B which are with probability at least $1-\varepsilon$ both equal to a string S with the properties

$$I(S;CZ^N) \ \leq \ \varepsilon \qquad \text{and} \qquad H(S) = \log|\mathcal{S}| \geq N \cdot (R-\varepsilon) \ .$$

Obviously, $\overline{S}(X;Y||Z) \leq S(X;Y||Z)$ holds. It is the goal of this section to show that equality holds for every distribution P_{XYZ}, i.e., that the attention can be totally restricted to the strong notion of secret-key rate.

2.2 Information Reconciliation and Privacy Amplification

In this section we analyze the two steps, called *information reconciliation* and *privacy amplification*, of a protocol allowing *strong* secret-key agreement whenever $I(X;Y) - I(X;Z) > 0$ or $I(Y;X) - I(Y;Z) > 0$ holds. More precisely, we show

$$\overline{S}(X;Y||Z) \geq \max\{\, I(X;Y) - I(X;Z)\,,\, I(Y;X) - I(Y;Z)\,\} \ . \qquad (3)$$

Assume $I(X;Y) > I(X;Z)$. The information-reconciliation phase of interactive error correction consists of the following step. For some suitable function $h : \mathcal{X}^N \to \{0,1\}^L$, Alice sends $h(X^N)$ to Bob for providing him (who knows Y^N) with a sufficient amount of information about X^N that allows him to reconstruct X^N with high probability. The existence of such a function (in a fixed universal class, see Definition 3) for L on the order of $N \cdot H(X|Y)$ is stated in Lemma 1, a weaker variant of which was formulated already in [13] without a proof. We give a proof in Section 2.3 as a first application of typical-sequences arguments. Note that this type of (one-way) information-reconciliation protocol is optimal with respect to the amount of exchanged information and efficient with respect to communication complexity, but not with respect to computational efficiency of Bob. There exist efficient interactive methods which however leak more information to the adversary (see [4] for various results on information reconciliation).

Definition 3 [7] A class G of functions $g : \mathcal{A} \longrightarrow \mathcal{B}$ is *universal* if, for any distinct x_1 and x_2 in $\mathcal{A}$, the probability that $g(x_1) = g(x_2)$ holds is at most $1/|\mathcal{B}|$ when g is chosen at random from G according to the uniform distribution.

The following example of a universal class is taken from [7].

Example 1. Let $1 \leq M \leq N$, let a be an element of $GF(2^N)$, and interpret $x \in \{0,1\}^N$ as an element of $GF(2^N)$ with respect to a fixed basis of the extension field over the prime field $GF(2)$. Consider the function $h_a : \{0,1\}^N \to \{0,1\}^M$ assigning to an argument x the first M bits (with respect to this basis representation) of the element ax of $GF(2^N)$, i.e., $h_a(x) := \mathrm{LSB}_M(a \cdot x)$. The class $\{h_a : a \in GF(2^N)\}$ is a universal class of functions mapping $\{0,1\}^N$ to $\{0,1\}^M$ with 2^N elements.

Lemma 1 *Let X and Y be random variables, and let $[(X_1, Y_1), \ldots, (X_N, Y_N)]$ be a block of N independent realizations of X and Y. Then for every $\varepsilon > 0$ and $\varepsilon' > 0$, for sufficiently large N, for every L satisfying $L/N > (1+\varepsilon)H(X|Y)$, and for every universal class $\mathcal{H}$ of functions mapping $\mathcal{X}^N$ to $\{0,1\}^L$, there exists a function h in $\mathcal{H}$ such that $[X_1, \ldots, X_N]$ can be decoded from $[Y_1, \ldots, Y_N]$ and $h(X^N)$ with error probability at most ε'.*

In the second protocol phase, privacy amplification, Alice and Bob compress the mutual but generally highly-insecure string X^N to a shorter string S with virtually-uniform distribution and about which Eve has essentially no information. (Note that Eve's total information about X^N consists of Z^N and $h(X^N)$ at this point.) Bennett *et. al.* [2] have shown that universal hashing allows for distilling a virtually-secure string whose length is roughly equal to the Rényi entropy of the original string in Eve's view.

Lemma 2 [2] *Let W be a random variable with range $\mathcal{W}$, and let G be the random variable corresponding to the random choice, according to the uniform distribution, of a function out of a universal class of functions mapping $\mathcal{W}$ to $\{0,1\}^M$. Then $H(G(W)|G) \geq H_2(G(W)|G) \geq M - 2^{M-H_2(W)}/\ln 2$.*

Lemma 2 states that if Alice and Bob share a particular string S and Eve's information about S corresponds to the distribution $P_{S|V=v}$ (where v denotes the particular value of her information V) about which Alice and Bob know nothing except a lower bound t on the Rényi entropy, i.e., $H_2(S|V=v) \geq t$, then Alice and Bob can generate a secret key S' of roughly t bits. More precisely, if Alice and Bob compress S to an $(t-s)$-bit key for some security parameter $s > 0$, then Eve's total information about this key is exponentially small in s (see Figure 1).

A natural problem that arises when combining information reconciliation and privacy amplification with universal hashing is to determine the effect of the error-correction information (leaked also to the adversary) on the Rényi entropy of the partially-secret string, given Eve's information. The following result, which was shown by Cachin [5], as an improvement of an earlier result by Cachin and Maurer [6], states that leaking t physical bits of arbitrary side information about a string cannot reduce its Rényi entropy by substantially more than t, except with exponentially small probability.

Lemma 3 [5] *Let X and Q be random variables, and let $s > 0$. Then with probability at least $1 - 2^{-(s/2-1)}$ (taken over $q \in \mathcal{Q}$), we have $H_2(X) - H_2(X|Q = q) \leq \log|\mathcal{Q}| + s$.*

2.3 Typical Sequences

In the following, we will make use of so-called typical-sequences arguments. Such arguments are based on the fact that if a large number of independent realizations of a random variable U is considered, then the actual probability of the particular outcome sequence is, with overwhelming probability, close to a certain "typical probability." There exist various definitions of typical sequences. The definition given below corresponds to a weak notion of typicality, dealing only with probabilities and not with the number of occurrences of the outcome symbols of the original random variable U in the sequence.

Definition 4 Let U be a random variable with probability distribution P_U and range $\mathcal{U}$. For $N \geq 0$, let $P_{U^N} = P_U^N$ be the distribution of the random variable U^N corresponding to N independent realizations of U. Then a sequence $u = (u_1, u_2, \ldots, u_N) \in \mathcal{U}^N$ is called (weakly) δ-typical if $2^{-N(H(U)+\delta)} \leq P_{U^N}(u) \leq 2^{-N(H(U)-\delta)}$.

Lemma 4 states that if N is large enough, then U^N is δ-typical with high probability. More precisely, the probability of the "non-typicality" event tends to zero faster than $1/N^2$. The lemma follows immediately from Theorem 12.69 in [8].

Lemma 4 [8] *For all $\delta, \varepsilon > 0$, we have $N \cdot (\text{Prob}\,[U^N$ is* not *δ-typical$])^{1/2} < \varepsilon$ for sufficiently large N.*

As a first application of typical sequences and of Lemma 4, we can now give a proof of Lemma 1 from the previous section.

Proof of Lemma 1. It is sufficient to show that the statement is true if the function h is chosen at random from the universal class $\mathcal{H}$ of functions mapping $\mathcal{X}^N$ to $\{0,1\}^L$. Then we can conclude from

$$\text{Prob}_{\,h \in_r \mathcal{H},\, [x_1,\ldots,x_N] \in \mathcal{X}^N}[\mathcal{E}] = \text{E}_H\left[\text{Prob}_{\,[x_1,\ldots,x_N] \in \mathcal{X}^N | H=h}[\mathcal{E}]\right] \leq \varepsilon'$$

(where $\mathcal{E}$ stands for the event of a decoding error when using the optimal strategy) that there exists a *specific* function $h_0 \in \mathcal{H}$ such that the probability of a decoding error is at most ε', given $H = h_0$.

We can assume $H(X|Y) > 0$ because the statement is trivial otherwise. Let $0 < \alpha < \varepsilon'/2$ and $0 < \delta < \varepsilon H(X|Y)/2$ be constants. For sufficiently large N, we have with probability at least $1 - 2\alpha$ that the outcomes x^N and y^N of the random variables X^N and Y^N satisfy both

$$2^{-N(H(XY)+\delta)} \leq P_{X^N Y^N}(x^N, y^N) \leq 2^{-N(H(XY)-\delta)} \tag{4}$$

and

$$2^{-N(H(Y)+\delta)} \leq P_{Y^N}(y^N) \leq 2^{-N(H(Y)-\delta)} \tag{5}$$

hold. From (4) and (5) one can conclude

$$2^{-N(H(X|Y)+2\delta)} \leq P_{X^N|Y^N}(x^N, y^N) \leq 2^{-N(H(X|Y)-2\delta)} \ . \tag{6}$$

For a particular fixed y^N, there are at most $2^{N(H(X|Y)+2\delta)}$ different x^N for which (6) can hold. Let $L > (1+\varepsilon)NH(X|Y)$. The probability that for a randomly-chosen $h \in \mathcal{H}$ and for given y^N and $h(x^N)$, there exists $(x')^N$, different from x^N, satisfying (6) and with $h(x^N) = h((x')^N)$, is at most

$$1 - (1 - 2^{-L})^{2^{N(H(X|Y)+2\delta)}} \leq 2^{-L} \cdot 2^{N(H(X|Y)+2\delta)} < 2^{-N(\varepsilon H(X|Y)-2\delta)} ,$$

which is arbitrarily small for sufficiently large N because $\delta < \varepsilon H(X|Y)/2$. From the above we conclude that the decoding-error probability is upper bounded by $\text{Prob}\,[\mathcal{E}] \leq 2\alpha + 2^{-N(\varepsilon H(X|Y)-2\delta)}$. This expression is smaller than ε' for sufficiently large N by the definition of α and δ. □

As another application of the typical-sequences technique, and as a further step towards proving equality of the secret-key rates with respect to the weak and strong definitions, we show that the weak definition can be extended by an additional condition requiring that the resulting key is close-to-uniformly distributed. More precisely, Lemma 5 states that the condition

$$\frac{1}{N}H(S) \geq \frac{1}{N}\log|\mathcal{S}| - \varepsilon \tag{7}$$

can be included into the definition of $S(X;Y||Z)$ without effect on its value. (Note that the condition (7) is much weaker than the uniformity condition in the definition of $\overline{S}(X;Y||Z)$.)

Lemma 5 *Let the* uniform *(weak) secret-key rate $S_u(X;Y||Z)$ be defined similarly to $S(X;Y||Z)$, but with the additional condition (7). Then $S_u(X;Y||Z) = S(X;Y||Z)$ holds.*

Proof. The idea is to carry out the key-generation procedure independently many times and to apply data compression. More precisely, secret-key agreement with respect to the definition of $S(X;Y||Z)$ is repeated M times. Clearly, we can assume that the resulting triples $[S_i, S_i', (Z^N C)_i]$ are independent for different values of i and can be considered as the random variables in a new random experiment. When repeating this experiment for a sufficient number of times and applying data compression to the resulting sequence of keys, thereby using that with high probability both $[S_1, S_2, \ldots]$ and $[S_1', S_2', \ldots]$ are typical sequences, one finally obtains key agreement that ends up in a highly-uniformly-distributed key.

Let $R := S(X;Y||Z)$. We show that for any $\varepsilon > 0$ (and for a sufficiently large number of realizations of the random variables) secret-key agreement at a rate at least $R - \varepsilon$ is possible even with respect to the stronger definition which includes the uniformity condition (7).

For parameters $\varepsilon' > 0$ and $N > 0$, both to be determined later, let secret-key agreement (not necessarily satisfying the new condition) be carried out M times independently. Let S_i and S_i', $i = 1, \ldots, M$, be the generated keys, and let C_i and $(Z^N)_i$ be the corresponding collection of messages sent over the public channel and the realizations of Z that Eve obtains, respectively. Then the triples $[S_i, S_i', (Z^N C)_i]$, $i = 1, \ldots, M$, are statistically independent and identically distributed. According to the definition of $S(X;Y||Z)$, we can achieve for every i

$$H(S_i)/N \geq R - \varepsilon' \ , \quad \text{Prob}\,[S_i \neq S_i'] < \tilde{\varepsilon} \ , \quad \text{and} \quad I(S_i;(Z^N C)_i)/N < \varepsilon' \ , \qquad (8)$$

where the constant $\tilde{\varepsilon}$ will be specified later. (Note that in order to make only $\tilde{\varepsilon}$ smaller and to leave ε' unchanged, it is not necessary to increase N because the second condition in (8) is stricter for larger N: The key can be subdivided into smaller pieces at the end, and for every such piece, the error probability is at most $\tilde{\varepsilon}$.)

Using the fact that for all $\alpha > 0$ and $\delta > 0$, the event $\mathcal{E}(\delta)$ that the sequence $[S_1, S_2, \ldots, S_M]$ is δ-typical has probability at least $1 - \alpha$ for sufficiently large M, we can transform the key vector $[S_1, \ldots, S_M]$ into an almost-uniformly-distributed key T as follows. If $\mathcal{E}(\delta)$ occurs, then let $T := [S_1, \ldots, S_M]$, otherwise $T := \Delta$ for some failure symbol Δ. The key T' is computed from $[S_1', \ldots, S_M']$ analogously. Then, T and T' have the following properties. First,

$$\log |\mathcal{T}| \leq M(H(S) + \delta) + 1 \qquad \text{and} \qquad H(T) \geq (1 - \alpha)M(H(S) - \delta)$$

follow from the definitions of T and of δ-typical sequences. For the quantities occurring in the definition of $S_u(X;Y||Z)$, we hence obtain

$$\begin{aligned} H(T)/MN &\geq (1 - \alpha)(R - \varepsilon' - \delta/N) \ , & (9) \\ \text{Prob}\,[T \neq T'] &< M\tilde{\varepsilon} \ , & (10) \\ I(T;(Z^N C)_{i=1,\ldots,M})/MN &< \varepsilon' \ , & (11) \\ (\log |\mathcal{T}| - H(T))/MN &\leq \alpha R + 2\delta/N \ . & (12) \end{aligned}$$

Because of Lemma 4 one can choose, for every sufficiently large N, constants α, δ, and ε' such that $\text{Prob}\,[\overline{\mathcal{E}(\delta)}] < \alpha$ (where $\overline{\mathcal{E}(\delta)}$ stands for the complementary event of $\mathcal{E}(\delta)$) for this choice of M, and such that the expressions on the right-hand sides of (11) and (12) are smaller than ε, whereas the right-hand side of (9) is greater than $R - \varepsilon$. Finally, $\tilde{\varepsilon}$ can be chosen as ε/M, such that the condition (10) is also satisfied.

We conclude that the uniform secret-key rate $S_u(X;Y||Z)$ is at least $R = S(X;Y||Z)$. This concludes the proof. □

Lemma 6 links Rényi entropy with typicality of sequences (and hence Shannon entropy). More precisely, the conditional Rényi entropy of a sequence of realizations of random variables is close to the length of the sequence times the conditional Shannon entropy of the original random variables, given a certain typicality event which occurs with high probability. Related arguments already appeared in [11] and [5].

Lemma 6 *Let P_{XZ} be the joint distribution of two random variables X and Z, let $0 < \delta \leq 1/2$, and let N be an integer. The event $\mathcal{F}(\delta)$ is defined as follows: First, the sequences x^N and $(x,z)^N$ must each be δ-typical, and second,*

z^N must be such that the probability, taken over $(x')^N$ according to the distribution $P_{X^N|Z^N=z^N}$, that $(x',z)^N$ is δ-typical is at least $1-\delta$. Then we have $N \cdot \text{Prob}\,[\overline{\mathcal{F}(\delta)}] \to 0$ for $N \to \infty$, and $H_2(X^N|Z^N = z^N, \mathcal{F}(\delta)) \geq N(H(X|Z) - 2\delta) + \log(1-\delta)$.

Proof. Because of Lemma 4, the event, denoted by $\mathcal{E}(\delta)$, that both x^N and $(x,z)^N$ are δ-typical has probability at least $1-\delta^2$ for some $N = N(\delta)$ with $N(\delta)\cdot\delta \to 0$. For this value of N, z^N has with probability at least $1 - \sqrt{\delta^2} = 1-\delta$ the property that $(x',z)^N$ is δ-typical with probability at least $1-\sqrt{\delta^2} = 1-\delta$, taken over $(x')^N$ distributed according to $P_{X^N|Z^N=z^N}$. Hence the probability of the complementary event $\overline{\mathcal{F}(\delta)}$ of $\mathcal{F}(\delta)$ is at most $\delta^2+\delta$, thus $N\cdot\text{Prob}\,[\overline{\mathcal{F}(\delta)}] \to 0$.

On the other hand, given that z^N and $(x',z)^N$ are δ-typical, we can conclude as in the proof of Lemma 1 that

$$2^{-N(H(X|Z)+2\delta)} \leq P_{X^N|Z^N}((x')^N, z^N) \leq 2^{-N(H(X|Z)-2\delta)}$$

holds. For a fixed value z^N, the Rényi entropy of X^N, given the events $Z^N = z^N$ and $\mathcal{F}(\delta)$, is lower bounded by the Rényi entropy of a uniform distribution over a set with $(1-\delta)\cdot 2^{N(H(X|Z)-2\delta)}$ elements: $H_2(X^N|Z^N = z^N, \mathcal{F}(\delta)) \geq N(H(X|Z) - 2\delta) + \log(1-\delta)$. □

2.4 Equality of Weakly- and Strongly-Defined Rates

In this section we prove the lower bound (3) on $\overline{S}(X;Y||Z)$ and the first main result, stating that the weak and strong secret-key rates are equal for any distribution. A result closely related to Lemma 7 was proved as the main result in [13]. We give a much shorter and simpler proof based on the results in Sections 2.2 and 2.3.

Lemma 7 *For all P_{XYZ}, $\overline{S}(X;Y||Z) \geq \max\{\, I(X;Y) - I(X;Z)\,, I(Y;X) - I(Y;Z)\,\}$ holds.*

Proof. We only prove that

$$I(X;Y) - I(X;Z) = H(X|Z) - H(X|Y) \tag{13}$$

is an achievable rate. The statement then follows by symmetry.

Let $\varepsilon > 0$, and let $\Delta > 0$ be determined later. We show that for the parameter ε, and for sufficiently large N, there exists a protocol which achieves the rate (13). Let $\delta < \varepsilon/4$ and $\alpha < \Delta/(2H(X))$ be constants, and let $\mathcal{F}(\delta)$ be the event defined in Lemma 6. Because of Lemma 6 we have for sufficiently large N that $N\cdot\text{Prob}\,[\overline{\mathcal{F}(\delta)}] < \alpha$. On the other hand,

$$H_2(X^N|Z^N = z^N, \mathcal{F}(\delta)) \geq N\cdot(H(X|Z) - 2\delta) + \log(1-\delta)$$

holds.

The protocol now consists of two messages sent from Alice to Bob, one for information reconciliation and the other one for privacy amplification (see Section 2.2). Let $\beta < \varepsilon/(2H(X|Y))$ be a positive constant. According to Lemma 1 there exists for sufficiently large N a function $h : \mathcal{X}^N \rightarrow \{0,1\}^L$, where $L := \lceil (1+\beta)NH(X|Y) \rceil$, such that X^N can be determined from Y^N and $h(X^N)$ with probability at least $1-\varepsilon/2$ (using the optimal strategy). Clearly, the value $h(X^N)$ reduces Eve's uncertainty in terms of Rényi entropy about X^N. We conclude from Lemma 3 for $s := 2\log(2NH(X)/\Delta) + 2$ that with probability at least $1 - 2^{-(s/2-1)}$,

$$\begin{aligned} & H_2(X^N|Z^N = z^N, h(X^N) = h(x^N), \mathcal{F}(\delta)) \\ & \qquad \geq N \cdot (H(X|Z) - 2\delta) + \log(1-\delta) - [(1+\beta) \cdot N \cdot H(X|Y) + 1 + s] \quad (14) \\ & \qquad = N \cdot (H(X|Z) - H(X|Y)) - 2\delta N - \beta N H(X|Y) - 1 - s + \log(1-\delta) \\ & \qquad =: \ Q \, . \end{aligned}$$

Finally, Alice and Bob use privacy amplification to transform their mutual information X^N into a highly-secret string $\tilde{S}$. Let $r := \lceil \log N \rceil$, and let $M := Q - r$ be the length of the resulting string $\tilde{S}$. If G is the random variable corresponding to the random choice of a universal hash function mapping $\mathcal{X}^N \rightarrow \{0,1\}^M$, and if $\tilde{S} := G(X^N)$, then we have $H(\tilde{S}|Z^N = z^N, h(X^N) = h(x^N), G, \mathcal{F}(\delta)) \geq M - 2^{-r}/\ln 2$ under the condition that inequality (14) holds. Hence we get for sufficiently large N

$$\begin{aligned} H(\tilde{S}|Z^N, h(X^N), G) & \geq (\mathrm{Prob}\,[\mathcal{F}(\delta)] - 2^{-(s/2-1)})(M - 2^{-r}/\ln 2) \\ & \geq M - 2^{-r}/\ln 2 - (\mathrm{Prob}\,[\overline{\mathcal{F}(\delta)}] + 2^{-(s/2-1)}) \cdot N \cdot H(X) \\ & > \log|\tilde{\mathcal{S}}| - \Delta \end{aligned}$$

by definition of r, α, and s. Let now S be a "uniformization" of $\tilde{S}$, i.e., a random variable S with range $\mathcal{S} = \tilde{\mathcal{S}} = \{0,1\}^M$ that can be generated by sending $\tilde{S}$ over some channel characterized by $P_{S|\tilde{S}}$, that is uniformly distributed and minimizes $\mathrm{Prob}\,[S \neq \tilde{S}]$ among all random variables with these properties. For $C = [h(X^N), G]$ and sufficiently small Δ, we can then conclude that

$$I(S; Z^N C) < \varepsilon \ , \quad H(S) = \log|\mathcal{S}| \ , \quad \text{and} \quad \mathrm{Prob}\,[S' \neq S] < \varepsilon$$

because of $H(\tilde{S}) \geq H(\tilde{S}|Z^N, h(X^N), G)$. The achievable key-generation rate with this protocol is hence at least

$$H(X|Z) - H(X|Y) - 2\delta - \beta H(X|Y) \geq I(X;Y) - I(X;Z) - \varepsilon \ .$$

Thus we obtain $\overline{S}(X;Y||Z) \geq I(X;Y) - I(X;Z)$, which concludes the proof. □

Theorem 1 is the main result of this section and states that the strong secret-key rate $\overline{S}(X;Y||Z)$ is always equal to the weak secret-key rate $S(X;Y||Z)$.

Theorem 1 *For all distributions P_{XYZ}, we have $\overline{S}(X;Y||Z) = S(X;Y||Z)$.*

Proof. Clearly, $\overline{S}(X;Y||Z) \leq S(X;Y||Z)$ holds. Let $R := S(X;Y||Z)$, and let $\varepsilon > 0$. According to the definition of the secret-key rate $S(X;Y||Z)$ (and because of Lemma 5), there exists, for sufficiently large N, a protocol with the following properties: Alice and Bob know, at the end of the protocol, strings S and S' such that $H(S) \geq NR - N\varepsilon$, $\text{Prob}\,[S \neq S'] < \varepsilon$, $I(S;Z^N C) \leq N\varepsilon$, and $H(S) \geq \log|\mathcal{S}| - N\varepsilon$ hold. From these equations, we can conclude by Fano's inequality [8] that

$$\begin{aligned} I(S;S') = H(S) - H(S|S') &\geq H(S) - h(\text{Prob}\,[S \neq S']) - \text{Prob}\,[S \neq S'](H(S) + N\varepsilon) \\ &> H(S)(1-\varepsilon) - h(\varepsilon) - N\varepsilon^2 \geq NR - NR\varepsilon - N\varepsilon - h(\varepsilon) \end{aligned}$$

(where h stands for the binary entropy function), hence $I(S;S') - I(S;Z^N C) \geq NR - NR\varepsilon - 2N\varepsilon - h(\varepsilon)$. Let us now consider the random experiment $[S, S', Z^N C]$ (where we assume that the realizations are independent). By applying Lemma 7 to the new distribution, we get

$$\overline{S}(X;Y||Z) \geq S(S;S'||Z^N C)/N \geq (I(S;S') - I(S;Z^N C))/N \geq R - R\varepsilon - 2\varepsilon - h(\varepsilon)/N$$

for every $\varepsilon > 0$, thus $\overline{S}(X;Y||Z) \geq S(X;Y||Z)$. □

3 Strengthening the Secrecy Capacity

This section is concerned with the model introduced by Wyner [18] and the generalization thereof by Csiszár and Körner [9], which served as a motivation for Maurer's [12] scenario treated in Section 2. In analogy to the weak definition of the secret-key rate, the original definition of the secrecy capacity is not satisfactory because the total amount of information about the resulting key that the adversary obtains can be unbounded. We show that also the definition of the secrecy capacity can be strengthened, without any effect on the actual value of this quantity, in the sense that the total amount of information the adversary obtains about the secret key is negligibly small. This solves an open problem stated in [10] and in [13].

A special difficulty that arises in this model as compared to the model of Section 2 is that no communication is "for free." More precisely, the noisy broadcast channel must be used for the entire communication (i.e., for the exchange of all the error-correction and privacy-amplification information), which at first sight appears to reduce the maximal achievable key-generation rate. However, the use of extractors (see Section 3.2) instead of universal hashing for privacy amplification allows to keep the fraction of channel uses for communicating the error-correction and privacy-amplification messages arbitrarily small.

3.1 Definition of the Secrecy Capacity $C_S(P_{YZ|X})$

Assume that the parties Alice and Bob, and the adversary Eve, are connected by a noisy broadcast channel with conditional output distribution $P_{YZ|X}$ [9]. (Wyner's wire-tap channel corresponds to the special case where $P_{YZ|X} = P_{Y|X} \cdot P_{Z|Y}$ holds.) The ability of generating mutual secret information was quantified in detail as follows.

Definition 5 **[18], [9]** Consider a memoryless broadcast channel characterized by the conditional joint distribution $P_{YZ|X}$. The *secrecy capacity* $C_S(P_{YZ|X})$ of the channel is the maximal real number $R \geq 0$ such that for every $\varepsilon > 0$, for sufficiently large N, and for $K := \lfloor (R - \varepsilon)N \rfloor$, there exists a possibly probabilistic (i.e., additionally depending on some random bits) encoding function $e : \{0,1\}^K \rightarrow \mathcal{X}^N$ together with a decoding function $d : \mathcal{Y}^N \rightarrow \{0,1\}^K$ such that if S is uniformly distributed over $\{0,1\}^K$, we have for $X^N = e(S)$ and $S' := d(Y^N)$ that $\text{Prob}\,[S' \neq S] < \varepsilon$ and

$$\frac{1}{K} H(S|Z^N) > 1 - \varepsilon \tag{15}$$

hold.

3.2 Privacy Amplification with Extractors

In order to show that the notion of secrecy used in the definition of C_S can be strengthened without reducing the secrecy capacity of the broadcast channel, we need a different technique for privacy amplification, requiring less information to be transmitted, namely only an asymptotically arbitrarily small fraction of the number of bits of the partially-secure string to be compressed. (Otherwise, the channel applications needed for sending this message would reduce the achievable key-generation rate.) We show that such a technique is given by so-called *extractors*. Roughly speaking, an extractor allows to efficiently isolate the randomness of some source into virtually-random bits, using a small additional number of perfectly-random bits as a catalyst, i.e., in such a way that these bits reappear as a part of the almost-uniform output. Extractors are of great importance in theoretical computer science, where randomness is often regarded as a resource. They have been studied intensively in the past years by many authors. For an introduction and some constructions, see for example [15], [17], or [16], and the references therein.

Recent results, described below, show that such functions allow, using only a small amount of true randomness, to distill (almost) the entire randomness, measured in terms of H_∞, of some string into an almost-uniformly-distributed string. A disadvantage of using extractors instead of universal hashing is that a string of length only roughly equal to the *min*-entropy instead of the generally greater *Rényi* entropy of the original random variable can be extracted. However, this drawback has virtually no effect in connection with typical sequences, i.e., almost-uniform distributions. (Notice that for uniform distributions, all the entropy measures are equal.)

Definition 6 A function $E : \{0,1\}^N \times \{0,1\}^d \rightarrow \{0,1\}^r$ is called a (δ', ε')-extractor if for any distribution P on $\{0,1\}^N$ with min-entropy $H_\infty(P) \geq \delta' N$, the variational distance of the distribution of $[V, E(X,V)]$ to the uniform distribution over $\{0,1\}^{d+r}$ is at most ε' when choosing X according to P and V independently according to the uniform distribution over $\{0,1\}^d$.

The following theorem was proved in [17]. It states that there exist extractors which distill virtually all the min-entropy out of a weakly-random source, thereby requiring only a small (i.e., "poly-logarithmic") number of truly-random bits. Note that Definition 6, and hence the statement of Lemma 8, is formally slightly stronger than the corresponding definition in [17] because it not only requires that the length of the extractor output is roughly equal to the min-entropy of the source plus the number of random bits, but even that these bits reappear as a part of the output. It is not difficult to see that the extractors described in [17] have this additional property.

Lemma 8 **[17]** *For every choice of the parameters N, $0 < \delta' < 1$, and $\varepsilon' > 0$, there exists a (δ', ε')-extractor $E : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^{\delta' N - 2\log(1/\varepsilon') - O(1)}$, where $d = O((\log(N/\varepsilon'))^2 \log(\delta' N))$.*

Lemma 9, which is a consequence of Lemma 8, is what we need in the proof of Theorem 2. The statement of Lemma 9 is related to Lemma 2, where universal hashing is replaced by extractors, and min-entropy must be used instead of Rényi entropy (see Figure 1).

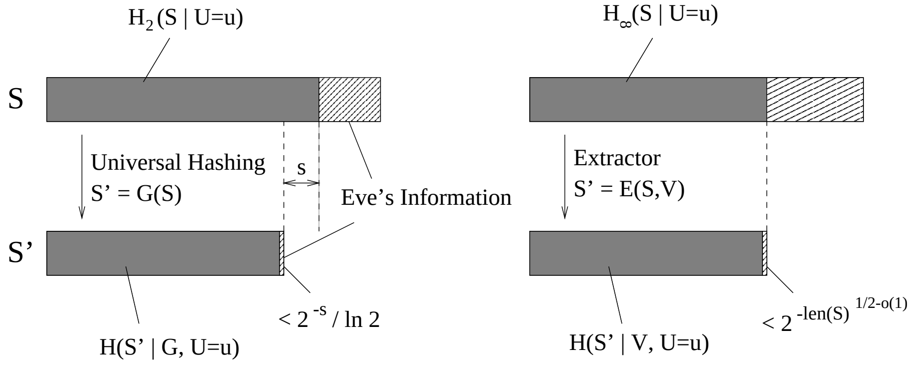

Fig. 1. Privacy Amplification: Universal Hashing Versus Extractors

Lemma 9 *Let $\delta', \Delta_1, \Delta_2 > 0$ be constants. Then there exists, for all sufficiently large N, a function $E : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^r$, where $d \leq \Delta_1 N$ and $r \geq (\delta' - \Delta_2)N$, such that for all random variables T with $\mathcal{T} \subseteq \{0,1\}^N$ and $H_\infty(T) > \delta' N$, we have*

$$H(E(T,V)|V) \geq r - 2^{-N^{1/2-o(1)}} . \qquad (16)$$

Proof. Let $\varepsilon'(N) := 2^{-\sqrt{N}/\log N}$. Then there exists N_0 such that for all $N \geq N_0$ we have a (δ', ε')-extractor E, mapping $\{0,1\}^{N+d}$ to $\{0,1\}^r$, where $d \leq \Delta_1 N$ (note that $d = O(N/\log N)$ holds for this choice of ε') and $r \geq (\delta' - \Delta_2)N$.

By definition, this means that for a uniformly-distributed d-bit string V and if $H_\infty(T) \geq \delta' N$, the distance of the distribution of $[V, E(T, V)]$ to the uniform distribution U_{d+r} over $\{0,1\}^{d+r}$ is at most $\varepsilon' = 2^{-\sqrt{N}/\log N}$. Because

$$d([V, E(T, V)], U_{d+r}) = \mathrm{E}_V[d(E(T, V), U_r)] \leq \varepsilon'$$

holds for uniformly distributed V, the distance of the distribution of $E(T, v)$ to the uniform distribution U_r (over $\{0,1\}^r$) is at most $\sqrt{\varepsilon'}$ with probability at least $1 - \sqrt{\varepsilon'}$ over v, i.e.,

$$P_V\left[d\left(E(T, V), U_r\right) \leq 2^{-\sqrt{N}/2\log N}\right] \geq 1 - 2^{-\sqrt{N}/2\log N} . \qquad (17)$$

Inequality (16) follows from (17) in a straight-forward way. □

Lemma 3 gives an upper bound on the effect of side information on the Rényi entropy of a random variable, and thus links information reconciliation and privacy amplification with universal hashing. We now need a similar result with respect to min-entropy H_∞. The proof of Lemma 10 is straight-forward and therefore omitted.

Lemma 10 *Let X and Q be random variables, and let $s > 0$. Then with probability at least $1 - 2^{-s}$ (taken over $q \in \mathcal{Q}$), we have $H_\infty(X) - H_\infty(X|Q = q) \leq \log|\mathcal{Q}| + s$.*

3.3 The Strong Secrecy Capacity $\overline{C_S}(P_{YZ|X})$

In this section we show that the definition of secrecy capacity in Csiszár and Körner's, hence also in Wyner's, model can be strengthened similarly to the weak and strong notions of secret-key rate: Not the rate, but the total amount of leaked information is negligible. Note that an additional uniformity condition is not necessary here since already the definition of C_S requires the key to be perfectly-uniformly distributed. Theorem 2 is the main result of this section.

Definition 7 For a distribution $P_{YZ|X}$, the *strong secrecy capacity* $\overline{C_S}(P_{YZ|X})$ is defined similarly to $C_S(P_{YZ|X})$, where the secrecy condition (15) is replaced by the stronger requirement $H(S|Z^N) > K - \varepsilon$.

Theorem 2 *For all conditional distributions $P_{YZ|X}$, we have $\overline{C_S}(P_{YZ|X}) = C_S(P_{YZ|X})$.*

Proof. The idea of the proof is to repeat the (weak) key generation a number of times and to compute from the block of resulting weak keys a secure string satisfying the stronger definition of secrecy capacity. More precisely, this is done by information reconciliation as described in Section 2.2, and by privacy amplification with extractors. Since the parties have, in contrast to the public-discussion model, no access to a noiseless public channel, all the error-correction and privacy-amplification information must be sent over the noisy channel specified by the conditional marginal distribution $P_{Y|X}(y,x) = \sum_{z\in\mathcal{Z}} P_{YZ|X}(y,z,x)$. However, the use of extractors instead of universal hashing for privacy amplification allows to keep the fraction of channel uses required for this communication negligibly small. This is precisely what is needed for showing equality of C_S and $\overline{C_S}$.

Let $R := C_S(P_{YZ|X})$. For a constant $\varepsilon' > 0$ and integers M and N to be determined later, assume that the key-generation procedure, with respect to the (weak) secrecy capacity C_S and parameters ε' and N, is repeated independently M times. Let $S^M := [S_1,\ldots,S_M]$ and $(S')^M := [S'_1,\ldots,S'_M]$ be the generated keys of Alice and Bob, respectively, and let $K = \lfloor (R-\varepsilon')N \rfloor$ be the length of (the binary strings) S_i and S'_i. From the fact that $\text{Prob}\,[S_i \neq S'_i] < \varepsilon'$ holds we conclude, by Fano's inequality, $H(S_i|S'_i) \leq \varepsilon' K + 1$ for all i, hence $H(S^M|(S')^M) \leq M(\varepsilon' K + 1)$.

For constants $\Delta_1, \Delta_2 > 0$, we conclude from Lemma 1 that there exists an error-correction-information function $h\,:\,(\{0,1\}^K)^M \longrightarrow \{0,1\}^{\lceil (1+\Delta_1)M(\varepsilon' K+1)\rceil}$ such that S^M can be determined from $(S')^M$ and $h(S^M)$ with probability at least $1-\Delta_2$ for sufficiently large M. Hence $\lceil (1+\Delta_1)M(\varepsilon' K + 1)\rceil$ message bits have to be transmitted over the channel $P_{Y|X}$ for error correction (see below).

According to the definition of the (weak) secrecy capacity C_S, we have $H(S_i|Z_i^N) \geq K(1-\varepsilon')$. For $\delta > 0$, let the event $\mathcal{F}(\delta)$, with respect to the random variables S and Z^N, be defined as in Lemma 6. For every $\alpha > 0$ we can achieve, for arbitrarily large (fixed) N and M, $MK \cdot \text{Prob}\,[\overline{\mathcal{F}(\delta)}] < \alpha$ and

$$H_\infty(S^M|(Z^N)^M = (z^N)^M, \mathcal{F}(\delta)) \geq M(K(1-\varepsilon') - 2\delta) + \log(1-\delta)\ .$$

The reason is that the statement of Lemma 6 also holds for the min-entropy H_∞ instead of H_2. The proof of this variant is exactly the same because it is ultimately based on uniform distributions, for which H_2 and H_∞ (and also H) are equal.

Let us now consider the effect of the error-correction information (partially) leaked to the adversary. According to Lemma 10, we have for $s > 0$ with probability at least $1 - 2^{-s}$

$$\begin{aligned} &H_\infty(S^M|(Z^N)^M = (z^N)^M, h(S^M) = h(s^M), \mathcal{F}(\delta)) \\ &\quad\geq\ M(K(1-\varepsilon') - 2\delta) + \log(1-\delta) - \lceil (1+\Delta_1)M(\varepsilon' K + 1)\rceil - s \\ &\quad\geq\ MK(1-\Delta_3) \end{aligned} \tag{18}$$

for some constant Δ_3 that can be made arbitrarily small by choosing N large enough, $s := \lceil \log M \rceil$, and Δ_1 as well as ε' small enough.

Let now for constants $\Delta_4, \Delta_5 > 0$ and sufficiently large M an extractor function E be given according to Lemma 9, i.e., $E\,:\,\{0,1\}^{MK} \times \{0,1\}^d \rightarrow \{0,1\}^r$ with $d \leq \Delta_4 MK$ and $r \geq MK(1-\Delta_3-\Delta_5)$ such that, for $\tilde{S} := E(S^M, V)$ the inequality

$$H(\tilde{S}|(Z^N)^M = (z^N)^M, h(S^M) = h(s^M), V, \mathcal{F}(\delta)) \geq r - 2^{-(MK)^{1/2-o(1)}}$$

holds for V uniformly distributed in $\{0,1\}^d$. Let S' be the key computed in the same way by Bob (where the random bits V are sent over to him by Alice using the channel $P_{Y|X}$ with an appropriate error-correcting code).

The resulting key $\tilde{S}$ of Alice is now close-to-uniformly, but not perfectly-uniformly distributed. Given the events $\mathcal{F}(\delta)$ and that inequality (18) holds, we have $H(\tilde{S}) \geq r - 2^{-(MK)^{1/2-o(1)}}$.

Let now, as in the proof of Lemma 7, S be the "uniformization" of $\tilde{S}$ (the random variable which is uniformly distributed in $\{0,1\}^r$ and jointly distributed with $\tilde{S}$ in such a way that $\text{Prob}\,[S \neq \tilde{S}]$ is minimized). It is clear that for any $\Delta_6 > 0$, $\text{Prob}\,[S \neq \tilde{S}] < \Delta_6$ can be achieved for sufficiently large M.

Let us finally consider the number of channel uses necessary for communicating the information for information reconciliation and privacy amplification. The number of bits to be transmitted is, according to the above, at most $\lceil(1+\Delta_1)M(\varepsilon' K+1)\rceil + \Delta_4 MK$. It is an immediate consequence of Shannon's channel-coding theorem (see for example [8]) that for arbitrary $\Delta_7, \Delta_8 > 0$ and sufficiently large M, the number of channel uses for transmitting these messages can be at most

$$\frac{MK((1+\Delta_1)\varepsilon' + \Delta_4) + (1+\Delta_1)M + 1}{C(P_{Y|X}) - \Delta_7}$$

(where $C(P_{Y|X})$ is the capacity of the channel $P_{Y|X}$ from Alice to Bob), keeping the probability of a decoding error below Δ_8. Note that $C(P_{Y|X}) > 0$ clearly holds when $C_S(P_{YZ|X}) > 0$. (If $C(P_{Y|X}) = 0$, the statement of the theorem is hence trivially satisfied.) Thus the total number of channel uses for the entire key generation can be made smaller than $MN(1+\Delta_9)$ for arbitrarily small $\Delta_9 > 0$ and sufficiently large N.

From the above we can now conclude that S is a perfectly uniformly distributed string of length $r = (1-o(1))RL$, where $L = (1+o(1))MN$ is the total number of channel uses. Furthermore, we have by construction $\text{Prob}\,[S' \neq S] = o(1)$ and finally

$$\begin{aligned} H(S|Z^L) = H(S) - I(S;Z^L) \ &\geq \ H(S) - I(\tilde{S};Z^L) \qquad (19) \\ = r - 2^{-(MK)^{1/2-o(1)}} - r \cdot (2^{-s} + \text{Prob}\,[\overline{\mathcal{F}(\delta)}]) \ &= \ r - o(1) \ . \end{aligned}$$

The inequality in (19) holds because $Z^L \to \tilde{S} \to S$ is a Markov chain. Hence the achievable rate with respect to the strong secrecy-capacity definition is of order $(1-o(1))R = (1-o(1))C_S(P_{YZ|X})$, thus $\overline{C_S}(P_{YZ|X}) = C_S(P_{YZ|X})$ holds. □

4 Concluding Remarks

The fact that the previous security definitions of information-theoretic key agreement in the noisy-channel models by Wyner [18] as well as Csiszár and Körner [9] and the correlated-randomness settings of Maurer [12] and Ahlswede-Csiszár [1] are unsatisfactory was our motivation for studying much stronger definitions which tolerate the adversary to obtain only a negligibly small amount of information about the generated key. We have shown that in all these models, the achievable key-generation rates with respect to the weak and strong definitions are asymptotically identical. Therefore, the old notions can be entirely replaced by the new definitions.

References

1. R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptography – Part I: secret sharing, *IEEE Transactions on Information Theory*, Vol. 39, No. 4, pp. 1121–1132, 1993.
2. C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, Generalized privacy amplification, *IEEE Transactions on Information Theory*, Vol. 41, No. 6, pp. 1915–1923, 1995.
3. C. H. Bennett, G. Brassard, and J.-M. Robert, Privacy amplification by public discussion, *SIAM Journal on Computing*, Vol. 17, pp. 210–229, 1988.
4. G. Brassard and L. Salvail, Secret-key reconciliation by public discussion, *Advances in Cryptology - EUROCRYPT '93*, Lecture Notes in Computer Science, Vol. 765, pp. 410–423, Springer-Verlag, 1994.
5. C. Cachin, *Entropy measures and unconditional security in cryptography*, Ph. D. Thesis, ETH Zürich, Hartung-Gorre Verlag, Konstanz, 1997.
6. C. Cachin and U. M. Maurer, Linking information reconciliation and privacy amplification, *Journal of Cryptology*, Vol. 10, No. 2, pp. 97–110, 1997.
7. J. L. Carter and M. N. Wegman, Universal classes of hash functions, *Journal of Computer and System Sciences*, Vol. 18, pp. 143–154, 1979.
8. T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley Series in Telecommunications, 1992.
9. I. Csiszár and J. Körner, Broadcast channels with confidential messages, *IEEE Transactions on Information Theory*, Vol. 24, No. 3, pp. 339–348, 1978.
10. M. van Dijk, *Secret key sharing and secret key generation*, Ph. D. Thesis, Technische Universiteit Eindhoven, 1997.
11. J. Håstad, R. Impagliazzo, L. Levin, and M. Luby, "Construction of a pseudorandom generator from any one-way function," ICSI Tech. Rep. 91–068, 1991.
12. U. M. Maurer, Secret key agreement by public discussion from common information, *IEEE Transactions on Information Theory*, Vol. 39, No. 3, pp. 733–742, May 1993.
13. U. M. Maurer, The strong secret key rate of discrete random triples, in *Communication and Cryptography – Two Sides of One Tapestry*, Kluwer Academic Publishers, pp. 271–285, 1994.
14. U. M. Maurer and S. Wolf, Unconditionally secure key agreement and the intrinsic conditional information, *IEEE Transactions on Information Theory*, Vol. 45, No. 2, pp. 499–514, 1999.
15. N. Nisan and D. Zuckerman, Randomness is linear in space, *Journal of Computer and System Sciences*, Vol. 52, No. 1, pp. 43–52, 1996.
16. L. Trevisan, Construction of Extractors Using Pseudorandom Generators, *Proc. of the 31st Symposium on Theory of Computing (STOC)*, ACM, pp. 141–148, 1999.
17. S. P. Vadhan, Extracting all the randomness from a weakly random source, *Electronic Colloquium on Computational Complexity*, Technical Report TR98-047, December 1998.

18. A. D. Wyner, The wire-tap channel, *Bell System Technical Journal*, Vol. 54, No. 8, pp. 1355–1387, 1975.