4 Secret-key capacity

In Chapter 3, we considered the transmission of information over a noisy broadcast channel subject to reliability and security constraints; we showed that appropriate coding schemes can exploit the presence of noise to confuse the eavesdropper and guarantee some amount of information-theoretic security. It is important to note that the wiretap channel model assumes that all communications occur over the channel, hence communications are inherently rate-limited and one-way. Consequently, the results obtained do not fully capture the role of noise for secrecy; in particular, for situations in which the secrecy capacity is zero, it is not entirely clear whether this stems from the lack of any "physical advantage" over the eavesdropper or the restrictions imposed on the communication schemes.

The objective of this chapter is to study more precisely the fundamental role of noise in information-theoretic security. Instead of studying how we can communicate messages securely over a noisy channel, we now analyze how much secrecy we can extract from the noise itself in the form of a secret key. Specifically, we assume that the legitimate parties and the eavesdropper observe the realizations of correlated random variables and that the legitimate parties attempt to agree on a secret key unknown to the eavesdropper. To isolate the role played by noise, we remove restrictions on communication schemes and we assume that the legitimate parties can distill their key by communicating over a *two-way*, *public*, *noiseless*, and *authenticated* channel at no cost. Contrary to the case in Chapter 3, in which the natural metric of interest was the number of message bits that one could transmit securely and reliably per channel use, here the relevant metric is the number of secret-key bits distilled per observation of the correlated random variables.

We start this chapter by introducing two standard models for secret-key agreement, called the source model and the channel model, and by defining key-distillation strategies (Section 4.1). We then discuss the fundamental limits of key generation for a source model (Section 4.2) and we study in detail a specific type of key-distillation strategy, which we call "sequential key-distillation strategies" (Section 4.3). For these strategies, we prove results under a strong secrecy condition and we show how to construct practical strategies. Finally, we study the fundamental limit of secret-key generation over a channel model (Section 4.4) and show that the fundamental limit remains unchanged under a strong secrecy condition (Section 4.5). In other words, we prove that strong secrecy comes "for free," that is, a secure rate achievable with weak secrecy is also achievable with strong secrecy. This result is crucial to justify a posteriori the cryptographic relevance of the results derived in Chapter 3 for a weak-secrecy condition.

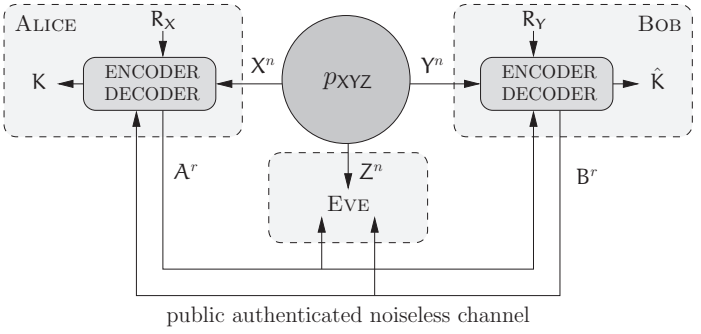

Figure 4.1 Source model for secret-key agreement.

4.1 Source and channel models for secret-key agreement

As illustrated in Figure 4.1, a *source model* for secret-key agreement represents a situation in which three parties, Alice, Bob, and Eve, observe the realizations of a DMS $(\mathcal{XYZ}, p_{XYZ})$ with three components. The DMS is assumed to be outside the control of all parties, but its statistics are known. By convention, component X is observed by Alice, component Y by Bob, and component Z by Eve. Alice and Bob's objective is to process their observations and agree on a key K about which Eve should have no information. To capture the essence of the problem, few restrictions are placed on the communication between Alice and Bob: they can exchange messages over a noiseless, two-way, and authenticated channel; however, to avoid trivializing the problem, the two-way channel is *public*, that is all messages are overheard by Eve and the existence of the public channel does not provide Alice and Bob with an explicit advantage over Eve. The only real simplifying assumption is the existence of an authentication mechanism that prevents Eve from tampering with communications over the public channel. Finally, we allow Alice and Bob to randomize the messages they transmit, which we model with sources of local randomness as done in Chapter 3. Alice has access to a DMS $\left(\mathcal{R}_{\mathcal{X}}, p_{R_X}\right)$ and Bob has access to a DMS $\left(\mathcal{R}_{\mathcal{Y}}, p_{R_Y}\right)$, which are mutually independent and independent of the DMS $(\mathcal{XYZ}, p_{XYZ})$. The rules by which Alice and Bob compute the messages they exchange over the public channel and agree on a key define a *key-distillation strategy*.[1]

Definition 4.1. *A $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ for a source model with DMS $(\mathcal{XYZ}, p_{XYZ})$ consists of*

- *a key alphabet $\mathcal{K} = [\![1, 2^{nR}]\!]$;*
- *an alphabet $\mathcal{A}$ used by Alice to communicate over the public channel;*
- *an alphabet $\mathcal{B}$ used by Bob to communicate over the public channel;*
- *a source of local randomness for Alice $\left(\mathcal{R}_{\mathcal{X}}, p_{R_X}\right)$;*
- *a source of local randomness for Bob $\left(\mathcal{R}_{\mathcal{Y}}, p_{R_Y}\right)$;*

[1] We use the term "strategy" instead of "code" because a key does not carry information by itself.

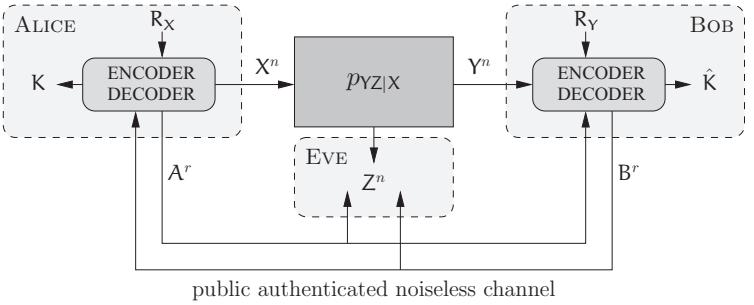

Figure 4.2 Channel model for secret-key agreement.

- *an integer* $r \in \mathbb{N}^*$ *that represents the number of rounds of communication;*
- r *encoding functions* $f_i : \mathcal{X}^n \times \mathcal{B}^{i-1} \times \mathcal{R}_\mathcal{X} \to \mathcal{A}$ *for* $i \in [\![1, r]\!]$*;*
- r *encoding functions* $g_i : \mathcal{Y}^n \times \mathcal{A}^{i-1} \times \mathcal{R}_\mathcal{Y} \to \mathcal{B}$ *for* $i \in [\![1, r]\!]$*;*
- *a key-distillation function* $\kappa_\mathrm{a} : \mathcal{X}^n \times \mathcal{B}^r \times \mathcal{R}_\mathcal{X} \to \mathcal{K}$*;*
- *a key-distillation function* $\kappa_\mathrm{b} : \mathcal{Y}^n \times \mathcal{A}^r \times \mathcal{R}_\mathcal{Y} \to \mathcal{K}$*;*

and operates as follows:

- *Alice observes* n *realizations of the source* x^n *while Bob observes* y^n*;*
- *Alice generates a realization* r_x *of her source of local randomness while Bob generates* r_y *from his;*
- *in round* $i \in [\![1, r]\!]$*, Alice transmits* $a_i = f_i\left(x^n, b^{i-1}, r_x\right)$ *while Bob transmits* $b_i = g_i\left(y^n, a^{i-1}, r_y\right)$*;*
- *after round* r*, Alice computes a key* $k = \kappa_\mathrm{a}\left(x^n, b^r, r_x\right)$ *while Bob computes a key* $\hat{k} = \kappa_\mathrm{b}\left(y^n, a^r, r_y\right)$.

By convention, we set $\mathcal{A}^0 \triangleq 0$ and $\mathcal{B}^0 \triangleq 0$. Note that the number of rounds r and the DMSs $\left(\mathcal{R}_\mathcal{X}, p_{\mathrm{R_X}}\right)$ and $\left(\mathcal{R}_\mathcal{Y}, p_{\mathrm{R_Y}}\right)$ can be optimized as part of the strategy design. The $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ is assumed known ahead of time to Alice, Bob, and Eve.

The source model assumes the existence of an uncontrollable external source of randomness and abstracts the physical origin of the randomness completely. Although this is a strong assumption, there are several practical situations in which this would be a reasonable model. For instance, in wireless sensor networks, devices may monitor a physical phenomenon (change in temperature or pressure) whose statistics may be known but whose complexity is such that we can reasonably assume that it cannot be controlled. Nevertheless, it is legitimate to wonder what happens if the source of randomness is partially controlled by one of the parties. The analysis of situations in which the source is partially controlled by the eavesdropper is not fully understood and is not covered in this book. We refer the interested reader to the bibliographical notes at the end of the chapter for references to existing results.

It is somewhat less arduous to study the situation in which the source is partially controlled by one of the legitimate parties. In this case, the model is called a *channel model* for secret-key agreement and is as illustrated in Figure 4.2. Instead of observing

the realizations of an external source, Bob and Eve now observe the outputs of a DMC $(\mathcal{X}, p_{\mathrm{YZ|X}}, \mathcal{Y}, \mathcal{Z})$ whose input is controlled by Alice. Alice and Bob have again access to a public, noiseless, two-way, and authenticated channel over which they can exchange messages to agree on a secret key. We also assume that Alice and Bob have access to sources of local randomness $\left(\mathcal{R}_{\mathcal{X}}, p_{\mathrm{R_X}}\right)$ and $\left(\mathcal{R}_{\mathcal{Y}}, p_{\mathrm{R_Y}}\right)$ to randomize their communications. Key-distillation strategies for the channel model are more sophisticated than those for the source model, because Alice can use the feedback provided by Bob to adapt the symbols she sends in the channel. Despite the similarity between the channel model for secret-key agreement and the wiretap channel studied in Chapter 3, note that the problems are different: in a channel model for secret-key agreement, the broadcast channel is used not only to communicate messages but also to generate randomness.

Definition 4.2. *A $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ for a channel model with DMC $(\mathcal{X}, p_{\mathrm{YZ|X}}, \mathcal{Y}, \mathcal{Z})$ consists of*

- *a key alphabet $\mathcal{K} = [\![1, 2^{nR}]\!]$;*
- *an alphabet $\mathcal{A}$ used by Alice to communicate over the public channel;*
- *an alphabet $\mathcal{B}$ used by Bob to communicate over the public channel;*
- *a source of local randomness for Alice $\left(\mathcal{R}_{\mathcal{X}}, p_{\mathrm{R_X}}\right)$;*
- *a source of local randomness for Bob $\left(\mathcal{R}_{\mathcal{Y}}, p_{\mathrm{R_Y}}\right)$;*
- *an integer $r \in \mathbb{N}^*$ that represents the number of rounds of communication;*
- *a set of n distinct integers $\{i_j\}_n \subseteq [\![1, r]\!]$ that represents the rounds in which Alice transmits a symbol over the channel;*
- *$r - n$ encoding functions $f_i : \mathcal{B}^{i-1} \times \mathcal{R}_{\mathcal{X}} \to \mathcal{A}$ for $i \in [\![1, r]\!] \setminus \{i_j\}_n$;*
- *$r - n$ encoding functions g_i for $i \in [\![1, r]\!] \setminus \{i_j\}_n$ of the form $g_i : \mathcal{Y}^j \times \mathcal{A}^{i-1} \times \mathcal{R}_{\mathcal{Y}} \to \mathcal{B}$ if $i \in [\![i_j + 1, i_{j+1} - 1]\!]$;*
- *n functions $h_j : \mathcal{B}^{i_j - 1} \times \mathcal{R}_{\mathcal{X}} \to \mathcal{X}$ for $j \in [\![1, n]\!]$ to generate channel inputs;*
- *a key-distillation function $\kappa_{\mathrm{a}} : \mathcal{X}^n \times \mathcal{B}^r \times \mathcal{R}_{\mathcal{X}} \to \mathcal{K}$;*
- *a key-distillation function $\kappa_{\mathrm{b}} : \mathcal{Y}^n \times \mathcal{A}^r \times \mathcal{R}_{\mathcal{Y}} \to \mathcal{K}$;*

and operates as follows:

- *Alice generates a realization r_x of her source of local randomness while Bob generates r_y from his;*
- *in round $i \in [\![1, i_1 - 1]\!]$, Alice transmits message $a_i = f_i\left(b^{i-1}, r_x\right)$ and Bob transmits message $b_i = g_i\left(a^{i-1}, r_y\right)$;*
- *in round i_j with $j \in [\![1, n]\!]$, Alice transmits symbol $x_j = h_j\left(b^{i_j - 1}, r_x\right)$ over the channel, and Bob and Eve observe the symbols y_j and Z_j, respectively;*
- *in round $i \in [\![i_j + 1, i_{j+1} - 1]\!]$, Alice transmits message $a_i = f_i\left(x^j, b^{i-1}, r_x\right)$ and Bob transmits message $b_i = g_i\left(y^j, a^{i-1}, r_y\right)$;*
- *after the last round, Alice computes a key $k = \kappa_{\mathrm{a}}(x^n, b^r, r_x)$ and Bob computes a key $\hat{k} = \kappa_{\mathrm{b}}\left(y^n, a^r, r_y\right)$.*

By convention, we set $i_{n+1} \triangleq r + 1, i_0 = 0, \mathcal{A}^0 = 0$, and $\mathcal{B}^0 \triangleq 0$. Note that the number of rounds r, the indices $\{i_j\}_n$ of the rounds during which a symbol is transmitted over the channel, and the sources of local randomness $\left(\mathcal{R}_{\mathcal{X}}, p_{\mathrm{R_X}}\right)$ and $\left(\mathcal{R}_{\mathcal{Y}}, p_{\mathrm{R_Y}}\right)$ can be

optimized as part of the strategy design. Again, the key-distillation strategy $\mathcal{S}_n$ is assumed known to Alice, Bob, and Eve ahead of time.

Remark 4.1. *A wiretap code $\mathcal{C}_n$ is a key-distillation strategy for a channel model since Alice can use $\mathcal{C}_n$ to transmit uniform secret keys to Bob directly without using the public channel. In general, key-distillation strategies that exploit the public channel are more powerful, but we will see an example of a channel model for which using a wiretap code is an optimal key-distillation strategy in Section 4.4.*

For both source and channel models, the performance of a $\left(2^{nR}, n\right)$ key-distillation strategy $\mathcal{S}_n$ is measured in terms of the average probability of error

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \triangleq \mathbb{P}\left[\mathsf{K} \neq \hat{\mathsf{K}} \middle| \mathcal{S}_n\right],$$

in terms of the information leakage to the eavesdropper,

$$\mathbf{L}(\mathcal{S}_n) \triangleq \mathbb{I}(\mathsf{K}; \mathsf{Z}^n \mathsf{A}^r \mathsf{B}^r | \mathcal{S}_n),$$

and in terms of the uniformity of the keys,

$$\mathbf{U}(\mathcal{S}_n) \triangleq \log\lceil 2^{nR} \rceil - \mathbb{H}(\mathsf{K}|\mathcal{S}_n).$$

Note that, by definition, $\mathbf{U}(\mathcal{S}_n) \geqslant 0$ with equality if and only if the key is exactly uniformly distributed.

Remark 4.2. *It is possible to combine the information leaked to the eavesdropper and the uniformity of the key into a single quantity called the* security index *and defined as*

$$\log\lceil 2^{nR} \rceil - \mathbb{H}(\mathsf{K}|\mathsf{Z}^n \mathsf{A}^r \mathsf{B}^r \mathcal{S}_n).$$

The security index is equal to zero if and only if the key is uniformly distributed and unknown to the eavesdropper. However, we choose to study $\mathbf{U}(\mathcal{S}_n)$ and $\mathbf{L}(\mathcal{S}_n)$ independently to emphasize that these are different constraints.

Definition 4.3. *A weak secret-key rate R is achievable for a source or channel model if there exists a sequence of $(2^{nR}, n)$ key-distillation strategies $\{\mathcal{S}_n\}_{n \geqslant 1}$ such that*

$$\lim_{n\to\infty} \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) = 0 \qquad \textit{(reliability)}, \tag{4.1}$$

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{L}(\mathcal{S}_n) = 0 \qquad \textit{(weak secrecy)}, \tag{4.2}$$

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{U}(\mathcal{S}_n) = 0 \qquad \textit{(weak uniformity)}. \tag{4.3}$$

The corresponding keys are called weak secret keys. If the strategies $\{\mathcal{S}_n\}_{n \geqslant 1}$ exploit public messages sent either from Alice to Bob only or from Bob to Alice only, the secret-key rate R is said to be achievable with one-way *communication; otherwise, R is said to be achievable with* two-way *communication.*

Condition (4.1) means that Alice and Bob should agree on a common key with high probability. Condition (4.2) requires that Eve, who has access to information through her randomness Z^n and the messages exchanged over the public channel $\mathsf{A}^r\mathsf{B}^r$, obtains

a negligible *rate* of information about the key; this condition is tantamount to the full secrecy rate condition studied in the context of wiretap channels. Finally, Condition (4.3) requires the key rate to be almost that of a uniform key, which is a necessary property of secret keys if they are are to be used to protect messages with a one-time pad as seen in Theorem 3.1. As already discussed in Section 3.3, Conditions (4.2) and (4.3) are called "weak" because they impose constraints on the rate of information leaked and on the rate at which the entropy of a key approaches that of a uniform distribution.

Remark 4.3. *In principle, we could attempt to characterize a rate–equivocation region for keys and consider partially secret keys for which* $(1/n)\mathbf{E}(\mathsf{K}|Z^n A^r \mathrm{B}^r \mathcal{S}_n) \geqslant R_{\mathrm{e}}$ *with* $R_{\mathrm{e}} \leqslant R$*; however, it is not clear what the cryptographic purpose of such keys would be. In addition, note that partially secret keys can be obtained by expanding secret keys with a known function or by adding known bits; therefore, without loss of generality, we restrict ourselves to the analysis of secret keys as in Definition 4.3.*

Before we investigate the fundamental limits of achievable weak secret-key rates, it is worth looking at how weak secret keys can be used. Consider a $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ such that

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \leqslant \epsilon, \quad \frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \epsilon, \quad \text{and} \quad \frac{1}{n}\mathbf{U}(\mathcal{S}_n) \leqslant \epsilon$$

for some $\epsilon > 0$. Assume the resulting weak secret key K is used in the one-time-pad encryption of a message $\mathsf{M} \in \mathcal{K}$ that is independent of all variables involved in the key-generation process. Since K is not exactly uniform and not totally unknown to the adversary, we should not expect to obtain the perfect secrecy discussed in Chapter 3. In fact, note that

$$\begin{aligned}
\mathbb{I}(\mathsf{M} \oplus \mathsf{K}, A^r, \mathrm{B}^r, Z^n; \mathsf{M}|\mathcal{S}_n) &= \mathbb{I}(A^r \mathrm{B}^r Z^n; \mathsf{M}|\mathcal{S}_n) + \mathbb{I}(\mathsf{M} \oplus \mathsf{K}; \mathsf{M}|A^r \mathrm{B}^r Z^n \mathcal{S}_n) \\
&\overset{(a)}{=} \mathbb{I}(\mathsf{M} \oplus \mathsf{K}; \mathsf{M}|A^r \mathrm{B}^r Z^n \mathcal{S}_n) \\
&\overset{(b)}{=} \mathbb{H}(\mathsf{M} \oplus \mathsf{K}|A^r \mathrm{B}^r Z^n \mathcal{S}_n) - \mathbb{H}(\mathsf{K}|\mathsf{M} A^r \mathrm{B}^r Z^n \mathcal{S}_n) \\
&\leqslant \log\lceil 2^{nR} \rceil - \mathbb{H}(\mathsf{K}|\mathsf{M} A^r \mathrm{B}^r Z^n \mathcal{S}_n) \\
&\overset{(c)}{\leqslant} \mathbb{H}(\mathsf{K}|\mathcal{S}_n) + n\epsilon - \mathbb{H}(\mathsf{K}|A^r \mathrm{B}^r Z^n \mathcal{S}_n) \\
&= \mathbf{L}(\mathcal{S}_n) + n\epsilon \\
&\overset{(d)}{\leqslant} n\delta(\epsilon),
\end{aligned} \tag{4.4}$$

where (a) follows from $\mathbb{I}(A^r \mathrm{B}^r Z^n; \mathsf{M}|\mathcal{S}_n) = 0$ since M is independent of $A^r \mathrm{B}^r Z^n$, (b) follows from $\mathbb{H}(\mathsf{M} \oplus \mathsf{K}|\mathsf{M} A^r \mathrm{B}^r Z^n \mathcal{S}_n) = \mathbb{H}(\mathsf{K}|\mathsf{M} A^r \mathrm{B}^r Z^n \mathcal{S}_n)$, (c) follows from $(1/n)\mathbf{U}(\mathcal{S}_n) \leqslant \epsilon$, and (d) follows from $(1/n)\mathbf{L}(\mathcal{S}_n) \leqslant \epsilon$. Therefore, we can prove only that this encryption guarantees the weak secrecy condition

$$\frac{1}{n}\mathbb{I}(\mathsf{M} \oplus \mathsf{K}, A^r, \mathrm{B}^r, Z^n; \mathsf{M}|\mathcal{S}_n) \leqslant \delta(\epsilon).$$

This result is somewhat unsatisfactory, but it can be improved if we strengthen the notion of the achievable rate as follows.

Definition 4.4. *A strong secret-key rate R is achievable for a source or channel model if there exists a sequence of $(2^{nR}, n)$ key-distillation strategies $\{\mathcal{S}_n\}_{n\geqslant 1}$ such that*

$$\lim_{n\to\infty} \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) = 0 \qquad \textit{(reliability)}, \tag{4.5}$$

$$\lim_{n\to\infty} \mathbf{L}(\mathcal{S}_n) = 0 \qquad \textit{(strong secrecy)}, \tag{4.6}$$

$$\lim_{n\to\infty} \mathbf{U}(\mathcal{S}_n) = 0 \qquad \textit{(strong uniformity)}. \tag{4.7}$$

The corresponding keys are called strong secret keys.

The secrecy condition and the uniformity condition for strong secret-key rates differ from their weak counterparts in that they do not involve any normalization by n; hence, achievable strong secret-key rates are also achievable weak secret-key rates. Consider now a $(2^{nR}, n)$ key-distillation strategy such that

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \leqslant \epsilon, \quad \mathbf{L}(\mathcal{S}_n) \leqslant \epsilon, \quad \text{and} \quad \mathbf{U}(\mathcal{S}_n) \leqslant \epsilon,$$

and assume as done earlier that the resulting strong secret key K is used for the one-time-pad encryption of a message M that is independent of all random variables involved in the key-distillation process. We can reiterate the calculation in (4.4) and the reader can check that we can then guarantee the strong secrecy condition

$$\mathbb{I}(\mathsf{M} \oplus \mathsf{K}, \mathsf{A}^r, \mathsf{B}^r, \mathsf{Z}^n; \mathsf{M}|\mathcal{S}_n) \leqslant \delta(\epsilon).$$

Note that this still does not match the perfect secrecy condition, but does guarantee a reasonable secrecy level if ϵ is small enough.

4.2 Secret-key capacity of the source model

In this section, we study the *secret-key capacity*, which is defined as the supremum of secret-key rates achievable for a source model. Since a secret-key rate is defined as a number of secret-key bits per realization of a DMS, the secret-key capacity does not account for the amount of communication required to distill keys, which could be arbitrarily large.

Definition 4.5. *The weak secret-key capacity of a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$ is*

$$C_{\mathrm{s}}^{\mathrm{SM}} \triangleq \sup\{R : R \text{ is an achievable weak secret-key rate}\}.$$

Similarly, the strong secret-key capacity of a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$ is

$$\overline{C}_{\mathrm{s}}^{\mathrm{SM}} \triangleq \sup\{R : R \text{ is an achievable strong secret-key rate}\}.$$

When required, we write $C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{XYZ}})$ in place of $C_{\mathrm{s}}^{\mathrm{SM}}$ to explicitly specify the underlying distribution of the DMS under consideration. It follows directly from the definition of

achievable secret-key rates that $\overline{C}_{\mathrm{s}}^{\mathrm{SM}} \leqslant C_{\mathrm{s}}^{\mathrm{SM}}$. We will show at the end of Section 4.3 that $\overline{C}_{\mathrm{s}}^{\mathrm{SM}} = C_{\mathrm{s}}^{\mathrm{SM}}$ but, until then, we restrict our attention to the weak secret-key capacity. Although this restriction will prove unnecessary, the study of weak secret-key capacity is less arduous than that of strong secret-key capacity and still provides useful insight into the design of key-distillation strategies. In addition, note that any upper bound we derive for $C_{\mathrm{s}}^{\mathrm{SM}}$ is automatically an upper bound for $\overline{C}_{\mathrm{s}}^{\mathrm{SM}}$.

A closed-form expression for the secret-key capacity for a general source model remains elusive. Nevertheless, it is possible to obtain simple upper and lower bounds that are useful in many situations.

Theorem 4.1 (Maurer, Ahlswede, and Csiszár)**.** *The weak secret-key capacity of a source model* $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$ *satisfies*

$$\mathbb{I}(\mathsf{X};\mathsf{Y}) - \min(\mathbb{I}(\mathsf{X};\mathsf{Z}), \mathbb{I}(\mathsf{Y};\mathsf{Z})) \leqslant C_{\mathrm{s}}^{\mathrm{SM}} \leqslant \min(\mathbb{I}(\mathsf{X};\mathsf{Y}), \mathbb{I}(\mathsf{X},\mathsf{Y}|\mathsf{Z})).$$

Moreover, the secret-key rate $\mathbb{I}(\mathsf{X};\mathsf{Y}) - \min(\mathbb{I}(\mathsf{X};\mathsf{Z}), \mathbb{I}(\mathsf{Y};\mathsf{Z}))$ *is achievable with one-way communication.*

Proof. We provide proofs of the result in Sections 4.2.1, 4.2.2, and 4.2.3. The proof in Section 4.2.1 leverages the results obtained for WTCs in Chapter 3, whereas the proof in Section 4.2.2 provides a more direct approach based on Slepian–Wolf codes. □

The lower bound $\mathbb{I}(\mathsf{X};\mathsf{Y}) - \min(\mathbb{I}(\mathsf{X};\mathsf{Z}), \mathbb{I}(\mathsf{Y};\mathsf{Z}))$ can be understood as the difference between the information rate between Alice and Bob and some information rate leaked to Eve. However, in contrast to the results obtained in Chapter 3, Alice and Bob can choose whether the information rate obtained by Eve is leaked from Alice ($\mathbb{I}(\mathsf{X};\mathsf{Y})$) or from Bob ($\mathbb{I}(\mathsf{Y};\mathsf{Z})$). We will see in the course of the proof that this result stems from the possibility of two-way communication over the public channel. Before we prove Theorem 4.1, it is also useful to note that, in general, the bounds in Theorem 4.1 are loose. In Section 4.3.1, we will see an example of a source model for which the lower bound is useless because $\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z}) < 0$ and $\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z}) < 0$. In Section 4.2.4, we will provide an example of a source model for which $\mathbb{I}(\mathsf{X};\mathsf{Y}|\mathsf{Z}) > 0$ and $\mathbb{I}(\mathsf{X};\mathsf{Y}) > 0$ while $C_{\mathrm{s}}^{\mathrm{SM}} = 0$. Nevertheless, there are several situations in which the bounds are tight.

Corollary 4.1. *Consider a source model with DMS* $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$*:*

- *if* Z *is independent of* (X,Y)*, then* $C_{\mathrm{s}}^{\mathrm{SM}} = \mathbb{I}(\mathsf{X};\mathsf{Y})$*;*
- *if* $\mathsf{X} \to \mathsf{Y} \to \mathsf{Z}$ *forms a Markov chain, then* $C_{\mathrm{s}}^{\mathrm{SM}} = \mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z})$*;*
- *if* $\mathsf{Y} \to \mathsf{X} \to \mathsf{Z}$ *forms a Markov chain, then* $C_{\mathrm{s}}^{\mathrm{SM}} = \mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z})$*.*

Proof. If Z is independent of X and Y, then both bounds in Theorem 4.1 are equal to $\mathbb{I}(\mathsf{X};\mathsf{Y})$. If $\mathsf{X} \to \mathsf{Y} \to \mathsf{Z}$ forms a Markov chain then

$$\begin{aligned}\mathbb{I}(\mathsf{X};\mathsf{Y}|\mathsf{Z}) &= \mathbb{I}(\mathsf{X};\mathsf{Y}\mathsf{Z}) - \mathbb{I}(\mathsf{X};\mathsf{Z}) \\ &= \mathbb{I}(\mathsf{X};\mathsf{Y}) + \mathbb{I}(\mathsf{X};\mathsf{Z}|\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z}) \\ &= \mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z}).\end{aligned}$$

Finally, if $\mathsf{Y} \to \mathsf{X} \to \mathsf{Z}$, then Z and Y are conditionally independent given X, and

$$\begin{aligned}\mathbb{I}(\mathsf{X};\mathsf{Y}|\mathsf{Z}) &= \mathbb{I}(\mathsf{XZ};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z}) \\ &= \mathbb{I}(\mathsf{X};\mathsf{Y}) + \mathbb{I}(\mathsf{Z};\mathsf{Y}|\mathsf{X}) - \mathbb{I}(\mathsf{Y};\mathsf{Z}) \\ &= \mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z}).\end{aligned}$$

□

Remark 4.4. *If we drop the term Z^n in the information leaked $\mathbf{L}(\mathcal{S}_n)$ and if we use $\mathbb{I}(\mathsf{K};\mathsf{A}^r\mathsf{B}^r|\mathcal{S}_n)$ as the measure of secrecy, then the secret-key capacity is called the* private-key capacity. *The private-key capacity measures the maximum key rate with respect to an eavesdropper who observes communications over the public channel only and who disregards the realizations of the source Z^n. The private-key capacity can also be viewed as a special case of the secret-key capacity for a source model in which Z is independent of X and Y.*

4.2.1 Secret-key distillation based on wiretap codes

In this section, we establish that the secret-key capacity is at least as large as $\mathbb{I}(\mathsf{X};\mathsf{Y}) - \min(\mathbb{I}(\mathsf{X};\mathsf{Z}), \mathbb{I}(\mathsf{Y};\mathsf{Z}))$ by leveraging the results of Chapter 3 about the secrecy capacity of WTCs. The basic idea of the proof is to analyze a specific key-distillation strategy that creates a *conceptual* WTC, for which we know the existence of wiretap codes and achievable secure communication rates.

We first show that $\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z})$ is an achievable secret-key rate. Assume that Alice wants to send a symbol $u \in \mathcal{X}$ that is independent of the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$ over the public channel. Instead of transmitting u directly, she observes a realization $x \in \mathcal{X}$ of the source and transmits $u \oplus x$, where $\oplus$ denotes addition modulo-$|\mathcal{X}|$ over $\mathcal{X}$. At the same time, Bob observes y and Eve observes z. In effect, this operation creates a memoryless WTC with input U, for which Bob receives the *pair* of symbols $(\mathsf{Y}, \mathsf{U} \oplus \mathsf{X})$ and Eve receives the *pair* of symbols $(\mathsf{Z}, \mathsf{U} \oplus \mathsf{X})$. We know from Corollary 3.4 that the secrecy capacity of this WTC is at least

$$\mathbb{I}(\mathsf{U};\mathsf{Y},\mathsf{U}\oplus\mathsf{X}) - \mathbb{I}(\mathsf{U};\mathsf{Z},\mathsf{U}\oplus\mathsf{X}) = \mathbb{H}(\mathsf{U}|\mathsf{Z},\mathsf{U}\oplus\mathsf{X}) - \mathbb{H}(\mathsf{U}|\mathsf{Y},\mathsf{U}\oplus\mathsf{X}),$$

where the distribution p_{U} over $\mathcal{X}$ can be chosen arbitrarily. Here, we choose p_{U} to be the uniform distribution over $\mathcal{X}$. Then,

$$\begin{aligned}\mathbb{H}(\mathsf{U}|\mathsf{Z},\mathsf{U}\oplus\mathsf{X}) &= \mathbb{H}(\mathsf{U},\mathsf{U}\oplus\mathsf{X}|\mathsf{Z}) - \mathbb{H}(\mathsf{U}\oplus\mathsf{X}|\mathsf{Z}) \\ &= \mathbb{H}(\mathsf{U}|\mathsf{Z}) + \mathbb{H}(\mathsf{U}\oplus\mathsf{X}|\mathsf{U},\mathsf{Z}) - \mathbb{H}(\mathsf{U}\oplus\mathsf{X}|\mathsf{Z}).\end{aligned}$$

Since U is independent of (X,Z), $\mathbb{H}(\mathsf{U}|\mathsf{Z}) = \mathbb{H}(\mathsf{U})$ and $\mathbb{H}(\mathsf{U}\oplus\mathsf{X}|\mathsf{U},\mathsf{Z}) = \mathbb{H}(\mathsf{X}|\mathsf{U}\mathsf{Z}) = \mathbb{H}(\mathsf{X}|\mathsf{Z})$. Additionally, since U is uniformly distributed over $\mathcal{X}$, the crypto lemma applies and $\mathbb{H}(\mathsf{U}\oplus\mathsf{X}|\mathsf{Z}) = \mathbb{H}(\mathsf{U})$. Therefore,

$$\mathbb{H}(\mathsf{U}|\mathsf{Z},\mathsf{U}\oplus\mathsf{X}) = \mathbb{H}(\mathsf{X}|\mathsf{Z}).$$

Similarly, we can show that $\mathbb{H}(\mathsf{U}|\mathsf{Y}, \mathsf{U} \oplus \mathsf{X}) = \mathbb{H}(\mathsf{X}|\mathsf{Y})$; therefore,

$$\mathbb{I}(\mathsf{U}; \mathsf{Y}, \mathsf{U} \oplus \mathsf{X}) - \mathbb{I}(\mathsf{U}; \mathsf{Z}, \mathsf{U} \oplus \mathsf{X}) = \mathbb{H}(\mathsf{X}|\mathsf{Z}) - \mathbb{H}(\mathsf{X}|\mathsf{Y}).$$

Since the secret-key capacity is at least as large as the secrecy capacity of the conceptual WTC, we conclude that

$$C_{\mathrm{s}}^{\mathrm{SM}} \geqslant \mathbb{H}(\mathsf{X}|\mathsf{Z}) - \mathbb{H}(\mathsf{X}|\mathsf{Y}) = \mathbb{I}(\mathsf{X}; \mathsf{Y}) - \mathbb{I}(\mathsf{X}; \mathsf{Z}). \tag{4.8}$$

Because the public channel is two-way, we can reverse the roles of Alice and Bob and create another conceptual WTC from Bob to Alice. Following the same arguments as above, we obtain the second lower bound

$$C_{\mathrm{s}}^{\mathrm{SM}} \geqslant \mathbb{H}(\mathsf{Y}|\mathsf{Z}) - \mathbb{H}(\mathsf{Y}|\mathsf{X}) = \mathbb{I}(\mathsf{Y}; \mathsf{X}) - \mathbb{I}(\mathsf{Y}; \mathsf{Z}). \tag{4.9}$$

By combining (4.8) and (4.9) we obtain

$$\begin{aligned} C_{\mathrm{s}}^{\mathrm{SM}} &\geqslant \max(\mathbb{I}(\mathsf{X}; \mathsf{Y}) - \mathbb{I}(\mathsf{X}; \mathsf{Z}), \mathbb{I}(\mathsf{X}; \mathsf{Y}) - \mathbb{I}(\mathsf{Y}; \mathsf{Z})) \\ &= \mathbb{I}(\mathsf{X}; \mathsf{Y}) - \min(\mathbb{I}(\mathsf{X}; \mathsf{Z}), \mathbb{I}(\mathsf{Y}; \mathsf{Z})). \end{aligned}$$

4.2.2 Secret-key distillation based on Slepian–Wolf codes

In this section, we rederive the achievable rates given by Theorem 4.1 with a more constructive proof that is based on Slepian–Wolf codes. Although this alternative approach does not improve on the bounds already obtained, it is useful because it provides operational insight into the design of key-distillation strategies without relying on the existence of wiretap codes. The use of Slepian–Wolf codes makes the derivation slightly more involved than that in Section 4.2.1, but it bears some similarity to the achievability proof of Theorem 3.2, which is based on the notion of an enhanced channel.

In principle, the definition of key-distillation strategies allows many exchanges of messages over the public channel; however, to make the analysis tractable, we study simpler strategies. The first simplification consists of restricting our attention to strategies that exploit a *single* and *one-way* round of communication over the public channel and that do not rely on local randomness. If we assume that Alice is the one transmitting the public message, then such a strategy involves only

- a *single* encoding function $f : \mathcal{X}^n \to \mathcal{A}$ to compute the message a sent over the public channel;
- Alice's key-distillation function $\kappa_{\mathrm{a}} : \mathcal{X}^n \to \mathcal{K}$;
- Bob's key-distillation function $\kappa_{\mathrm{b}} : \mathcal{Y}^n \times \mathcal{A} \to \mathcal{K}$.

The second simplification consists of requiring Bob to decode Alice's observation x^n on the basis of his own observation y^n and the public message a instead of computing k directly. We will show that these simple strategies suffice to achieve the rates in Theorem 4.1 by means of a random-binning argument; however, before we do so, it is useful to develop an additional desirable property that these strategies should

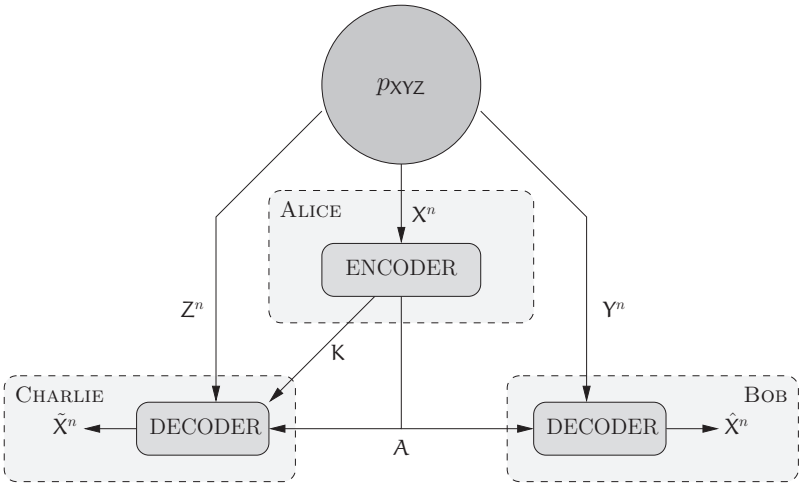

Figure 4.3 Enhanced source model for secret-key agreement.

possess. We expand the information rate $(1/n)\mathbf{L}(\mathcal{S}_n)$ leaked to the eavesdropper as follows:

$$\begin{aligned}\frac{1}{n}\mathbf{L}(\mathcal{S}_n) &= \frac{1}{n}\mathbb{I}(\mathsf{K}; Z^n \mathsf{A}|\mathcal{S}_n)\\ &= \frac{1}{n}\mathbb{I}(\mathsf{K}X^n; Z^n \mathsf{A}|\mathcal{S}_n) - \frac{1}{n}\mathbb{I}(X^n; Z^n \mathsf{A}|\mathsf{K}\mathcal{S}_n)\\ &= \frac{1}{n}\mathbb{I}(X^n; Z^n \mathsf{A}|\mathcal{S}_n) + \frac{1}{n}\mathbb{I}(\mathsf{K}; Z^n \mathsf{A}|X^n\mathcal{S}_n) - \frac{1}{n}\mathbb{I}(X^n; Z^n \mathsf{A}|\mathsf{K}\mathcal{S}_n)\\ &= \frac{1}{n}\mathbb{I}(X^n; Z^n \mathsf{A}|\mathcal{S}_n) - \frac{1}{n}\mathbb{I}(X^n; Z^n \mathsf{A}|\mathsf{K}\mathcal{S}_n),\end{aligned}$$

where the last inequality follows from $\mathbb{I}(\mathsf{K}; Z^n \mathsf{A}|X^n\mathcal{S}_n) = 0$ since K is a function of X^n. Note that, no matter how A is computed, the leakage is minimized if $(1/n)\mathbb{I}(X^n; Z^n \mathsf{A}|\mathsf{K}\mathcal{S}_n)$ is maximized, which happens if $(1/n)\mathbb{H}(X^n|Z^n \mathsf{A}\mathsf{K}\mathcal{S}_n)$ is small.

Therefore, we shall prove the existence of a sequence of $(2^{nR}, n)$ strategies $\{\mathcal{S}_n\}_{n\geqslant 1}$ with a single and one-way round of communication such that

$$\lim_{n\to\infty} \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) = 0, \qquad \lim_{n\to\infty} \frac{1}{n}\mathbb{H}(X^n|Z^n \mathsf{A}\mathsf{K}\mathcal{S}_n) = 0, \tag{4.10}$$

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{U}(\mathcal{S}_n) = 0, \qquad \lim_{n\to\infty} \frac{1}{n}\mathbf{L}(\mathcal{S}_n) = 0. \tag{4.11}$$

We now show that the two conditions in (4.10) can be combined into a single reliability constraint by considering the enhanced source model illustrated in Figure 4.3. This model enhances the original secret-key agreement problem by introducing a *virtual receiver*, hereafter named Charlie, who obtains the same observation Z^n as Eve, overhears the message A over the public channel, and has access to K through an error-free side channel.

Definition 4.6. *An* $(2^{nR}, 2^{nR_{\mathrm{p}}}, n)$ *strategy* $\mathcal{S}_n$ *for the enhanced source model consists of*

- *two index sets* $\mathcal{K} = [\![1, 2^{nR}]\!]$ *and* $\mathcal{A} = [\![1, 2^{nR_{\mathrm{p}}}]\!]$*;*
- *an encoding function* $f : \mathcal{X}^n \to \mathcal{A}$*;*
- *a key-distillation function* $\kappa_{\mathrm{a}} : \mathcal{X}^n \to \mathcal{K}$*;*
- *a decoding function* $g : \mathcal{Y}^n \times \mathcal{A} \to \mathcal{X}^n \cup \{\mathbf{?}\}$ *for Bob;*
- *a decoding function* $h : \mathcal{Z}^n \times \mathcal{A} \times \mathcal{K} \to \mathcal{X}^n \cup \{\mathbf{?}\}$ *for Charlie.*

Bob's key-distillation function is implicitly defined as $\kappa_{\mathrm{b}} \triangleq \kappa_{\mathrm{a}} \circ g$ if $g(y^n, a) \neq \mathbf{?}$. The reliability performance of a $(2^{nR}, 2^{nR_{\mathrm{p}}}, n)$ strategy $\mathcal{S}_n$ is measured in terms of its average probability of error

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \triangleq \mathbb{P}\left[\hat{\mathsf{X}}^n \neq \mathsf{X}^n \text{ or } \tilde{\mathsf{X}}^n \neq \mathsf{X}^n \middle| \mathcal{S}_n\right],$$

its secrecy performance is measured in terms of the leakage

$$\mathbf{L}(\mathcal{S}_n) \triangleq \mathbb{I}(\mathsf{K}; \mathsf{Z}^n\mathsf{A}|\mathcal{S}_n),$$

and the uniformity of keys is measured in terms of

$$\mathbf{U}(\mathcal{S}_n) \triangleq \log\lceil 2^{nR}\rceil - \mathbb{H}(\mathsf{K}|\mathcal{S}_n).$$

Note that a $(2^{nR}, 2^{nR_{\mathrm{p}}}, n)$ strategy $\mathcal{S}_n$ for the enhanced source model is a $(2^{nR}, n)$ secret-key distillation strategy for the original source model, which is subject to a more stringent reliability constraint and for which we are controlling explicitly the rate R_{p} of communication over the public channel. By construction, the probability of error for the original source model is at most $\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n)$ since

$$\mathbb{P}\left[\hat{\mathsf{X}}^n \neq \mathsf{X}^n \middle| \mathcal{S}_n\right] \leqslant \mathbb{P}\left[\hat{\mathsf{X}}^n \neq \mathsf{X}^n \text{ or } \tilde{\mathsf{X}}^n \neq \mathsf{X}^n \middle| \mathcal{S}_n\right] = \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n).$$

In addition, Fano's inequality guarantees that $(1/n)\mathbb{H}(\mathsf{X}^n|\mathsf{A}\mathsf{Z}^n\mathsf{K}\mathcal{S}_n) \leqslant \delta(\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n))$. Therefore, the two constraints in (4.10) are automatically satisfied if $\lim_{n\to\infty} \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) = 0$ for the enhanced source model.

We are now ready to develop a random-binning argument and show the existence of a sequence of $(2^{nR}, 2^{nR_{\mathrm{p}}}, n)$ strategies $\{\mathcal{S}_n\}_{n\geqslant 1}$ such that

$$\lim_{n\to\infty} \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) = 0, \quad \lim_{n\to\infty} \frac{1}{n}\mathbf{L}(\mathcal{S}_n) = 0, \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n}\mathbf{U}(\mathcal{S}_n) = 0$$

for some appropriate choice of R and R_{p}.

Without loss of generality, we assume that $\mathbb{H}(\mathsf{X}) > \mathbb{H}(\mathsf{X}|\mathsf{Z}) - \mathbb{H}(\mathsf{X}|\mathsf{Y}) > 0$. Let $\epsilon > 0$ and $n \in \mathbb{N}^*$. Let $R > 0$ and $R_{\mathrm{p}} > 0$ be rates to be specified later. We construct a $(2^{nR}, 2^{nR_{\mathrm{p}}}, n)$ strategy as follows.

- *Codebook construction.* For each sequence $x^n \in \mathcal{T}_\epsilon^n(\mathsf{X})$, draw two indices uniformly at random in the sets $[\![1, 2^{nR_{\mathrm{p}}}]\!]$ and $[\![1, 2^{nR}]\!]$; these index assignments define the functions $f : \mathcal{X}^n \to [\![1, 2^{nR_{\mathrm{p}}}]\!]$ and $\kappa_{\mathrm{a}} : \mathcal{X}^n \to [\![1, 2^{nR}]\!]$, which are revealed to all parties.
- *Alice's encoder.* Given an observation x^n, if $x^n \in \mathcal{T}_\epsilon^n(\mathsf{X})$, set $k = \kappa_{\mathrm{a}}(x^n)$ and $a = f(x^n)$; otherwise, set $k = 1$ and $a = 1$.

- *Bob's decoder.* Given an observation y^n, output $\hat{x}^n$ if it is the unique sequence such that $(\hat{x}^n, y^n) \in \mathcal{T}_\epsilon^n(\mathsf{XY})$ and $f(\hat{x}^n) = a$; otherwise, output an error **?**; if there is no error, distill a key $\hat{k} = \kappa_{\mathrm{a}}(\hat{x}^n)$.
- *Charlie's decoder.* Given an observation z^n, output $\tilde{x}^n$ if it is the unique sequence such that $(\tilde{x}^n, z^n) \in \mathcal{T}_\epsilon^n(\mathsf{XZ})$, $f(\tilde{x}^n) = a$ and $\kappa_{\mathrm{a}}(\tilde{x}^n) = k$; otherwise, output an error **?**.

The random variable that represents the strategy defined by the randomly chosen index assignments is denoted by S_n. We proceed to bound the quantities $\mathbb{E}[\mathbf{P}_{\mathrm{e}}(\mathsf{S}_n)]$, $\mathbb{E}[(1/n)\mathbf{L}(\mathsf{S}_n)]$, and $\mathbb{E}[(1/n)\mathbf{U}(\mathsf{S}_n)]$ separately.

The upper bound for $\mathbb{E}[\mathbf{P}_{\mathrm{e}}(\mathsf{S}_n)]$ is obtained with the approach used in Section 2.3.1 for the Slepian–Wolf theorem. $\mathbb{E}[\mathbf{P}_{\mathrm{e}}(\mathsf{S}_n)]$ can be expressed in terms of the events

$$\begin{aligned}
\mathcal{E}_0 &= \left\{\mathsf{X}^n \notin \mathcal{T}_\epsilon^n(\mathsf{X}) \text{ or } (\mathsf{X}^n, \mathsf{Y}^n) \notin \mathcal{T}_\epsilon^n(\mathsf{XY})\right\},\\
\mathcal{E}_1 &= \left\{\exists x^n \neq \mathsf{X}^n : f(x^n) = \mathsf{A} \text{ and } (x^n, \mathsf{Y}^n) \in \mathcal{T}_\epsilon^n(\mathsf{XY})\right\},\\
\mathcal{E}_2 &= \left\{\exists x^n \neq \mathsf{X}^n : \kappa_{\mathrm{a}}(x^n) = \mathsf{K},\, f(x^n) = \mathsf{A} \text{ and } (x^n, \mathsf{Z}^n) \in \mathcal{T}_\epsilon^n(\mathsf{XZ})\right\}
\end{aligned}$$

since $\mathbb{E}[\mathbf{P}_{\mathrm{e}}(\mathsf{S}_n)] = \mathbb{P}[\mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_2]$. By the union bound, we obtain

$$\mathbb{E}[\mathbf{P}_{\mathrm{e}}(\mathsf{S}_n)] \leqslant \mathbb{P}[\mathcal{E}_0] + \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2],$$

and, following exactly the same approach as that used in Section 2.3.1 to prove the Slepian–Wolf theorem, we can show that, if

$$R_{\mathrm{p}} > \mathbb{H}(\mathsf{X}|\mathsf{Y}) + \delta(\epsilon) \quad \text{and} \quad R + R_{\mathrm{p}} > \mathbb{H}(\mathsf{X}|\mathsf{Z}) + \delta(\epsilon), \tag{4.12}$$

then $\mathbb{E}[\mathbf{P}_{\mathrm{e}}(\mathsf{S}_n)] \leqslant \delta_\epsilon(n)$.

Next, we develop an upper bound for $\mathbb{E}[(1/n)\mathbf{L}(\mathsf{S}_n)]$. We expand $\mathbb{E}[(1/n)\mathbf{L}(\mathsf{S}_n)]$ as

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\mathbf{L}(\mathsf{S}_n)\right] &= \frac{1}{n}\mathbb{I}(\mathsf{K}; \mathsf{A}\mathsf{Z}^n|\mathsf{S}_n)\\
&= \frac{1}{n}\mathbb{H}(\mathsf{K}|\mathsf{S}_n) + \frac{1}{n}\mathbb{H}(\mathsf{A}\mathsf{Z}^n|\mathsf{S}_n) - \frac{1}{n}\mathbb{H}(\mathsf{A}\mathsf{K}\mathsf{Z}^n|\mathsf{S}_n)\\
&= \frac{1}{n}\mathbb{H}(\mathsf{K}|\mathsf{S}_n) + \frac{1}{n}\mathbb{H}(\mathsf{A}|\mathsf{S}_n) + \frac{1}{n}\mathbb{H}(\mathsf{Z}^n|\mathsf{A}\mathsf{S}_n) - \frac{1}{n}\mathbb{H}(\mathsf{A}\mathsf{K}\mathsf{Z}^n|\mathsf{S}_n)\\
&\leqslant \frac{1}{n}\mathbb{H}(\mathsf{K}|\mathsf{S}_n) + \frac{1}{n}\mathbb{H}(\mathsf{A}|\mathsf{S}_n) + \mathbb{H}(\mathsf{Z}) - \frac{1}{n}\mathbb{H}(\mathsf{A}\mathsf{K}\mathsf{Z}^n|\mathsf{S}_n),
\end{aligned} \tag{4.13}$$

where the last inequality follows from $\mathbb{H}(\mathsf{Z}^n|\mathsf{A}\mathsf{S}_n) \leqslant \mathbb{H}(\mathsf{Z}^n) = n\mathbb{H}(\mathsf{Z})$. We bound the remaining terms on the right-hand side separately. By construction of a $(2^{nR}, 2^{nR_{\mathrm{p}}}, n)$ strategy,

$$\frac{1}{n}\mathbb{H}(\mathsf{K}|\mathsf{S}_n) \leqslant R + \delta(n), \quad \frac{1}{n}\mathbb{H}(\mathsf{A}|\mathsf{S}_n) \leqslant R_{\mathrm{p}} + \delta(n). \tag{4.14}$$

In addition,

$$\begin{aligned}
\frac{1}{n}\mathbb{H}(\mathsf{A}\mathsf{K}\mathsf{Z}^n|\mathsf{S}_n) &= \frac{1}{n}\mathbb{H}(\mathsf{A}\mathsf{K}\mathsf{Z}^n\mathsf{X}^n|\mathsf{S}_n) - \frac{1}{n}\mathbb{H}(\mathsf{X}^n|\mathsf{A}\mathsf{K}\mathsf{Z}^n\mathsf{S}_n)\\
&= \frac{1}{n}\mathbb{H}(\mathsf{Z}^n\mathsf{X}^n|\mathsf{S}_n) + \frac{1}{n}\mathbb{H}(\mathsf{A}\mathsf{K}|\mathsf{X}^n\mathsf{Z}^n\mathsf{S}_n) - \frac{1}{n}\mathbb{H}(\mathsf{X}^n|\mathsf{A}\mathsf{K}\mathsf{Z}^n\mathsf{S}_n)\\
&= \frac{1}{n}\mathbb{H}(\mathsf{Z}^n\mathsf{X}^n|\mathsf{S}_n) - \frac{1}{n}\mathbb{H}(\mathsf{X}^n|\mathsf{A}\mathsf{K}\mathsf{Z}^n\mathsf{S}_n),
\end{aligned}$$

where the last equality follows from $\mathbb{H}(\mathsf{AK}|Z^nX^nS_n) = 0$ because A and K are functions of X^n. By virtue of Fano's inequality,

$$\begin{aligned}\frac{1}{n}\mathbb{H}(X^n|\mathsf{AK}Z^nS_n) &= \sum_{\mathcal{S}_n} p_{S_n}(\mathcal{S}_n)\,\frac{1}{n}\mathbb{H}(X^n|\mathsf{AK}Z^n\mathcal{S}_n)\\ &\leqslant \sum_{\mathcal{S}_n} p_{S_n}(\mathcal{S}_n)\left(\frac{1}{n} + \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n)\log|\mathcal{X}|\right)\\ &= \delta(n) + \mathbb{E}[\mathbf{P}_{\mathrm{e}}(S_n)]\log|\mathcal{X}|\\ &= \delta_\epsilon(n).\end{aligned}$$

Therefore,

$$\frac{1}{n}\mathbb{H}(\mathsf{AK}Z^n|S_n) \geqslant \frac{1}{n}\mathbb{H}(Z^nX^n|S_n) - \delta_\epsilon(n) = \mathbb{H}(ZX) - \delta(\epsilon). \tag{4.15}$$

On substituting (4.14) and (4.15) into (4.13), we obtain

$$\mathbb{E}\left[\frac{1}{n}\mathbf{L}(S_n)\right] \leqslant R + R_{\mathrm{p}} + \mathbb{H}(Z) - \mathbb{H}(ZX) + \delta_\epsilon(n),$$

for any R and R_{p} satisfying (4.12). In particular, the choice

$$R_{\mathrm{p}} = \mathbb{H}(X|Y) + \delta(\epsilon) \quad \text{and} \quad R = \mathbb{H}(X|Z) - \mathbb{H}(X|Y) + \delta(\epsilon) \tag{4.16}$$

is compatible with the constraints (4.12) and yields

$$\mathbb{E}\left[\frac{1}{n}\mathbf{L}(S_n)\right] = \delta(\epsilon) + \delta_\epsilon(n).$$

Finally, we develop an upper bound for $\mathbb{E}[(1/n)\mathbf{U}(S_n)]$ by establishing a lower bound for $(1/n)\mathbb{H}(\mathsf{K}|S_n)$. Let us introduce the random variable Ξ, such that

$$\Xi \triangleq \begin{cases} 1 & \text{if } X^n \in \mathcal{T}_\epsilon^n(X),\\ 0 & \text{otherwise.}\end{cases}$$

By construction, Ξ is independent of S_n and, by the AEP, $\mathbb{P}[\Xi = 1] \geqslant 1 - \delta_\epsilon(n)$. Next, notice that,

$$\begin{aligned}\frac{1}{n}\mathbb{H}(\mathsf{K}|S_n) &\geqslant \frac{1}{n}\mathbb{H}(\mathsf{K}|S_n\,\Xi)\\ &\geqslant \mathbb{P}[\Xi = 1]\frac{1}{n}\mathbb{H}(\mathsf{K}|S_n,\,\Xi = 1)\\ &\geqslant (1 - \delta_\epsilon(n))\frac{1}{n}\mathbb{H}(\mathsf{K}|S_n,\,\Xi = 1).\end{aligned} \tag{4.17}$$

For a specific strategy $\mathcal{S}_n$, define $\mathsf{K}_{\mathcal{S}_n}$ as the random variable with distribution

$$p_{\mathsf{K}_{\mathcal{S}_n}} \triangleq p_{\mathsf{K}|S_n=\mathcal{S}_n,\,\Xi=1}.$$

Then $\mathbb{H}(\mathsf{K}|S_n = \mathcal{S}_n,\,\Xi = 1) = \mathbb{H}\big(\mathsf{K}_{\mathcal{S}_n}\big)$, which can be written explicitly in terms of the probability $p_{\mathsf{K}_{\mathcal{S}_n}}$ as

$$\frac{1}{n}\mathbb{H}\big(\mathsf{K}_{\mathcal{S}_n}\big) = -\frac{1}{n}\sum_{k\in\mathcal{K}} p_{\mathsf{K}_{\mathcal{S}_n}}(k)\log p_{\mathsf{K}_{\mathcal{S}_n}}(k).$$

By virtue of the symmetry of the random-binning construction, the quantity $\mathbb{E}_{S_n}\left[-p_{K_{S_n}}(k)\log p_{K_{S_n}}(k)\right]$ is independent of k; therefore,

$$\mathbb{E}_{S_n}\left[\frac{1}{n}\mathbb{H}\left(K_{S_n}\right)\right] = \frac{1}{n}\sum_{k\in\mathcal{K}}\mathbb{E}_S\left[-p_{K_{S_n}}(k)\log p_{K_{S_n}}(k)\right]$$
$$= \frac{\lceil 2^{nR}\rceil}{n}\mathbb{E}_S\left[-p_{K_{S_n}}(1)\log p_{K_{S_n}}(1)\right]. \tag{4.18}$$

Intuitively, because we use a random binning in which keys are assigned uniformly at random, we expect $p_{K_{S_n}}(1)$ to be on the order of 2^{-nR} for most strategies $\mathcal{S}_n$. This idea is formalized in the following lemma, whose proof is relegated to the appendix to this chapter.

Lemma 4.1. *Let χ be a function of a key-distillation strategy $\mathcal{S}_n$ defined as*

$$\chi(\mathcal{S}_n) = \begin{cases} 1 & \text{if } \left|p_{K_{S_n}}(1) - 2^{-nR}\right| \leqslant \epsilon 2^{-nR}, \\ 0 & \text{otherwise.} \end{cases}$$

If $R < \mathbb{H}(X) - \delta(\epsilon)$ then $\mathbb{P}_{S_n}[\chi(S_n) = 1] \geqslant 1 - \delta_\epsilon(n)$.

Using Lemma 4.1 with R as defined by (4.16), we then bound $\mathbb{E}_{S_n}\left[-p_{K_{S_n}}(1)\log p_{K_{S_n}}(1)\right]$ as

$$\mathbb{E}_{S_n}\left[-p_{K_{S_n}}(1)\log p_{K_{S_n}}(1)\right] \geqslant \mathbb{E}_{S_n}\left[-p_{K_{S_n}}(1)\log p_{K_{S_n}}(1)|\chi(S_n)=1\right]\mathbb{P}_{S_n}[\chi(S_n)=1]$$
$$\geqslant (1-\delta_\epsilon(n))(1-\epsilon)2^{-nR}\log\left(\frac{2^{nR}}{1+\epsilon}\right). \tag{4.19}$$

On combining (4.17), (4.18), and (4.19), we obtain

$$\frac{1}{n}\mathbb{H}(K|S_n) \geqslant R - \delta(\epsilon),$$

and, therefore,

$$\mathbb{E}\left[\frac{1}{n}\mathbf{U}(S_n)\right] \leqslant \delta_\epsilon(n) + \delta(\epsilon).$$

By applying the selection lemma to the random variable S_n and the functions $\mathbf{P}_e$, $\mathbf{L}$, and $\mathbf{U}$, we conclude that there exists a specific strategy $\mathcal{S}_n$ with rate R given by (4.16) and such that

$$\mathbf{P}_e(\mathcal{S}_n) \leqslant \delta_\epsilon(n), \quad \frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \delta_\epsilon(n) + \delta(\epsilon), \quad \text{and} \quad \frac{1}{n}\mathbf{U}(\mathcal{S}_n) \leqslant \delta_\epsilon(n) + \delta(\epsilon).$$

Hence, there exists a sequence of $(2^{nR}, n)$ key-distillation strategies $\{\mathcal{S}_n\}_{n\geqslant 1}$ with rate

$$R = \mathbb{H}(X|Z) - \mathbb{H}(X|Y) + \delta(\epsilon) = \mathbb{I}(X;Y) - \mathbb{I}(X;Z) + \delta(\epsilon)$$

and such that

$$\lim_{n\to\infty}\mathbf{P}_e(\mathcal{S}_n) = 0, \quad \lim_{n\to\infty}\frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \delta(\epsilon), \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{n}\mathbf{U}(\mathcal{S}_n) \leqslant \delta(\epsilon).$$

Since ϵ can be chosen arbitrarily small, we conclude that the secret-key capacity is at least $\mathbb{I}(X;Y) - \mathbb{I}(X;Z)$.

Note that, if $\mathbb{I}(X;Y) - \mathbb{I}(Y;Z) > 0$, we can reproduce the proof by swapping the roles of Alice and Bob and swapping X and Y in the equations. Therefore, the secret-key capacity is also at least $\mathbb{I}(Y;X) - \mathbb{I}(Y;Z)$.

Remark 4.5. *The existence of a public channel of unlimited capacity in the model is convenient because it allows us to focus solely on secrecy without having to account for the cost of communication and processing; however, this approach has a subtle drawback. In the real world, there exists no channel of unlimited capacity, and the public channel used in the model would be obtained through multiple uses of a side channel with finite capacity. Consequently, if a key-distillation strategy requires many rounds of communication over the public channel, the side channel would have to be used many times. Hence, the* effective *secret-key rate, obtained by normalizing the key size by the number of random realizations of the sources plus the number of uses of the side channel, may be much lower than what is predicted by the results obtained thus far. This motivates the study of key-distillation strategies with* rate-limited *public communication, that is, key-distillation strategies for which the messages exchanged over the public channel are subject to a rate constraint R_{p}. The construction of strategies based on Slepian–Wolf codes described above does not allow us to control precisely the rate of messages sent over the public channel. Actually, in* (4.16)*, we implicitly* required *the public rate R_{p} to be at least $\mathbb{H}(X|Y)$ (or $\mathbb{H}(Y|X)$ if the roles of Alice and Bob are swapped). The key idea to handle rate-limited public communication is to construct a Slepian–Wolf-based strategy that operates on a* quantized *version of X^n instead of on X^n directly. The quantization allows us to control the rate of public communication and to adjust it so that it falls below the rate constraint R_{p}. Specifically, one can combine the strategy described above with a Wyner–Ziv compression scheme, which can be thought of as a special case of vector quantization. We refer the reader to the bibliographical notes for further references.*

4.2.3 Upper bound for secret-key capacity

In this section, we show that $C_{\mathrm{s}}^{\mathrm{SM}} \leqslant \min(\mathbb{I}(X;Y), \mathbb{I}(X;Y|Z))$ with a converse argument. Let R be an achievable weak secret-key rate and let $\epsilon > 0$. For n sufficiently large, there exists a $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ such that

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \leqslant \epsilon, \quad \frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \epsilon, \quad \text{and} \quad \frac{1}{n}\mathbf{U}(\mathcal{S}_n) \leqslant \epsilon.$$

For clarity, we drop the conditioning on the strategy $\mathcal{S}_n$ in all subsequent calculations. By virtue of Fano's inequality, we have

$$\frac{1}{n}\mathbb{H}\left(K|\hat{K}A^r B^r Z^n\right) \leqslant \frac{1}{n}\mathbb{H}\left(K|\hat{K}\right) \leqslant \delta(\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n)) \leqslant \delta(\epsilon).$$

First, we show that $R \leqslant \mathbb{I}(\mathsf{X};\mathsf{Y}|\mathsf{Z}) + \delta(\epsilon)$. Note that

$$\begin{aligned}
R &\leqslant \frac{1}{n}\log\lceil 2^{nR}\rceil \\
&= \frac{1}{n}\mathbb{H}(\mathsf{K}) + \frac{1}{n}\mathbf{U}(\mathcal{S}_n) \\
&\leqslant \frac{1}{n}\mathbb{H}(\mathsf{K}) + \epsilon \\
&= \frac{1}{n}\mathbb{H}(\mathsf{K}|A^r\mathsf{B}^r\mathsf{Z}^n) + \frac{1}{n}\mathbf{L}(\mathcal{S}_n) + \epsilon \\
&\leqslant \frac{1}{n}\mathbb{H}(\mathsf{K}|A^r\mathsf{B}^r\mathsf{Z}^n) + \delta(\epsilon) \\
&= \frac{1}{n}\mathbb{I}\left(\mathsf{K};\hat{\mathsf{K}}|A^r\mathsf{B}^r\mathsf{Z}^n\right) + \frac{1}{n}\mathbb{H}\left(\mathsf{K}|\hat{\mathsf{K}}A^r\mathsf{B}^r\mathsf{Z}^n\right) + \delta(\epsilon) \\
&\leqslant \frac{1}{n}\mathbb{I}\left(\mathsf{K};\hat{\mathsf{K}}|A^r\mathsf{B}^r\mathsf{Z}^n\right) + \delta(\epsilon) \\
&\leqslant \frac{1}{n}\mathbb{I}(\mathsf{X}^n\mathsf{R}_{\mathsf{X}};\mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}|A^r\mathsf{B}^r\mathsf{Z}^n) + \delta(\epsilon),
\end{aligned} \tag{4.20}$$

where the last inequality follows from the data-processing inequality applied to the Markov chain $\mathsf{K} \to \mathsf{X}^n\mathsf{R}_{\mathsf{X}}\mathsf{B}^r \to \mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}A^r \to \hat{\mathsf{K}}$. We upper bound $(1/n)\mathbb{I}(\mathsf{X}^n\mathsf{R}_{\mathsf{X}};\mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}|A^r\mathsf{B}^r\mathsf{Z}^n)$ by using the following lemma.

Lemma 4.2. *Let $r \in \mathbb{N}^*$. Let $\mathsf{S} \in \mathcal{S}$, $\mathsf{T} \in \mathcal{T}$, $\mathsf{U} \in \mathcal{U}$, $\mathsf{V}^r \in \mathcal{V}^r$, and $\mathsf{W}^r \in \mathcal{W}^r$ be random vectors such that*

$$\forall i \in [\![1, r]\!] \quad \begin{cases} \mathsf{V}_i \text{ is a function of } \mathsf{S} \text{ and } \mathsf{W}^{i-1}, \\ \mathsf{W}_i \text{ is a function of } \mathsf{T} \text{ and } \mathsf{V}^{i-1}. \end{cases}$$

Then, $\mathbb{I}(\mathsf{S};\mathsf{T}|\mathsf{V}^r\mathsf{W}^r\mathsf{U}) \leqslant \mathbb{I}(\mathsf{S};\mathsf{T}|\mathsf{U})$.

Proof. We upper bound $\mathbb{I}(\mathsf{S};\mathsf{T}|\mathsf{V}^r\mathsf{W}^r\mathsf{U})$ as follows:

$$\begin{aligned}
\mathbb{I}(\mathsf{S};\mathsf{T}|\mathsf{V}^r\mathsf{W}^r\mathsf{U}) &\leqslant \mathbb{I}\big(\mathsf{S}\mathsf{V}_r;\mathsf{T}|\mathsf{V}^{r-1}\mathsf{W}^r\mathsf{U}\big) \\
&\leqslant \mathbb{I}\big(\mathsf{S}\mathsf{V}_r;\mathsf{T}\mathsf{W}_r|\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big) \\
&= \mathbb{I}\big(\mathsf{S};\mathsf{T}\mathsf{W}_r|\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big) + \mathbb{I}\big(\mathsf{V}_r;\mathsf{T}\mathsf{W}_r|\mathsf{S}\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big) \\
&= \mathbb{I}\big(\mathsf{S};\mathsf{T}|\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big) + \mathbb{I}\big(\mathsf{S};\mathsf{W}_r|\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{T}\mathsf{U}\big) \\
&\quad + \mathbb{I}\big(\mathsf{V}_r;\mathsf{T}\mathsf{W}_r|\mathsf{S}\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big).
\end{aligned} \tag{4.21}$$

Since V_r is a function of S and W^{r-1} and W_r is a function of T and V^{r-1}, note that

$$\mathbb{I}\big(\mathsf{V}_r;\mathsf{T}\mathsf{W}_r|\mathsf{S}\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big) = 0 \quad \text{and} \quad \mathbb{I}\big(\mathsf{S};\mathsf{W}_r|\mathsf{T}\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big) = 0. \tag{4.22}$$

On substituting (4.22) into (4.21), we obtain

$$\mathbb{I}(\mathsf{S};\mathsf{T}|\mathsf{V}^r\mathsf{W}^r\mathsf{U}) \leqslant \mathbb{I}\big(\mathsf{S};\mathsf{T}|\mathsf{V}^{r-1}\mathsf{W}^{r-1}\mathsf{U}\big).$$

By induction over r, we conclude that $\mathbb{I}(\mathsf{S};\mathsf{T}|\mathsf{V}^r\mathsf{W}^r\mathsf{U}) \leqslant \mathbb{I}(\mathsf{S};\mathsf{T}|\mathsf{U})$. □

By using Lemma 4.2 with $S \triangleq X^n R_X$, $T \triangleq Y^n R_Y$, $U \triangleq Z^n$, $V^r \triangleq A^r$, and $W^r \triangleq B^r$, we obtain

$$\mathbb{I}(X^n R_X; Y^n R_Y | A^r B^r Z^n) \leqslant \mathbb{I}(X^n R_X; Y^n R_Y | Z^n).$$

Since R_X and R_Y are independent of the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$, we have

$$\mathbb{I}(X^n R_X; Y^n R_Y | Z^n) = \mathbb{I}(X^n; Y^n | Z^n) = n\mathbb{I}(X; Y|Z);$$

therefore,

$$\mathbb{I}(X^n R_X; Y^n R_Y | A^r B^r Z^n) \leqslant n\mathbb{I}(X; Y|Z). \tag{4.23}$$

On substituting (4.23) into (4.20), we obtain

$$R \leqslant \mathbb{I}(X; Y|Z) + \delta(\epsilon).$$

Finally, we show that $R \leqslant \mathbb{I}(X; Y) + \delta(\epsilon)$. Note that, by assumption,

$$\frac{1}{n}\mathbb{I}(K; A^r B^r) \leqslant \frac{1}{n}\mathbb{I}(K; A^r B^r Z^n) = \frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \delta(\epsilon).$$

Therefore, all the steps leading to (4.20) and (4.23) can be reapplied without conditioning on the variable Z^n, which yields $R \leqslant \mathbb{I}(X; Y) + \delta(\epsilon)$. Since $\epsilon > 0$ can be chosen arbitrarily small, we obtain the desired upper bound

$$C_s^{SM} \leqslant \min(\mathbb{I}(X; Y), \mathbb{I}(X; Y|Z)).$$

4.2.4 Alternative upper bounds for secret-key capacity

In general, the upper bound $C_s^{SM} \leqslant \mathbb{I}(X; Y|Z)$ established in Theorem 4.2 is loose, but the cause of this looseness is buried in the technical details of the proof; hence, it is worth developing an intuitive understanding of the bound before we try to improve it.

We start by showing that, for any source model, the quantity $\mathbb{I}(X; Y|Z)$ admits an operational interpretation: it is the secret-key capacity obtained by providing Bob with an explicit advantage over Eve. In fact, consider a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$ and secret-key capacity $C_s^{SM}(p_{XYZ})$. Assume we provide an advantage to Bob by giving him access to Eve's observation Z, which creates a new source model in which Bob observes $Y' = (YZ)$ instead of Y. Since a key-distillation strategy for the original source model remains a key-distillation strategy for the new source model, it holds that $C_s^{SM}(p_{XY'Z}) \geqslant C_s^{SM}(p_{XYZ})$; however, because $X \to Y' \to Z$ forms a Markov chain, Corollary 4.1 applies and

$$C_s^{SM}(p_{XY'Z}) = \mathbb{I}\big(X; Y'|Z\big) = \mathbb{I}(X; Y|Z).$$

On the basis of this operational interpretation, a natural approach to improve the bound $C_s^{SM}(p_{XYZ}) \leqslant \mathbb{I}(X; Y|Z)$ is to reduce the advantage given to Bob.

A first possibility to mitigate Bob's advantage is to analyze more precisely how Eve could further process her observations. Specifically, consider a key-distillation

strategy $\mathcal{S}_n$ for a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$, with key K and public messages $\mathsf{A}^r\mathsf{B}^r$. We allow Eve to send her observations Z^n through an arbitrary DMC $\big(\mathcal{Z}, p_{\bar{\mathsf{Z}}|\mathsf{Z}}, \bar{\mathcal{Z}}\big)$, which results in a new source model with DMS $\big(\mathcal{X}\mathcal{Y}\bar{\mathcal{Z}}, p_{\mathsf{XY}\bar{\mathsf{Z}}}\big)$. Because $\mathsf{K}\mathsf{A}^r\mathsf{B}^r \to \mathsf{Z}^n \to \bar{\mathsf{Z}}^n$ forms a Markov chain, the data-processing inequality ensures that

$$\frac{1}{n}\mathbb{I}\big(\mathsf{K}; \mathsf{A}^r\mathsf{B}^r\bar{\mathsf{Z}}^n|\mathcal{S}_n\big) \leqslant \frac{1}{n}\mathbb{I}(\mathsf{K}; \mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n|\mathcal{S}_n) = \frac{1}{n}\mathbf{L}(\mathcal{S}_n).$$

Hence, we have that K is also a secret key for the new source model and $C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{XYZ}}) \leqslant C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{XY}\bar{\mathsf{Z}}})$. By virtue of Theorem 4.1, we also have $C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{XY}\bar{\mathsf{Z}}}) \leqslant \mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big)$. Since the DMC $\big(\mathcal{Z}, p_{\bar{\mathsf{Z}}|\mathsf{Z}}, \bar{\mathcal{Z}}\big)$ is arbitrary, it must hold that

$$C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{XYZ}}) \leqslant \inf_{p_{\bar{\mathsf{Z}}|\mathsf{Z}}} \mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big).$$

This inequality motivates the definition of a new measure of information.

Definition 4.7. *For a DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$, the intrinsic conditional information between X and Y given Z is*

$$\mathbb{I}(\mathsf{X}; \mathsf{Y} \downarrow \mathsf{Z}) \triangleq \inf_{p_{\bar{\mathsf{Z}}|\mathsf{Z}}} \mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big).$$

Intuitively, the intrinsic conditional information measures the information between X and Y that remains after Eve has chosen the best memoryless processing of the observation Z. The following theorem shows that the intrinsic conditional information is an upper bound for the secret-key capacity that is at least as tight as Theorem 4.1.

Theorem 4.2 (Maurer)**.** *The secret-key capacity of a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$ satisfies*

$$C_{\mathrm{s}}^{\mathsf{SM}} \leqslant \mathbb{I}(\mathsf{X}; \mathsf{Y} \downarrow \mathsf{Z}) \leqslant \min(\mathbb{I}(\mathsf{X}; \mathsf{Y}), \mathbb{I}(\mathsf{X}; \mathsf{Y}|\mathsf{Z})).$$

Proof. The inequality $C_{\mathrm{s}}^{\mathsf{SM}} \leqslant \inf_{p_{\bar{\mathsf{Z}}|\mathsf{Z}}} \mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big)$ has already been proved in the paragraphs above. To establish the second inequality, note that $\mathbb{I}(\mathsf{X}; \mathsf{Y} \downarrow \mathsf{Z}) \leqslant \mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big)$ for any choice of transition probabilities $p_{\bar{\mathsf{Z}}|\mathsf{Z}}$; therefore, we prove that $\mathbb{I}(\mathsf{X}; \mathsf{Y} \downarrow \mathsf{Z}) \leqslant \min(\mathbb{I}(\mathsf{X}; \mathsf{Y}), \mathbb{I}(\mathsf{X}; \mathsf{Y}|\mathsf{Z}))$ by constructing specific DMCs $(\mathcal{Z}, p_{\bar{\mathsf{Z}}|\mathsf{Z}}, \bar{\mathcal{Z}})$ for which $\mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big)$ takes the values $\mathbb{I}(\mathsf{X}; \mathsf{Y})$ and $\mathbb{I}(\mathsf{X}; \mathsf{Y}|\mathsf{Z})$:

- if we set $\bar{\mathcal{Z}} = \mathcal{Z}$ and $p_{\bar{\mathsf{Z}}|\mathsf{Z}}(\bar{z}|z) = 1/|\mathcal{Z}|$ for all $\bar{z}$, then $\bar{\mathsf{Z}}$ is independent of X, Y, and Z so that $\mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big) = \mathbb{I}(\mathsf{X}; \mathsf{Y})$;
- if we set $p_{\bar{\mathsf{Z}}|\mathsf{Z}}(\bar{z}|z) = \mathbb{1}(\bar{z} - z)$, then $\mathsf{Z} = \bar{\mathsf{Z}}$ with probability one and $\mathbb{I}\big(\mathsf{X}; \mathsf{Y}|\bar{\mathsf{Z}}\big) = \mathbb{I}(\mathsf{X}; \mathsf{Y}|\mathsf{Z})$.

Therefore,

$$\mathbb{I}(\mathsf{X}; \mathsf{Y} \downarrow \mathsf{Z}) \leqslant \min(\mathbb{I}(\mathsf{X}; \mathsf{Y}), \mathbb{I}(\mathsf{X}; \mathsf{Y}|\mathsf{Z})). \qquad \square$$

The following example shows that Theorem 4.2 is useful because $\mathbb{I}(\mathsf{X}; \mathsf{Y} \downarrow \mathsf{Z})$ can be strictly tighter than $\mathbb{I}(\mathsf{X}; \mathsf{Y}|\mathsf{Z})$.

Example 4.1. Let $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3\}$ and let p_{XY} be defined in the table below.

X \ Y	0	1	2	3
0	1/8	1/8	0	0
1	1/8	1/8	0	0
2	0	0	1/4	0
3	0	0	0	1/4

Define Z as

$$Z \triangleq \begin{cases} X \oplus Y & \text{if} \quad X \in \{0, 1\}, \\ X & \text{if} \quad X \in \{2, 3\}. \end{cases}$$

One can verify that $\mathbb{I}(X;Y) = \frac{3}{2}$ and $\mathbb{I}(X;Y|Z) = \frac{1}{2}$. The particularity of this DMS is that, for $X \in \{0, 1\}$, individual knowledge of Z or Y does not resolve any uncertainty about X but joint knowledge of Z and Y determines X completely. Consider now the channel $\left(\mathcal{Z}, p_{\bar{Z}|Z}, \bar{\mathcal{Z}}\right)$, such that

$$p_{\bar{Z}|Z}(0|0) = p_{\bar{Z}|Z}(0|1) = p_{\bar{Z}|Z}(1|0) = p_{\bar{Z}|Z}(1|1) = \frac{1}{2},$$

$$p_{\bar{Z}|Z}(2|2) = p_{\bar{Z}|Z}(3|3) = 1.$$

Notice that, if $Z \in \{0, 1\}$, then $\bar{Z}$ is obtained by sending Z through a BSC with cross-over probability $\frac{1}{2}$ and, therefore, $\bar{Z}$ becomes independent of Z. As a result, for $X \in \{0, 1\}$, knowledge of $\bar{Z}$ still does not resolve any uncertainty about X, but knowledge of both $\bar{Z}$ and Y does not resolve any uncertainty, either. Hence, $\mathbb{I}\left(X;Y|\bar{Z}\right) = 0$ and, consequently, $\mathbb{I}(X;Y \downarrow Z) = 0$ and $C_s^{SM} = 0$.

A second possibility to improve the result of Theorem 4.2 is not only to analyze how Eve could further process her observations but also to provide her with some side information represented by a correlated DMS $(\mathcal{U}, p_U)$; however, the side information must be introduced carefully if we want to retain an upper bound for $C_s^{SM}(p_{XYZ})$ because, in general, the secret-key capacity is reduced if Eve has access to side information.

Proposition 4.1. *Consider a source model with DMS* $\left(\mathcal{X}\mathcal{Y}\tilde{\mathcal{Z}}, p_{XY\tilde{Z}}\right)$*, in which the eavesdropper has access to the observations* $\tilde{Z} = (Z, U) \in \mathcal{Z} \times \mathcal{U}$*. Then,*

$$C_s^{SM}(p_{XY\tilde{Z}}) \geqslant C_s^{SM}(p_{XYZ}) - \mathbb{H}(U).$$

Proof. We prove the inequality by constructing a key-distillation strategy for the DMS $\left(\mathcal{X}\mathcal{Y}\tilde{\mathcal{Z}}, p_{XY\tilde{Z}}\right)$ from a key-distillation strategy for the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$.

Let R be an achievable weak secret-key rate for the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$. For any $\epsilon > 0$, there exists a $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$, such that

$$\mathbf{P}_e(\mathcal{S}_n) \leqslant \delta(\epsilon), \quad \frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \delta(\epsilon), \quad \text{and} \quad \frac{1}{n}\mathbf{U}(\mathcal{S}_n) \leqslant \delta(\epsilon). \tag{4.24}$$

Using Fano's inequality, we obtain

$$\frac{1}{n}\mathbb{H}\Big(\mathsf{K}|\hat{\mathsf{K}}\mathcal{S}_n\Big) \leqslant \delta(\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n)) \leqslant \delta(\epsilon). \tag{4.25}$$

In addition,

$$\begin{aligned}\frac{1}{n}\mathbb{I}(\mathsf{K};\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n\mathsf{U}^n|\mathcal{S}_n) &= \frac{1}{n}\mathbb{I}(\mathsf{K};\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n|\mathcal{S}_n) + \frac{1}{n}\mathbb{I}(\mathsf{K};\mathsf{U}^n|\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n\mathcal{S}_n)\\ &\leqslant \frac{1}{n}\mathbb{I}(\mathsf{K};\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n|\mathcal{S}_n) + \frac{1}{n}\mathbb{H}(\mathsf{U}^n)\\ &\leqslant \frac{1}{n}\mathbf{L}(\mathcal{S}_n) + \mathbb{H}(\mathsf{U})\\ &\leqslant \delta(\epsilon) + \mathbb{H}(\mathsf{U}).\end{aligned} \tag{4.26}$$

We now turn our attention back to the DMS $\big(\mathcal{X}\mathcal{Y}\tilde{\mathcal{Z}}, p_{\mathsf{X}\mathsf{Y}\tilde{\mathsf{Z}}}\big)$. We construct a key-distillation strategy by first running m *independent* repetitions of the key-distillation strategy $\mathcal{S}_n$, from which Alice obtains i.i.d. sequences K^m, Bob obtains i.i.d. sequences $\hat{\mathsf{K}}^m$, and Eve observes A^{rm}, B^{rm}, Z^{nm}, and U^{nm}. Effectively, this creates a source model with DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$ in which $\mathsf{X}' \triangleq \mathsf{K}$, $\mathsf{Y}' \triangleq \hat{\mathsf{K}}$, and $\mathsf{Z}' \triangleq \mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n\mathsf{U}^n$. Since a key-distillation strategy for the DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$ is a specific instance of key-distillation strategy for the DMS $\big(\mathcal{X}\mathcal{Y}\tilde{\mathcal{Z}}, p_{\mathsf{X}\mathsf{Y}\tilde{\mathsf{Z}}}\big)$, we have

$$C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{X}\mathsf{Y}\tilde{\mathsf{Z}}}) \geqslant \frac{1}{n}C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'}), \tag{4.27}$$

where the normalization by n appears because each realization of the DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$ relies on n realizations of the DMS $\big(\mathcal{X}\mathcal{Y}\tilde{\mathcal{Z}}, p_{\mathsf{X}\mathsf{Y}\tilde{\mathsf{Z}}}\big)$. Note that Theorem 4.1 guarantees that

$$C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'}) \geqslant \mathbb{I}\big(\mathsf{X}';\mathsf{Y}'\big) - \mathbb{I}\big(\mathsf{X}';\mathsf{Z}'\big).$$

Next, we use (4.24), (4.25), and (4.26) to lower bound $\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}')$ as

$$\begin{aligned}\mathbb{I}\big(\mathsf{X}';\mathsf{Y}'\big) - \mathbb{I}\big(\mathsf{X}';\mathsf{Z}'\big) &= \mathbb{I}\Big(\mathsf{K};\hat{\mathsf{K}}|\mathcal{S}_n\Big) - \mathbb{I}(\mathsf{K};\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n\mathsf{U}^n|\mathcal{S}_n)\\ &= \mathbb{H}(\mathsf{K}|\mathcal{S}_n) - \mathbb{H}\Big(\mathsf{K}|\hat{\mathsf{K}}\mathcal{S}_n\Big) - \mathbb{I}(\mathsf{K};\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n\mathsf{U}^n|\mathcal{S}_n)\\ &\geqslant nR - n\mathbb{H}(\mathsf{U}) - n\delta(\epsilon).\end{aligned}$$

All in all, we obtain

$$C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{X}\mathsf{Y}\tilde{\mathsf{Z}}}) \geqslant R - \mathbb{H}(\mathsf{U}) - \delta(\epsilon).$$

Since $\epsilon > 0$ can be arbitrarily small and R can be chosen arbitrarily close to $C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{X}\mathsf{Y}\mathsf{Z}})$, we get the desired result

$$C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{X}\mathsf{Y}\tilde{\mathsf{Z}}}) \geqslant C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{X}\mathsf{Y}\mathsf{Z}}) - \mathbb{H}(\mathsf{U}). \qquad \square$$

Proposition 4.1 means that providing Eve with side information U reduces the secret-key capacity by at most $\mathbb{H}(\mathsf{U})$. This motivates the definition of another measure of information.

Definition 4.8. *For a DMS* $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$*, the reduced intrinsic conditional information between* X *and* Y *given* Z *is*

$$\mathbb{I}(X;Y\downarrow\downarrow Z) \triangleq \inf_{p_{U|XYZ}} (\mathbb{I}(X;Y\downarrow ZU) + \mathbb{H}(U)).$$

Intuitively, the reduced intrinsic conditional information measures the information between X and Y that remains after the best memoryless processing of Z and with the best memoryless side information U; however, the term $\mathbb{H}(U)$ is introduced to compensate for the decrease of $\mathbb{I}(X;Y\downarrow Z)$ caused by the side information. The next theorem shows that $\mathbb{I}(X;Y\downarrow\downarrow Z)$ is an upper bound for C_s^{SM} that is at least as good as $\mathbb{I}(X;Y\downarrow Z)$.

Theorem 4.3 (Renner and Wolf)**.** *The secret-key capacity of a source model with DMS* $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$ *satisfies*

$$C_s^{SM} \leqslant \mathbb{I}(X;Y\downarrow\downarrow Z) \leqslant \mathbb{I}(X;Y\downarrow Z).$$

Proof. The inequality $\mathbb{I}(X;Y\downarrow\downarrow Z) \leqslant \mathbb{I}(X;Y\downarrow Z)$ follows on choosing $U = 0$ in the definition of $\mathbb{I}(X;Y\downarrow\downarrow Z)$. The inequality $C_s^{SM} \leqslant \mathbb{I}(X;Y\downarrow\downarrow Z)$ follows on noting that, for any U, Proposition 4.1 and Theorem 4.2 ensure that

$$C_s^{SM}(p_{XYZ}) \leqslant C_s^{SM}(p_{XY(ZU)}) + \mathbb{H}(U) \leqslant \mathbb{I}(X;Y\downarrow ZU) + \mathbb{H}(U)$$

and, therefore,

$$C_s^{SM}(p_{XYZ}) \leqslant \inf_{p_{U|XYZ}} (\mathbb{I}(X;Y\downarrow ZU) + \mathbb{H}(U)) = \mathbb{I}(X;Y\downarrow\downarrow Z). \qquad \square$$

As shown in the following example, the result of Theorem 4.3 is useful because $\mathbb{I}(X;Y\downarrow\downarrow Z)$ can provide a better bound than can $\mathbb{I}(X;Y\downarrow Z)$.

Example 4.2. Let $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3\}$ and let us consider again the joint probability distribution p_{XY} defined in the table below.

X \ Y	0	1	2	3
0	1/8	1/8	0	0
1	1/8	1/8	0	0
2	0	0	1/4	0
3	0	0	0	1/4

Let Z be defined as

$$Z \triangleq \begin{cases} X \oplus Y & \text{if} \quad X \in \{0, 1\}, \\ X \bmod 2 & \text{if} \quad X \in \{2, 3\}. \end{cases}$$

In contrast to Example 4.1, knowledge of Z never fully resolves the uncertainty about X or Y. One can verify that $\mathbb{I}(X;Y) = \mathbb{I}(X;Y|Z) = \frac{3}{2}$. We now introduce the side information U as

$$U \triangleq \left\lfloor \frac{X}{2} \right\rfloor.$$

Let us consider the channel $\left(\mathcal{Z} \times \mathcal{U}, p_{\bar{Z}|ZU}, \{0, 1, 2\}\right)$ such that

$$p_{\bar{Z}|ZU}(2|0, 0) = p_{\bar{Z}|ZU}(2|1, 0) = p_{\bar{Z}|ZU}(0|0, 1) = p_{\bar{Z}|ZU}(1|1, 1).$$

One can check that $\mathbb{I}\left(X; Y|\bar{Z}\right) = 0$ and therefore $\mathbb{I}(X; Y \downdownarrows Z) = 0$ as well.

To establish that $\mathbb{I}(X; Y \downarrow Z) > 0$, consider an arbitrary channel $\left(\mathcal{Z}, p_{\bar{Z}|Z}, \bar{\mathcal{Z}}\right)$ and consider an output symbol $\bar{z}^*$. Let

$$p \triangleq p_{\bar{Z}|Z}(\bar{z}^*|0) \quad \text{and} \quad q \triangleq p_{\bar{Z}|Z}(\bar{z}^*|1).$$

One can check that

$$p_{XY|Z}(0, 0|\bar{z}^*) = p_{XY|Z}(0, 1|\bar{z}^*) = p_{XY|Z}(1, 0|\bar{z}^*) = p_{XY|Z}(1, 1|\bar{z}^*) = \frac{p}{4(p+q)},$$

$$p_{XY|Z}(2, 2|\bar{z}^*) = p_{XY|Z}(3, 3|\bar{z}^*) = \frac{p}{2(p+q)},$$

and $\mathbb{I}\left(X; Y|\bar{Z} = \bar{z}^*\right) = \frac{3}{2}$. Consequently, $\mathbb{I}\left(X; Y|\bar{Z}\right) = \frac{3}{2}$ as well, and, since the channel $\left(\mathcal{Z}, p_{\bar{Z}|Z}, \bar{\mathcal{Z}}\right)$ was arbitrary, $\mathbb{I}(X; Y \downarrow Z) = \frac{3}{2}$.

Note that neither the intrinsic conditional information nor the reduced intrinsic conditional information really helps determine a generic expression for C_s^{SM}.

4.3 Sequential key distillation for the source model

The analysis of wiretap codes and key-distillation strategies is complex because messages or keys are subject to *simultaneous* reliability and secrecy constraints. Random-coding and random-binning arguments allow us to circumvent this difficulty and to establish achievability results, but they provide limited insight into the design of practical schemes. Even the proof in Section 4.2.2, which exploits a strong connection between key-distillation strategies and Slepian–Wolf codes, implicitly requires the key-distillation function and the public-message-encoding function to be designed *jointly*; this joint design makes it difficult to find such functions in practice. Hence, to further simplify the design of practical schemes, it is legitimate to wonder whether one could handle the reliability and secrecy requirements *independently*. This makes little sense for wiretap codes, but the idea is not totally contrived for key-distillation strategies because keys are random sequences that do not carry any information by themselves; there is a lot of leeway in the construction of key-distillation strategies and Alice and Bob are free to remove, combine, or shuffle their observations.

In this section, we show that, for a source model, it is indeed possible to design key-distillation strategies that handle reliability and secrecy independently. Such strategies, which we call *sequential key-distillation strategies* because they operate in sequential phases, play a particularly important role for three reasons:

- they incur no loss of optimality, since they can achieve all rates below the secret-key capacity;

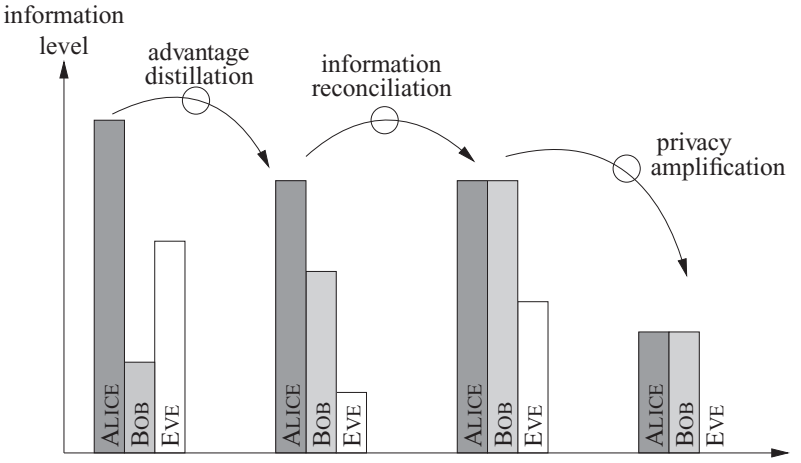

Figure 4.4 Evolution of information during the phases of a sequential key-generation strategy.

- they achieve strong secret-key rates, which allow us to prove $C_{\mathrm{s}}^{\mathrm{SM}} = \overline{C}_{\mathrm{s}}^{\mathrm{SM}}$;
- their analysis eventually leads to explicit and practical constructions.

Specifically, a sequential key-distillation strategy is a key-distillation strategy that operates in four successive phases.

1. *Randomness sharing.* Alice, Bob, and Eve observe n realizations of a DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$.
2. *Advantage distillation.* If needed, Alice and Bob exchange messages over the public channel to process their observations and to "distill" observations for which they have an advantage over Eve.
3. *Information reconciliation.* Alice and Bob exchange messages over the public channel to process their observations and agree on a common bit sequence.
4. *Privacy amplification.* Alice and Bob publicly agree on a deterministic function they apply to their common sequence to generate a secret key.

Before we describe and analyze these phases precisely, it is useful to understand intuitively the role played by each of them in the key-distillation strategy. Figure 4.4 illustrates the evolution of each party's information about Alice's initial source observations during the different phases. The amount of information is represented qualitatively by the height of the bars in the figure. After the randomness-sharing phase, we assume that Eve has an advantage over Bob; hence, Bob's information is lower than Eve's information. During the advantage-distillation phase, Alice and Bob interact over the public channel to distill observations for which they have an advantage over Eve. Since the observations for which Eve has an advantage are discarded, Alice's information decreases; however, Bob's information now exceeds Eve's information. During the information-reconciliation phase, Alice provides Bob with side information that enables him to correct all the discrepancies between his sequence and Alice's; as a result, Bob's information increases to reach the level of Alice's, but, since the error-correction information is public, Eve's

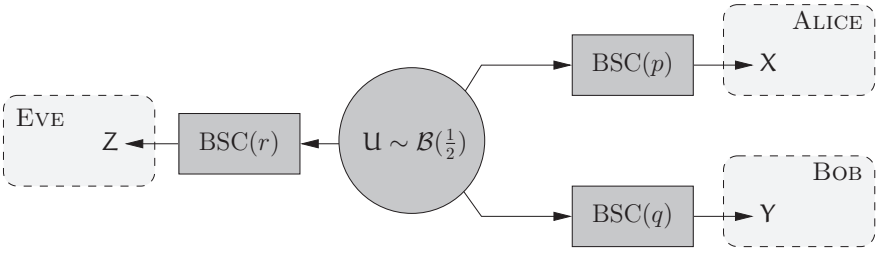

Figure 4.5 Satellite source model.

information increases as well. Finally, during the privacy-amplification phase, Alice and Bob generate a smaller sequence about which Eve has no information.

4.3.1 Advantage distillation

For some source models, the lower bound for the secret-key capacity provided by Theorem 4.1 is negative, and hence useless. Figure 4.5 illustrates such a source model, which is commonly called a "satellite" source model. It consists of a DMS $(\mathcal{U}, p_{\mathsf{U}})$ with $\mathsf{U} \sim \mathcal{B}(\frac{1}{2})$ that broadcasts sequences of bits to Alice, Bob, and Eve through independent binary symmetric channels with respective cross-over probabilities $p > 0$, $q > 0$, and $r > 0$. This source model can be thought of as a satellite transmitting to three base stations on Earth, one of which is an eavesdropper. It is assumed that Eve's cross-over probability r satisfies $r < p$ and $r < q$ so that

$$\mathbb{I}(\mathsf{X};\mathsf{Y}) < \mathbb{I}(\mathsf{X};\mathsf{Z}) \quad \text{and} \quad \mathbb{I}(\mathsf{X};\mathsf{Y}) < \mathbb{I}(\mathsf{Y};\mathsf{Z}).$$

In other words, Eve has an advantage over both Alice and Bob because the mutual information between Z and X and the mutual information between Z and Y are higher than that between X and Y.

The basic premise of advantage distillation is that Alice and Bob may reverse Eve's advantage by exchanging messages over the public channel. In fact, the mutual information $\mathbb{I}(\mathsf{X};\mathsf{Z})$ or $\mathbb{I}(\mathsf{Y};\mathsf{Z})$ measures only Eve's *average* advantage over Alice and Bob. Although $\mathbb{I}(\mathsf{X};\mathsf{Y}) < \mathbb{I}(\mathsf{X};\mathsf{Z})$ and $\mathbb{I}(\mathsf{X};\mathsf{Y}) < \mathbb{I}(\mathsf{Y};\mathsf{Z})$, there may exist some realizations of the DMS for which Eve's observations are loosely correlated to Alice and Bob's. One can think of an advantage-distillation protocol as a procedure to distill the realizations for which Alice and Bob have an advantage over Eve. Formally, an advantage-distillation protocol is defined as follows.

Definition 4.9. *An advantage-distillation protocol $\mathcal{D}_n$ for a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$ consists of*

- *two alphabets $\mathcal{X}'$ and $\mathcal{Y}'$;*
- *a source of local randomness $\left(\mathcal{R}_{\mathcal{X}}, p_{\mathsf{R}_{\mathsf{X}}}\right)$ for Alice;*
- *a source of local randomness $\left(\mathcal{R}_{\mathcal{Y}}, p_{\mathsf{R}_{\mathsf{Y}}}\right)$ for Bob;*
- *an integer $r \in \mathbb{N}^*$ that represents the number of rounds of communication;*
- *r encoding functions $f_i : \mathcal{X}^n \times \mathcal{B}^{i-1} \times \mathcal{R}_{\mathcal{X}} \to \mathcal{A}$ for $i \in [\![1, r]\!]$;*

- *r encoding functions* $g_i : \mathcal{Y}^n \times \mathcal{A}^{i-1} \times \mathcal{R}_\mathcal{Y} \to \mathcal{B}$ *for* $i \in [\![1, r]\!]$;
- *a function* $\theta_\mathrm{a} : \mathcal{X}^n \times \mathcal{B}^r \times \mathcal{R}_\mathcal{X} \to \mathcal{X}'$;
- *a function* $\theta_\mathrm{b} : \mathcal{Y}^n \times \mathcal{A}^r \times \mathcal{R}_\mathcal{Y} \to \mathcal{Y}'$;

and operates as follows:

- *Alice observes n realizations of the source* x^n *while Bob observes* y^n;
- *Alice generates a realization* r_x *of her source of local randomness while Bob generates* r_y *from his;*
- *in round* $i \in [\![1, r]\!]$, *Alice transmits* $a_i = f_i(x^n, b^{i-1}, r_x)$ *while Bob transmits* $b_i = g_i(y^n, a^{i-1}, r_y)$;
- *after round r, Alice distills* $x' = \theta_\mathrm{a}(x^n, b^r, r_x)$ *while Bob distills* $y' = \theta_\mathrm{b}(y^n, a^r, r_y)$.

By convention, $\mathcal{A}^0 \triangleq 0$ and $\mathcal{B}^0 \triangleq 0$. The number of rounds r and the sources of local randomness can be optimized during the design of the advantage-distillation protocol. By repeating an advantage-distillation protocol multiple times, Alice and Bob distill the realizations of a new DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$ with components X', Y', and $\mathsf{Z}' \triangleq \mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r$. Ideally, this new DMS provides Alice and Bob with an advantage over Eve in the sense that

$$\mathbb{I}(\mathsf{X}';\mathsf{Y}') \geqslant \mathbb{I}(\mathsf{X}';\mathsf{Z}') \quad \text{or} \quad \mathbb{I}(\mathsf{X}';\mathsf{Y}') \geqslant \mathbb{I}(\mathsf{Y}';\mathsf{Z}').$$

Hence, it is natural to measure the performance of an advantage-distillation protocol in terms of the quantity

$$\mathbf{R}(\mathcal{D}_n) \triangleq \frac{1}{n} \max\big(\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}'), \mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{Y}';\mathsf{Z}')\big),$$

which we call the *advantage-distillation rate*.[2] Notice that we have introduced a normalization by n to express the rate in bits per observation of the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$. The advantage-distillation rate captures an inherent trade-off in the design of an advantage-distillation protocol. On the one hand, Alice and Bob want to exchange messages to maximize $\mathbb{I}(\mathsf{X}';\mathsf{Y}')$, that is, they want to extract the observations for which their realizations are highly correlated; on the other hand, they must also minimize $\mathbb{I}(\mathsf{X}';\mathsf{Z}')$ or $\mathbb{I}(\mathsf{Y}';\mathsf{Z}')$, that is, they must choose their messages carefully in order to avoid revealing the values of their observations to Eve.

Definition 4.10. *An advantage-distillation rate R is achievable if there exists a sequence of advantage-distillation protocols* $\{\mathcal{D}_n\}_{n\geqslant 1}$ *such that*

$$\varliminf_{n\to\infty} \mathbf{R}(\mathcal{D}_n) \geqslant R.$$

Definition 4.11. *The advantage-distillation capacity* D^{SM} *of a source model with DMS* $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$ *is*

$$D^{\mathsf{SM}} \triangleq \sup\{R : R \text{ is an achievable advantage-distillation rate}\}.$$

[2] Our definition allows an advantage-distillation rate to take negative values, which of course has little interest.

When required, we write $D^{\mathrm{SM}}(p_{\mathrm{XYZ}})$ in place of D^{SM} to specify explicitly the distribution of the DMS. We do not attempt to characterize the advantage distillation of a source model exactly; rather, we relate it to the secret-key capacity of the same source model.

Proposition 4.2 (Muramatsu)**.** *For a source model with DMS* $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$*,*

$$D^{\mathrm{SM}} = C_{\mathrm{s}}^{\mathrm{SM}}.$$

Proof. A $(2^{nR}, n)$ key-distillation strategy for a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$ can be viewed as an advantage-distillation protocol for which $\mathsf{X}' = \mathsf{K}$, $\mathsf{Y}' = \hat{\mathsf{K}}$, and $\mathsf{Z}' = \mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n$. If a secret-key rate R is achievable, there exists a sequence of key-distillation strategies $\{\mathcal{S}_n\}_{n\geqslant 1}$ such that

$$\lim_{n\to\infty} \mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) = 0, \quad \lim_{n\to\infty} \frac{1}{n}\mathbf{L}(\mathcal{S}_n) = 0, \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n}\mathbf{U}(\mathcal{S}_n) = 0.$$

By virtue of Fano's inequality,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{H}\big(\mathsf{K}|\hat{\mathsf{K}}\mathcal{S}_n\big) \leqslant \lim_{n\to\infty} \delta(\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n)) = 0.$$

and, from the definition of $\mathbf{L}(\mathcal{S}_n)$,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{I}(\mathsf{K}; \mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r|\mathcal{S}_n) = \lim_{n\to\infty} \frac{1}{n}\mathbf{L}(\mathcal{S}_n) = 0.$$

Hence, the sequence of advantage-distillation rate $\{\mathbf{R}(\mathcal{S}_n)\}_{n\geqslant 1}$ satisfies

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{R}(\mathcal{S}_n) &= \lim_{n\to\infty} \left(\frac{1}{n}\mathbb{I}\big(\mathsf{X}';\mathsf{Y}'\big) - \frac{1}{n}\mathbb{I}\big(\mathsf{X}';\mathsf{Z}'\big) \right) \\
&= \lim_{n\to\infty} \left(\frac{1}{n}\mathbb{H}(\mathsf{K}|\mathcal{S}_n) - \frac{1}{n}\mathbb{H}\Big(\mathsf{K}|\hat{\mathsf{K}}\mathcal{S}_n\Big) - \frac{1}{n}\mathbb{I}(\mathsf{K}; \mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r|\mathcal{S}_n) \right) \\
&= \lim_{n\to\infty} \left(\frac{1}{n}\log\lceil 2^{nR}\rceil - \frac{1}{n}\mathbf{U}(\mathcal{S}_n) - \frac{1}{n}\mathbb{H}\Big(\mathsf{K}|\hat{\mathsf{K}}\mathcal{S}_n\Big) - \frac{1}{n}\mathbb{I}(\mathsf{K}; \mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r|\mathcal{S}_n) \right) \\
&= R.
\end{aligned}$$

Therefore, $R \leqslant D^{\mathrm{SM}}$; and, since R is an arbitrary achievable secret-key rate, $C_{\mathrm{s}}^{\mathrm{SM}} \leqslant D^{\mathrm{SM}}$.

To prove the reverse inequality, notice that repeated application of an advantage-distillation protocol $\mathcal{D}_n$ creates a source model with DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathrm{X'Y'Z'}})$, whose secret-key capacity is $C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{X'Y'Z'}})$; by Theorem 4.1, $C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{X'Y'Z'}})$ satisfies

$$C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{X'Y'Z'}}) \geqslant \max\big(\mathbb{I}\big(\mathsf{X}';\mathsf{Y}'\big) - \mathbb{I}\big(\mathsf{X}';\mathsf{Z}'\big), \mathbb{I}\big(\mathsf{X}';\mathsf{Y}'\big) - \mathbb{I}\big(\mathsf{Y}';\mathsf{Z}'\big)\big).$$

The secret-key capacity $C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{X'Y'Z'}})$ is expressed in bits per observation of the source $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathrm{X'Y'Z'}})$; hence, the corresponding secret-key rate in bits per observation of the source $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$ is $(1/n)C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{X'Y'Z'}})$ and, by definition, it cannot exceed $C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{XYZ}})$. Therefore,

$$\begin{aligned}
C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{XYZ}}) &\geqslant \frac{1}{n}C_{\mathrm{s}}^{\mathrm{SM}}(p_{\mathrm{X'Y'Z'}}) \\
&\geqslant \frac{1}{n}\max\big(\mathbb{I}\big(\mathsf{X}';\mathsf{Y}'\big) - \mathbb{I}\big(\mathsf{X}';\mathsf{Z}'\big), \mathbb{I}\big(\mathsf{X}';\mathsf{Y}'\big) - \mathbb{I}\big(\mathsf{Y}';\mathsf{Z}'\big)\big) \\
&= \mathbf{R}(\mathcal{D}_n).
\end{aligned}$$

Since the advantage-distillation rate $\mathbf{R}(\mathcal{D}_n)$ can be arbitrarily close to $D^{\mathsf{SM}}(p_{\mathrm{XYZ}})$, we obtain

$$C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathrm{XYZ}}) \geqslant D^{\mathsf{SM}}(p_{\mathrm{XYZ}}). \qquad \square$$

Proposition 4.2 does not provide an explicit characterization of the advantage-distillation capacity, but it shows that the secret-key capacity is equal to the maximum rate at which Alice and Bob can "generate an advantage" over Eve. This formalizes the intuition that the amount of secrecy that Alice and Bob can extract from their observations is related to the advantage they have over Eve. In addition, the proof shows that there is no loss of optimality in starting a key-distillation strategy with an advantage-distillation phase.

Unfortunately, there is no known generic procedure for designing advantage-distillation protocols because the distillation heavily depends on the specific statistics of the underlying DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$. Nevertheless, we illustrate the concept by analyzing a protocol for the satellite scenario of Figure 4.5. This protocol, which is called the "repetition" protocol, is due to Maurer and operates as follows.

1. Alice, Bob, and Eve observe m realizations of the DMS denoted by X^m, Y^m, and Z^m, respectively.
2. Alice generates a bit $\mathrm{V} \sim \mathcal{B}(\frac{1}{2})$.
3. Alice creates a vector $\overline{\mathrm{V}}^m \triangleq (\mathrm{V}, \ldots, \mathrm{V})$ that consists of m repetitions of the same bit V, and transmits the bit-wise sum $\overline{\mathrm{V}}^m \oplus \mathrm{X}^m$ over the public channel.
4. Upon reception of $\overline{\mathrm{V}}^m \oplus \mathrm{X}^m$, Bob uses his observations Y^m to compute $\overline{\mathrm{V}}^m \oplus \mathrm{X}^m \oplus \mathrm{Y}^m$ and defines

$$\tilde{\mathrm{Y}} \triangleq \begin{cases} 0 & \text{if} \quad \overline{\mathrm{V}}^m \oplus \mathrm{X}^m \oplus \mathrm{Y}^m = (0, 0, \ldots, 0), \\ 1 & \text{if} \quad \overline{\mathrm{V}}^m \oplus \mathrm{X}^m \oplus \mathrm{Y}^m = (1, 1, \ldots, 1), \\ ? & \text{else.} \end{cases}$$

   Bob then sends Alice a message F over the public channel defined as

$$\mathsf{F} \triangleq \begin{cases} 1 & \text{if} \quad \tilde{\mathrm{Y}} \in \{0, 1\}, \\ 0 & \text{if} \quad \tilde{\mathrm{Y}} = ?. \end{cases}$$

   Notice that F carries information about $\tilde{\mathrm{Y}}$ but does not reveal its exact value.
5. Upon reception of F, Alice defines $\tilde{\mathrm{X}}$ as

$$\tilde{\mathrm{X}} \triangleq \begin{cases} \mathrm{V} & \text{if} \quad \mathsf{F} = 1, \\ ? & \text{if} \quad \mathsf{F} = 0. \end{cases}$$

By reiterating the repetition protocol multiple times, Alice and Bob effectively create a new DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathrm{X}'\mathrm{Y}'\mathrm{Z}'})$ with components $\mathrm{X}' \triangleq \tilde{\mathrm{X}}\mathsf{F}$, $\mathrm{Y}' \triangleq \tilde{\mathrm{Y}}$, and $\mathrm{Z}' \triangleq \big(\mathrm{Z}^m, \overline{\mathrm{V}}^m \oplus \mathrm{X}^m, \mathsf{F}\big)$. The protocol can be thought of as a *post-selection* procedure with a repetition code, by which Alice and Bob retain a bit only if their observations X^m and Y^m are highly correlated or anticorrelated. Since Eve is a passive eavesdropper and has no control over the post-selection, she cannot bias Alice or Bob towards selecting observations that would be favorable to her.

We now compute $\mathbb{I}(\mathrm{X}'; \mathrm{Y}')$ and $\mathbb{I}(\mathrm{X}'; \mathrm{Z}')$ obtained with the repetition protocol in order to characterize its advantage-distillation rate.

Proposition 4.3. *The mutual information between Alice and Bob after advantage distillation with the repetition protocol is*

$$\mathbb{I}(X';Y') = \mathbb{H}_b(\alpha) + \alpha\,(1 - \mathbb{H}_b(\beta)),$$

with

$$\alpha \triangleq (\bar{p}\bar{q} + pq)^m + (\bar{p}q + p\bar{q})^m, \qquad \beta \triangleq \frac{(\bar{p}q + p\bar{q})^m}{(\bar{p}\bar{q} + pq)^m + (\bar{p}q + p\bar{q})^m},$$

and $\bar{p} \triangleq (1 - p)$ *and* $\bar{q} = (1 - q)$.

Proof. The mutual information between Alice and Bob after the repetition protocol can be written

$$\mathbb{I}(X';Y') = \mathbb{I}(\tilde{X}\mathsf{F};\tilde{Y}) = \mathbb{I}(\mathsf{F};\tilde{Y}) + \mathbb{I}(\tilde{X};\tilde{Y}|\mathsf{F}).$$

By construction, $\mathbb{H}(\mathsf{F}|\tilde{Y}) = 0$ and $\mathbb{I}(\tilde{X};\tilde{Y}|\mathsf{F} = 0) = 0$; therefore,

$$\mathbb{I}(X';Y') = \mathbb{H}(\mathsf{F}) + \mathbb{P}[\mathsf{F} = 1]\mathbb{I}(\tilde{X};\tilde{Y}|\mathsf{F} = 1). \tag{4.28}$$

We compute each term on the right-hand side of (4.28) separately. By construction, $\mathbb{P}[\mathsf{F} = 1]$ is the probability that Bob obtains $\tilde{Y} \in \{0, 1\}$, which can be computed explicitly as

$$\begin{aligned}\alpha \triangleq \mathbb{P}[\mathsf{F} = 1] &= \mathbb{P}[X^m \oplus Y^m = 0 \quad \text{or} \quad X^m \oplus Y^m = 1] \\ &= (\bar{p}\bar{q} + pq)^m + (\bar{p}q + p\bar{q})^m. \end{aligned} \tag{4.29}$$

The probability that $\tilde{Y}$ differs from $\tilde{X}$ given $\mathsf{F} = 1$ is simply the probability that $X^m \oplus Y^m = \bar{1}$ given $\mathsf{F} = 1$; hence,

$$\beta \triangleq \mathbb{P}\left[\tilde{X} \neq \tilde{Y}|\mathsf{F} = 1\right] = \frac{(\bar{p}q + p\bar{q})^m}{(\bar{p}\bar{q} + pq)^m + (\bar{p}q + p\bar{q})^m}. \tag{4.30}$$

The conditional probabilities $p_{\tilde{Y}|\tilde{X}\mathsf{F}}$ are those of a BSC with cross-over probability $\mathbb{P}\left[\tilde{X} \neq \tilde{Y}|\mathsf{F} = 1\right]$; therefore,

$$\mathbb{I}(\tilde{X};\tilde{Y}|\mathsf{F} = 1) = 1 - \mathbb{H}_b(\beta). \tag{4.31}$$

On combining (4.29), (4.30), and (4.31) in (4.28), we obtain

$$\mathbb{I}(X';Y') = \mathbb{H}_b(\alpha) + \alpha\,(1 - \mathbb{H}_b(\beta)). \qquad \square$$

Proposition 4.4. *The mutual information between Alice and Eve after advantage distillation with the repetition protocol is*

$$\mathbb{I}(X';Z') = \mathbb{H}_b(\alpha) + \alpha \sum_{k=0}^{m} \binom{m}{k} p_k \left(1 - \mathbb{H}_b\left(\frac{p_k}{p_k + p_{m-k}}\right)\right),$$

with $\alpha \triangleq (\bar{p}\bar{q} + pq)^m + (\bar{p}q + p\bar{q})^m$ *and*

$$p_k \triangleq \frac{1}{\alpha}(\bar{p}\bar{q}r + pq\bar{r})^k(\bar{p}\bar{q}\bar{r} + pqr)^{m-k} + \frac{1}{\alpha}(\bar{p}qr + p\bar{q}\bar{r})^k(\bar{p}q\bar{r} + p\bar{q}r)^{m-k}.$$

Proof. Because Eve may not perform any hard decision on the bit V, evaluating Eve's probability of error[3] does not suffice to compute $\mathbb{I}(X';Z')$. Nevertheless, the satellite source model and the repetition protocol are simple enough that we can compute $\mathbb{I}(X';Z')$ in closed form. Using the chain rule,

$$\begin{aligned}\mathbb{I}(X';Z') &= \mathbb{I}(\tilde{X}F;Z^m,\overline{V}^m\oplus X^m,F)\\ &= \mathbb{H}(\tilde{X}F)-\mathbb{H}(\tilde{X}F|Z^m,\overline{V}^m\oplus X^m,F)\\ &= \mathbb{H}(F)+\mathbb{H}(\tilde{X}|F)-\mathbb{H}(\tilde{X}|Z^m,\overline{V}^m\oplus X^m,F)\,. \end{aligned}\tag{4.32}$$

Notice that the pair $(Z^m,\overline{V}^m\oplus X^m)$ uniquely determines $(Z^m,\overline{V}^m\oplus X^m\oplus Z^n)$ and vice versa, so that $\mathbb{H}(\tilde{X}|Z^m,\overline{V}^m\oplus X^m,F)=\mathbb{H}(\tilde{X}|Z^m,\overline{V}^m\oplus X^m\oplus Z^m,F)$. In addition, Z^m and X^m are observations of the same i.i.d. sequence U^m through indendent BSCs. Hence, we can write $Z^n=U^n\oplus E_z^m$ and $X^n=U^n\oplus E_x^m$, where E_z^m and E_x^m are the independent error patterns introduced by the BSCs. Consequently,

$$\begin{aligned}&\mathbb{H}(\tilde{X}|Z^m,\overline{V}^m\oplus X^m,F)\\ &\quad=\mathbb{H}(\tilde{X}|Z^m,\overline{V}^m\oplus X^m\oplus Z^m,F)\\ &\quad=\mathbb{H}(\tilde{X}|U^m\oplus E_z^m,\overline{V}^m\oplus E_x^m\oplus E_z^m,F)\\ &\quad=\mathbb{H}(\tilde{X},U^m\oplus E_z^m|\overline{V}^m\oplus E_x^m\oplus E_z^m,F)-\mathbb{H}(U^m\oplus E_z^m|\overline{V}^m\oplus E_x^m\oplus E_z^m,F)\\ &\quad=\mathbb{H}(\tilde{X}|\overline{V}^m\oplus E_x^m\oplus E_z^m,F)+\mathbb{H}(U^m\oplus E_z^m|\overline{V}^m\oplus E_x^m\oplus E_z^m,F,\tilde{X})\\ &\qquad-\mathbb{H}(U^m\oplus E_z^m|\overline{V}^m\oplus E_x^m\oplus E_z^m,F).\end{aligned}$$

Since the sequence U^m is uniformly distributed, the crypto lemma ensures that

$$\mathbb{H}(U^m\oplus E_z^m|\overline{V}^m\oplus E_x^m\oplus E_z^m,F,\tilde{X})=\mathbb{H}(U^m\oplus E_z^m|\overline{V}^m\oplus E_x^m\oplus E_z^m,F)=m,$$

and, therefore,

$$\mathbb{H}(\tilde{X}|Z^m,\overline{V}^m\oplus X^m,F)=\mathbb{H}(\tilde{X}|\overline{V}^m\oplus E_x^m\oplus E_z^m,F).\tag{4.33}$$

On substituting (4.33) into (4.32), we obtain

$$\begin{aligned}\mathbb{I}(X';Z') &= \mathbb{H}(F)+\mathbb{H}(\tilde{X}|F)-\mathbb{H}(\tilde{X}|\overline{V}^m\oplus E_x^m\oplus E_z^m,F)\\ &= \mathbb{H}(F)+\mathbb{I}(\tilde{X};\overline{V}^m\oplus E_x^m\oplus E_z^m|F)\\ &= \mathbb{H}(F)+\mathbb{P}[F=1]\mathbb{I}(\tilde{X};\overline{V}^m\oplus E_x^m\oplus E_z^m|F=1),\end{aligned}\tag{4.34}$$

where we have used $\mathbb{I}(\tilde{X};Z^m,\overline{V}^m\oplus X^m|F=0)=0$ to obtain the last equality. Given $F=1$, note that $\tilde{X}=V$ and that the weight $W\triangleq w(\overline{V}^m\oplus E_x^m\oplus E_z^m)$ of the sequence $\overline{V}^m\oplus E_x^m\oplus E_z^m$ is a sufficient statistic for $\tilde{X}$ given $\overline{V}^m\oplus E_x^m\oplus E_z^m$. Hence,

$$\mathbb{I}(X';Z')=\mathbb{H}(F)+\mathbb{P}[F=1]\mathbb{I}(\tilde{X};W|F=1).$$

[3] If we were to compute the probability of error for Eve under maximum-likelihood decoding, Fano's inequality would yield only a *lower* bound for $\mathbb{I}(X';Z')$.

Finally, all we need in order to compute $\mathbb{I}\big(\tilde{\mathsf{X}};W|\mathsf{F}=1\big)$ is the joint distribution $\mathbb{P}\big[W,\tilde{\mathsf{X}}|\mathsf{F}=1\big]$. For any weight $k\in[\![0,m]\!]$,

$$\begin{aligned}\mathbb{P}\big[W=k,\tilde{\mathsf{X}}=0|\mathsf{F}=1\big]&=\mathbb{P}\big[w(\overline{\mathsf{V}}^m\oplus\mathsf{E}_x^m\oplus\mathsf{E}_z^m)=k,\tilde{\mathsf{X}}=0|\mathsf{F}=1\big]\\&=\mathbb{P}\big[w\left(\mathsf{E}_x^m\oplus\mathsf{E}_z^m\right)=k,\mathsf{V}=0|\mathsf{F}=1\big]\\&=\mathbb{P}\big[w\left(\mathsf{E}_x^m\oplus\mathsf{E}_z^m\right)=k|\mathsf{F}=1\big]\mathbb{P}[\mathsf{V}=0].\end{aligned}$$

By construction, $\mathbb{P}[\mathsf{V}=0]=\frac{1}{2}$, and $\mathbb{P}\big[w\left(\mathsf{E}_x^m\oplus\mathsf{E}_z^m\right)=k|\mathsf{F}=1\big]$ can be written as

$$\begin{aligned}&\frac{\mathbb{P}\big[w\left(\mathsf{E}_x^m\oplus\mathsf{E}_z^m\right)=k,\mathsf{F}=1\big]}{\mathbb{P}[\mathsf{F}=1]}\\&=\frac{1}{\alpha}\big(\mathbb{P}\big[w\big(\mathsf{E}_x^m\oplus\mathsf{E}_z^m\big)=k,\mathsf{X}^m\oplus\mathsf{Y}^m=0\big]+\mathbb{P}\big[w\big(\mathsf{E}_x^m\oplus\mathsf{E}_z^m\big)=k,\mathsf{X}^m\oplus\mathsf{Y}^m=1\big]\big)\\&=\frac{1}{\alpha}\left(\binom{m}{k}(\bar{p}\bar{q}r+pq\bar{r})^k(\bar{p}\bar{q}\bar{r}+pqr)^{m-k}+\binom{m}{k}(\bar{p}qr+p\bar{q}\bar{r})^k(\bar{p}q\bar{r}+p\bar{q}r)^{m-k}\right).\end{aligned}$$

Upon defining p_k as

$$p_k\triangleq\frac{1}{\alpha}(\bar{p}\bar{q}r+pq\bar{r})^k(\bar{p}\bar{q}\bar{r}+pqr)^{m-k}+\frac{1}{\alpha}(\bar{p}qr+p\bar{q}\bar{r})^k(\bar{p}q\bar{r}+p\bar{q}r)^{m-k},$$

we obtain

$$\mathbb{P}\big[W=k,\tilde{\mathsf{X}}=0|\mathsf{F}=1\big]=\frac{1}{2}\binom{m}{k}p_k. \tag{4.35}$$

Similarly,

$$\mathbb{P}\big[W=k,\tilde{\mathsf{X}}=1|\mathsf{F}=1\big]=\frac{1}{2}\binom{m}{k}p_{m-k}, \tag{4.36}$$

and, on combining (4.35) and (4.36), we have

$$\mathbb{P}[W=k|\mathsf{F}=1]=\frac{1}{2}\binom{m}{k}(p_k+p_{m-k}).$$

We can now evaluate $\mathbb{I}(X';W|\mathsf{F}=1)$ explicitly as

$$\begin{aligned}&\sum_{k=0}^{m}\left(\mathbb{P}\big[W=k,\tilde{\mathsf{X}}=0|\mathsf{F}=1\big]\log\left(\frac{\mathbb{P}\big[W=k|\tilde{\mathsf{X}}=0,\mathsf{F}=1\big]}{\mathbb{P}[W=k|\mathsf{F}=1]}\right)\right.\\&\qquad\left.+\,\mathbb{P}\big[W=k,\tilde{\mathsf{X}}=1|\mathsf{F}=1\big]\log\left(\frac{\mathbb{P}\big[W=k|\tilde{\mathsf{X}}=1,\mathsf{F}=1\big]}{\mathbb{P}[W=k|\mathsf{F}=1]}\right)\right)\\&=\sum_{k=0}^{m}\left(\frac{1}{2}\binom{m}{k}p_k\log\left(\frac{2p_k}{p_{m-k}+p_k}\right)+\frac{1}{2}\binom{m}{k}p_{m-k}\log\left(\frac{2p_{m-k}}{p_{m-k}+p_k}\right)\right)\\&=\frac{1}{2}\sum_{k=0}^{m}\binom{m}{k}(p_k+p_{m-k})+\frac{1}{2}\sum_{k=0}^{m}\binom{m}{k}(p_k+p_{m-k})\left(-\mathbb{H}_{\mathrm{b}}\left(\frac{p_k}{p_k+p_{m-k}}\right)\right)\\&=\sum_{k=0}^{m}\binom{m}{k}p_k\left(1-\mathbb{H}_{\mathrm{b}}\left(\frac{p_k}{p_k+p_{m-k}}\right)\right),\end{aligned}$$

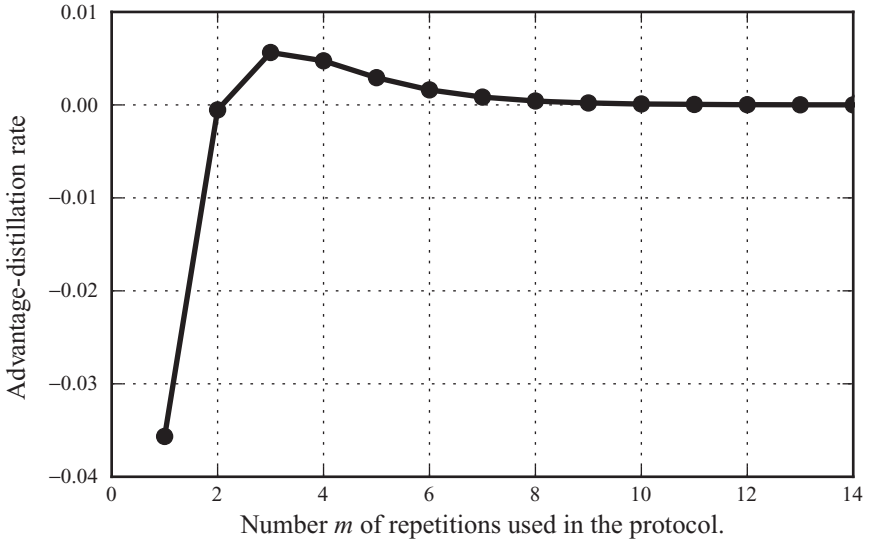

Figure 4.6 Advantage-distillation rate for different values of the repetition-protocol parameter m. The parameters of the satellite source are $p = q = 0.2$ and $r = 0.15$.

where the last equality follows because $\sum_{k=0}^{m} \binom{m}{k} p_k = \sum_{k=0}^{m} \binom{m}{k} p_{m-k}$ and

$$\mathbb{H}_{\mathrm{b}}\left(\frac{p_k}{p_k + p_{m-k}}\right) = \mathbb{H}_{\mathrm{b}}\left(1 - \frac{p_k}{p_k + p_{m-k}}\right) = \mathbb{H}_{\mathrm{b}}\left(\frac{p_{m-k}}{p_k + p_{m-k}}\right).$$

On substituting (4.29) and (4.13) into (4.34), we obtain the desired result

$$\mathbb{I}(\mathsf{X}';\mathsf{Z}') = \mathbb{H}_{\mathrm{b}}(\alpha) + \alpha \sum_{k=0}^{m} \binom{m}{k} p_k \left(1 - \mathbb{H}_{\mathrm{b}}\left(\frac{p_k}{p_k + p_{m-k}}\right)\right). \qquad \square$$

Figure 4.6 illustrates the advantage-distillation rate of the protocol for a satellite source with $p = q = 0.2$ and $r = 0.15$ for various values of the repetition parameter m. Note that, on choosing m large enough, the protocol achieves a strictly positive advantage-distillation rate; however, the protocol is inefficient and the advantage-distillation rates are quite low. For instance, with $p = q = 0.2$, $r = 0.15$, and $m = 3$, the advantage-distillation rate is on the order of 0.005 bits per observation of the DMS. This rate is relatively low because the post-selection with a repetition code is extremely wasteful of source observations. It is possible to improve these rates by using slightly better codes, and we refer the reader to the bibliographical notes at the end of this chapter for additional details.

4.3.2 Information reconciliation

After the advantage-distillation phase, Alice, Bob, and Eve, obtain the realizations of a DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$. The objective of the information-reconciliation phase (reconciliation for short) is to allow Alice and Bob to agree on a common sequence S,

but, at this stage, the sequence S is not subject to any secrecy constraint. To simplify the notation, we assume that the reconciliation protocol operates on the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$ instead of on the DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathrm{X'Y'Z'}})$ obtained with advantage distillation.

We place few restrictions on how reconciliation should be performed. The common sequence S could be a function of Alice and Bob's observations and of messages exchanged interactively over the public authenticated channel. Alice and Bob are also allowed to randomize their operations using sources of local randomness. Formally, a reconciliation protocol is defined as follows.

Definition 4.12. *A reconciliation protocol $\mathcal{R}_n$ for a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathrm{XYZ}})$ consists of*

- *an alphabet $\mathcal{S} = [\![1, S]\!]$;*
- *a source of local randomness $\left(\mathcal{R}_{\mathcal{X}}, p_{\mathrm{R_X}}\right)$ for Alice;*
- *a source of local randomness $\left(\mathcal{R}_{\mathcal{Y}}, p_{\mathrm{R_Y}}\right)$ for Bob;*
- *an integer $r \in \mathbb{N}^*$ that represents the number of rounds of communication;*
- *r encoding functions $f_i : \mathcal{X}^n \times \mathcal{B}^{i-1} \times \mathcal{R}_{\mathcal{X}} \to \mathcal{A}$ for $i \in [\![1, r]\!]$;*
- *r encoding functions $g_i : \mathcal{Y}^n \times \mathcal{A}^{i-1} \times \mathcal{R}_{\mathcal{Y}} \to \mathcal{B}$ for $i \in [\![1, r]\!]$;*
- *a function $\eta_{\mathrm{a}} : \mathcal{X}^n \times \mathcal{B}^r \times \mathcal{R}_{\mathcal{X}} \to \mathcal{S}$;*
- *a function $\eta_{\mathrm{b}} : \mathcal{Y}^n \times \mathcal{A}^r \times \mathcal{R}_{\mathcal{Y}} \to \mathcal{S}$;*

and operates as follows:

- *Alice observes n realizations of the source x^n while Bob observes y^n;*
- *Alice generates a realization r_x of her source of local randomness while Bob generates r_y from his;*
- *in round $i \in [\![1, r]\!]$, Alice transmits $a_i = f_i(x^n, b^{i-1}, r_x)$ while Bob transmits $b_i = g_i(y^n, a^{i-1}, r_y)$;*
- *after round r, Alice computes $s = \eta_{\mathrm{a}}(x^n, b^r, r_x)$ while Bob computes $\hat{s} = \eta_{\mathrm{b}}(y^n, a^r, r_y)$.*

By convention, $\mathcal{A}^0 \triangleq 0$ and $\mathcal{B}_0 \triangleq 0$. The number of rounds r and the sources of local randomness can be optimized during the design of the reconciliation protocol. Note that the definition of a reconciliation protocol is slightly different from that of an advantage-distillation protocol because the objective is not the same. The goal of the reconciliation protocol is to guarantee that Alice and Bob agree on a common sequence; hence, the output alphabets of the functions η_{a} and η_{b} are the same. The goal of an advantage-distillation protocol is merely to generate a new source; hence, the output alphabets of the functions θ_{a} and θ_{b} in Definition 4.9 could be different.

The reliability performance of a reconciliation protocol is measured in terms of the average probability of error

$$\mathbf{P}_{\mathrm{e}}(\mathcal{R}_n) \triangleq \mathbb{P}\left[\mathrm{S} \neq \hat{\mathrm{S}} \middle| \mathcal{R}_n\right].$$

In addition, since the common sequence S generated by a reconciliation protocol is eventually processed to generate a secret key, it is desirable that the protocol leaks as

little information as possible over the public channel. Hence, a reasonable measure of performance is the difference between the entropy of the common sequence $\mathbb{H}(S)$ and the amount of public information $\mathbb{H}(A^r B^r)$ exchanged over the public channel. The quantity

$$\mathbf{R}(\mathcal{R}_n) \triangleq \frac{1}{n}(\mathbb{H}(S|\mathcal{R}_n) - \mathbb{H}(A^r B^r|\mathcal{R}_n))$$

is called the *reconciliation rate* of a reconciliation protocol. The choice of this measure is fully justified when we discuss privacy amplification in Section 4.3.3.

Definition 4.13. *A reconciliation rate R is achievable if there exists a sequence of reconciliation protocols $\{\mathcal{R}_n\}_{n\geqslant 1}$ such that*

$$\lim_{n\to\infty} \mathbf{P}_e(\mathcal{R}_n) = 0 \quad \textit{and} \quad \underline{\lim}_{n\to\infty} \mathbf{R}(\mathcal{R}_n) \geqslant R.$$

Definition 4.14. *The reconciliation capacity R^{SM} of a DMS $(\mathcal{XY}, p_{XY})$ is*

$$R^{\mathsf{SM}} \triangleq \sup\{R : R \text{ is an achievable reconciliation rate}\}.$$

Proposition 4.5. *The reconciliation capacity of a source model with DMS $(\mathcal{XYZ}, p_{XYZ})$ is*

$$R^{\mathsf{SM}} = \mathbb{I}(X;Y).$$

In addition the reconciliation rates below R^{SM} are achievable with one-way communication and without sources of local randomness.

Proof. We start by proving that the reconciliation capacity cannot exceed $\mathbb{I}(X;Y)$. Let R be an achievable reconciliation rate. By definition, for any $\epsilon > 0$, there exists a reconciliation protocol $\mathcal{R}_n$ such that

$$\mathbf{P}_e(\mathcal{R}_n) \leqslant \delta(\epsilon) \quad \text{and} \quad \mathbf{R}(\mathcal{R}_n) \geqslant R - \delta(\epsilon). \tag{4.37}$$

In the remainder of the proof, we omit the conditioning on $\mathcal{R}_n$ to simplify the notation. Fano's inequality guarantees that $(1/n)\mathbb{H}(S|\hat{S}) \leqslant \delta(\mathbf{P}_e(\mathcal{R}_n)) \leqslant \delta(\epsilon)$; therefore,

$$\begin{aligned} R &\leqslant \mathbf{R}(\mathcal{R}_n) + \delta(\epsilon) \\ &= \frac{1}{n}\left(\mathbb{H}(S) - \mathbb{H}(A^r B^r)\right) + \delta(\epsilon) \\ &\leqslant \frac{1}{n}\mathbb{H}(S) - \frac{1}{n}\mathbb{H}(S|\hat{S}) - \frac{1}{n}\mathbb{H}(A^r B^r) + \delta(\epsilon) \\ &= \frac{1}{n}\mathbb{I}\left(S;\hat{S}\right) - \frac{1}{n}\mathbb{H}(A^r B^r) + \delta(\epsilon). \end{aligned} \tag{4.38}$$

Since $S \to X^n A^r B^r R_X \to Y^n A^r B^r R_Y \to \hat{S}$ forms a Markov chain, the data-processing inequality ensures that

$$\mathbb{I}\left(S;\hat{S}\right) \leqslant \mathbb{I}(X^n R_X A^r B^r; Y^n R_Y A^r B^r). \tag{4.39}$$

We further expand $\mathbb{I}(X^n R_X A^r B^r; Y^n R_Y A^r B^r)$ as

$$\begin{aligned}
&\mathbb{I}(X^n R_X A^r B^r; Y^n R_Y A^r B^r) \\
&\quad = \mathbb{I}(X^n R_X; Y^n R_Y A^r B^r) + \mathbb{I}(A^r B^r; Y^n R_Y A^r B^r | X^n R_X) \\
&\quad = \mathbb{I}(X^n R_X; Y^n R_Y | A^r B^r) + \mathbb{I}(X^n R_X; A^r B^r) + \mathbb{I}(A^r B^r; Y^n R_Y A^r B^r | X^n R_X) \\
&\quad = \mathbb{I}(X^n R_X; Y^n R_Y | A^r B^r) + \mathbb{H}(A^r B^r) - \mathbb{H}(A^r B^r | X^n R_X) \\
&\qquad + \mathbb{I}(A^r B^r; Y^n R_Y A^r B^r | X^n R_X). \qquad (4.40)
\end{aligned}$$

Note that

$$\mathbb{I}(A^r B^r; Y^n R_Y A^r B^r | X^n R_X) = \mathbb{H}(A^r B^r | X^n R_X). \qquad (4.41)$$

By applying Lemma 4.2 with $S \triangleq X^n R_X$, $T \triangleq Y^n R_Y$, $V^r \triangleq A^r$, and $W^r \triangleq B^r$, we obtain

$$\mathbb{I}(X^n R_X; Y^n R_Y | A^r B^r) \leqslant \mathbb{I}(X^n R_X; Y^n R_Y).$$

Since the DMSs $\left(\mathcal{R}_{\mathcal{X}}, p_{R_X}\right)$ and $\left(\mathcal{R}_{\mathcal{Y}}, p_{R_Y}\right)$ are mutually independent and independent of the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$, we have

$$\mathbb{I}(X^n R_X; Y^n R_Y | A^r B^r) \leqslant n\mathbb{I}(X; Y). \qquad (4.42)$$

On combining (4.41) and (4.42) in (4.40), we have

$$\mathbb{I}(X^n R_X A^r B^r; Y^n R_Y A^r B^r) \leqslant n\mathbb{I}(X; Y) + \mathbb{H}(A^r B^r), \qquad (4.43)$$

and, using (4.43) in (4.39) and (4.38), we obtain

$$R \leqslant \mathbb{I}(X; Y) + \delta(\epsilon).$$

Since ϵ can be chosen arbitrarily small and R can be chosen arbitrarily close to R^{SM}, it must hold that $R^{\mathsf{SM}} \leqslant \mathbb{I}(X; Y)$.

We now show that all reconciliation rates below R^{SM} are achievable. This result follows directly from Corollary 2.4. In fact, for any $\epsilon > 0$, Corollary 2.4 guarantees the existence of a $(2^{nR}, n)$ code $\mathcal{C}_n$ that compresses X^n into a message A at rate $R \leqslant \mathbb{H}(X|Y) + \delta(\epsilon)$ and such that X^n can be retrieved from Y^n and A with probability of error $\mathbf{P}_{\mathrm{e}}(\mathcal{C}_n) \leqslant \delta(\epsilon)$. Such a code can be viewed as a reconciliation protocol $\mathcal{R}_n$ without sources of local randomness, for which $S = X^n$ and in which there is a single public message A of about $n\mathbb{H}(X|Y)$ bits exchanged over the public channel. The corresponding reconciliation rate is

$$\begin{aligned}
\mathbf{R}(\mathcal{R}_n) &= \frac{1}{n}(\mathbb{H}(S) - \mathbb{H}(A)) \\
&\geqslant \frac{1}{n}\mathbb{H}(X^n) - R \\
&= \mathbb{I}(X; Y) - \delta(\epsilon).
\end{aligned}$$

Hence, all reconciliation rates $R < \mathbb{I}(X; Y)$ are achievable without sources of local randomness and with one-way communication. □

The achievability proof of Proposition 4.5 carries over directly if the roles of Alice and Bob are interchanged. By having Alice estimate Y^n and Bob send information over the public channel, Alice and Bob can recover a sequence of length $n\mathbb{H}(Y)$ while disclosing about $n\mathbb{H}(Y|X)$ bits. The rate of this reconciliation protocol is again on the order of $\mathbb{I}(X;Y)$. Reconciliation protocols for which $S = X^n$ are called *direct* reconciliation protocols, while those for which $S = Y^n$ are called *reverse* reconciliation protocols. Both direct and reverse reconciliation protocols can achieve the reconciliation capacity, but the keys that can be distilled subsequently might not be identical; this issue is discussed further in Section 4.3.3.

Although Proposition 4.5 ensures the existence of reconciliation protocols achieving reconciliation rates arbitrarily close to $\mathbb{I}(X;Y)$, this limit cannot be exactly attained. Any practical finite-length reconciliation protocol introduces an overhead and discloses strictly more than $n\mathbb{H}(X|Y)$ bits over the public channel. It is convenient to account for this overhead by defining the *efficiency* of a reconciliation protocol as follows.

Definition 4.15. *The efficiency of a reconciliation protocol $\mathcal{R}_n$ for a source model with DMS $(\mathcal{XYZ}, p_{XYZ})$ is*

$$\beta \triangleq \frac{\mathbb{H}(S) - r\log(|\mathcal{A}|\,|\mathcal{B}|)}{n\mathbb{I}(X;Y)}. \tag{4.44}$$

The quantity $r\log(|\mathcal{A}|\,|\mathcal{B}|)$ represents the number of bits required in order to describe all messages exchanged over the public channel. Note that $\beta \leqslant 1$ because

$$\begin{aligned}\frac{1}{n}(\mathbb{H}(S) - r\log(|\mathcal{A}|\,|\mathcal{B}|)) &\leqslant \frac{1}{n}(\mathbb{H}(S) - \mathbb{H}(A^r B^r)) \\ &= \mathbf{R}(\mathcal{R}_n) \\ &\leqslant R^{\mathrm{SM}} \\ &= \mathbb{I}(X;Y).\end{aligned}$$

In terms of efficiency, Proposition 4.5 states that there exist reconciliation protocols with efficiency arbitrarily close to one.

Remark 4.6. *With continuous correlated sources, lossless source coding with side information is not possible, since the discrepancies between continuous sources cannot be corrected exactly. Unlike traditional source coding problems, which can be analyzed in a rate-distortion framework, reconciliation* requires *Alice and Bob to agree on a common sequence for further processing. Consequently, the natural way of handling continuous sources is to quantize them to revert back to a discrete case. Assuming again that Alice's randomness X^n is chosen as the common sequence S, Alice can generate a quantized version X_{d}^n of X^n with a scalar quantizer. The upper bound of Proposition 4.5 applies even if Y^n is not quantized, and reconciliation rates can be no greater than $\mathbb{I}(X_{\mathrm{d}};Y)$. By Corollary 2.4, the upper bound can be approached with one-way communication only. Additionally, by choosing a fine enough quantizer, reconciliation rates can be made arbitrarily close to $\mathbb{I}(X;Y)$, and quantization incurs a negligible loss.*

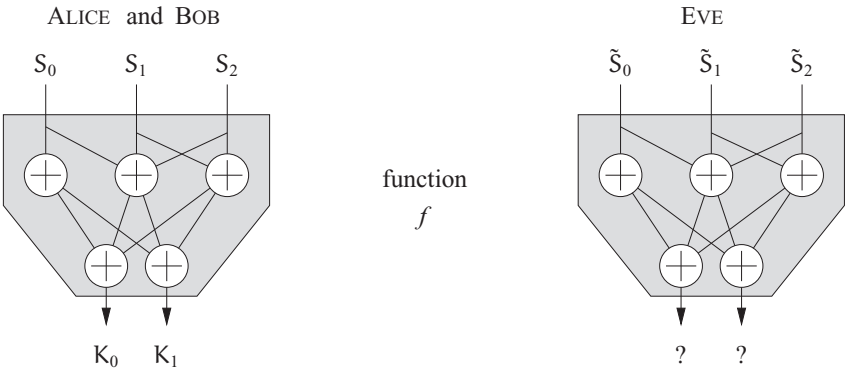

Figure 4.7 Principle of privacy amplification. Alice, Bob, and Eve apply a known function f to their respective sequences S and $\tilde{\mathsf{S}}$. Because of the discrepancies between $\tilde{\mathsf{S}}$ and S, the outputs $\mathsf{K} = f(\mathsf{S})$ and $f(\tilde{\mathsf{S}})$ are different. Eve cannot predict how her errors propagate and, for a well-chosen f, she obtains no information about K.

4.3.3 Privacy amplification

Privacy amplification is the final step of a sequential key-distillation strategy that allows Alice and Bob to distill a secret key. Specifically, the role of privacy amplification is to process the sequence S obtained by Alice and Bob after reconciliation to extract a shorter sequence of k bits that is provably unknown to Eve. Without loss of generality, we assume throughout this section that S is a binary sequence of n bits.

Before we analyze privacy amplification in detail, it is useful to develop an intuitive understanding of why and how this operation is possible. First, note that privacy amplification is straightforward in certain cases. For instance,

- if Alice and Bob know that Eve has no information about S, then the sequence S itself can be used as a key;
- if Alice and Bob know that Eve has access to S, then no secret key can be distilled;
- if Alice and Bob know that Eve has access to m bits of S then the remaining $n - m$ bits of S can be used as a secret key.

In general, privacy amplification is not so simple because Alice and Bob know a bound for Eve's information that cannot be tied to bits of S directly. Nevertheless, we show that this bound is all Alice and Bob need to *extract* bits from S about which Eve has little knowledge. Why this operation is possible can be understood intuitively as follows. For simplicity, assume that Eve computes her best estimate $\tilde{\mathsf{S}}$ of S on the basis of her observations of the source and all messages exchanged over the public channel by the advantage-distillation and reconciliation protocols. Unless Eve's information about S is exactly $\mathbb{H}(\mathsf{S})$, her estimate $\tilde{\mathsf{S}}$ differs from S in some positions. The key idea is that Eve cannot determine the *location* of the discrepancies; consequently, as illustrated in Figure 4.7, if Alice and Bob apply a deterministic transformation f to their sequence S to shuffle and remove some bits, Eve cannot predict how her errors propagate and affect her outcome $f(\tilde{\mathsf{S}})$. We will show that there exist transformations that propagate Eve's

errors so much that all possible outcomes of the transformation f become equally likely from Eve's perspective.

The precise analysis of privacy amplification is slightly involved because it does not rely on the Shannon entropy but on alternative measures called the *collision entropy* and the *min-entropy*. The detailed study of these entropies goes well beyond the scope of this book, and the following section presents only the properties that are useful in the context of secret-key distillation. We refer the interested reader to the bibliographical notes at the end of the chapter for further references.

Collision entropy and min-entropy

The *collision entropy* and the *min-entropy* are convenient metrics because they are tailored to the actual functions used for privacy amplification and because they are more sensitive than the Shannon entropy to deviations from uniform distributions. The latter property allows us to establish the achievability of strong secret-key rates rather than just weak secret-key rates. As an illustration, the following example exhibits a random variable that is not uniform but whose Shannon entropy is hardly distinguishable from that of a uniform random variable.

Example 4.3. Consider a random variable $\mathsf{K} \in [\![1, 2^k]\!]$ with probability distribution

$$\mathbb{P}[\mathsf{K} = 1] = 2^{-k/4} \quad \text{and} \quad \mathbb{P}[\mathsf{K} = i] = \frac{1 - 2^{-k/4}}{2^k - 1} \quad \text{for } i \neq 1.$$

In other words, K is approximately uniform because the realization $\mathsf{K} = 1$ has probability $2^{-k/4}$ while the others have probability $\sim 2^{-k}$. If k is large, this slight non-uniformity is not well captured by the Shannon entropy $\mathbb{H}(\mathsf{K})$. In fact,

$$\begin{aligned}
\mathbb{H}(\mathsf{K}) &= 2^{-k/4}\frac{k}{4} - \left(2^k - 1\right)\frac{1 - 2^{-k/4}}{2^k - 1}\log\left(\frac{1 - 2^{-k/4}}{2^k - 1}\right) \\
&= 2^{-k/4}\frac{k}{4} + \left(1 - 2^{-k/4}\right)\log\left(2^k - 1\right) - \left(1 - 2^{-k/4}\right)\log\left(1 - 2^{-k/4}\right) \\
&= k + 2^{-k/4}\left(\frac{k}{4} - k\right) + \left(1 - 2^{-k/4}\right)\log\left(\frac{1 - 2^{-k}}{1 - 2^{-k/4}}\right).
\end{aligned}$$

Therefore, $\lim_{k\to\infty}(1/k)\mathbb{H}(\mathsf{K}) = 1$, which obscures the fact that the random variable K is not exactly uniform.

Definition 4.16. *The collision entropy of a discrete random variable* $\mathsf{X} \in \mathcal{X}$ *is*

$$\mathbb{H}_{\mathrm{c}}(\mathsf{X}) \triangleq -\log \mathbb{E}[p_{\mathsf{X}}(\mathsf{X})] = -\log\left(\sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x)^2\right).$$

For two discrete random variables $\mathsf{X} \in \mathcal{X}$ *and* $\mathsf{Y} \in \mathcal{Y}$*, the conditional collision entropy of* X *given* Y *is*

$$\mathbb{H}_{\mathrm{c}}(\mathsf{X}|\mathsf{Y}) \triangleq \sum_{y\in\mathcal{Y}} p_{\mathsf{Y}}(y)\, \mathbb{H}_{\mathrm{c}}(\mathsf{X}|\mathsf{Y} = y).$$

The collision entropy bears such a name because it is a function of the collision probability $\sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x)^2$, which measures the probability of obtaining the same realization of a random variable twice in two independent experiments.

Proposition 4.6. *For any discrete random variable* $\mathsf{X} \in \mathcal{X}$*, the collision entropy satisfies* $\mathbb{H}(\mathsf{X}) \geqslant \mathbb{H}_{\mathrm{c}}(\mathsf{X}) \geqslant 0$*. If* X *is uniformly distributed over* $\mathcal{X}$*, then* $\mathbb{H}(\mathsf{X}) = \mathbb{H}_{\mathrm{c}}(\mathsf{X}) = \log|\mathcal{X}|$.

Proof. Using Jensen's inequality and the convexity of the function $x \mapsto -\log x$, we obtain

$$\mathbb{H}(\mathsf{X}) = \mathbb{E}_{\mathsf{X}}[-\log p_{\mathsf{X}}(\mathsf{X})] \geqslant -\log \mathbb{E}_{\mathsf{X}}[p_{\mathsf{X}}(\mathsf{X})] = \mathbb{H}_{\mathrm{c}}(\mathsf{X}).$$

In addition, since $p_{\mathsf{X}}(x) \leqslant 1$ for all $x \in \mathcal{X}$, it holds that

$$\sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x)^2 \leqslant \sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x) = 1,$$

and, therefore, $\mathbb{H}_{\mathrm{c}}(\mathsf{X}) \geqslant 0$. If X is uniformly distributed, then $p_{\mathsf{X}}(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$ and we obtain $\mathbb{H}_{\mathrm{c}}(\mathsf{X}) = \log|\mathcal{X}|$ by direct calculation. □

Remark 4.7. *Many properties of the Shannon entropy* $\mathbb{H}(\mathsf{X})$ *do not hold for the collision entropy* $\mathbb{H}_{\mathrm{c}}(\mathsf{X})$*. For instance, conditioning may* increase *the collision entropy; that is, some random variables are such that* $\mathbb{H}_{\mathrm{c}}(\mathsf{X}|\mathsf{Y}) > \mathbb{H}_{\mathrm{c}}(\mathsf{X})$*. In such cases,* Y *is called the* spoiling knowledge.

Definition 4.17. *The min-entropy of a discrete random variable* $\mathsf{X} \in \mathcal{X}$ *is*

$$\mathbb{H}_{\infty}(\mathsf{X}) = -\log\left(\max_{x\in\mathcal{X}} p_{\mathsf{X}}(x)\right).$$

For two discrete random variables $\mathsf{X} \in \mathcal{X}$ *and* $\mathsf{Y} \in \mathcal{Y}$*, the conditional min-entropy of* X *given* Y *is*

$$\mathbb{H}_{\infty}(\mathsf{X}|\mathsf{Y}) \triangleq \sum_{y\in\mathcal{Y}} p_{\mathsf{Y}}(y)\mathbb{H}_{\infty}(\mathsf{X}|\mathsf{Y}=y).$$

Proposition 4.7. *For any discrete random variable* $\mathsf{X} \in \mathcal{X}$*, the min-entropy satisfies* $\mathbb{H}_{\mathrm{c}}(\mathsf{X}) \geqslant \mathbb{H}_{\infty}(\mathsf{X}) \geqslant 0$*. If* X *is uniformly distributed over* $\mathcal{X}$ *then* $\mathbb{H}_{\infty}(\mathsf{X}) = \mathbb{H}_{\mathrm{c}}(\mathsf{X}) = \mathbb{H}(\mathsf{X}) = \log|\mathcal{X}|$.

Proof. Since $p_{\mathsf{X}}(x) \leqslant 1$ for all $x \in \mathcal{X}$, it is obvious that $\mathbb{H}_{\infty}(\mathsf{X}) \geqslant 0$. Also,

$$\begin{aligned}\mathbb{H}_{\mathrm{c}}(\mathsf{X}) = -\log\left(\sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x)^2\right) &\geqslant -\log\left(\sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x)\max_{x} p_{\mathsf{X}}(x)\right)\\ &= -\log\left(\max_{x} p_{\mathsf{X}}(x)\right) = \mathbb{H}_{\infty}(\mathsf{X}).\end{aligned}$$

If X is uniformly distributed, we obtain $\mathbb{H}_{\infty}(\mathsf{X}) = \log|\mathcal{X}|$ by direct calculation. □

To illustrate that the collision entropy $\mathbb{H}_{\mathrm{c}}(\mathsf{X})$ and the min-entropy $\mathbb{H}_{\infty}(\mathsf{X})$ are indeed stronger measures of uniformity that $\mathbb{H}(\mathsf{X})$, we revisit Example 4.3.

Example 4.4. Consider the random variable $\mathsf{K} \in [\![1, 2^k]\!]$ with distribution

$$\mathbb{P}[\mathsf{K} = 1] = 2^{-k/4} \qquad \text{and} \qquad \mathbb{P}[\mathsf{K} = i] = \frac{1 - 2^{-k/4}}{2^k - 1} \quad \text{for } i \neq 1,$$

for which we showed that $\lim_{k\to\infty}(1/k)\mathbb{H}(\mathsf{K}) = 1$ in Example 4.3. The collision entropy of K is

$$\begin{aligned}\mathbb{H}_{\mathrm{c}}(\mathsf{K}) &= -\log\left(2^{-k/2} + \left(2^k - 1\right)\left(\frac{1 - 2^{-k/4}}{2^k - 1}\right)^2\right) \\ &= \frac{k}{2} - \log\left(\frac{2^k + 2^{k/2} - 2 \times 2^{k/4}}{2^k - 1}\right).\end{aligned}$$

Hence, $\lim_{k\to\infty}(1/k)\mathbb{H}_{\mathrm{c}}(\mathsf{K}) = \frac{1}{2}$.

For k large enough, the min-entropy of K is

$$\mathbb{H}_\infty(\mathsf{K}) = -\log 2^{-k/4} = \frac{k}{4}.$$

Therefore, $\lim_{k\to\infty}(1/k)\mathbb{H}_\infty(\mathsf{K}) = \frac{1}{4}$.

Remark 4.8. *The collision entropy, the min-entropy, and the Shannon entropy are special cases of the Rényi entropy. For a discrete random variable* X*, the Rényi entropy of order* α *is*

$$R_\alpha(\mathsf{X}) \triangleq \frac{1}{1 - \alpha} \log\left(\sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x)^\alpha\right).$$

One can check directly that

$$\mathbb{H}_{\mathrm{c}}(\mathsf{X}) = R_2(\mathsf{X}), \quad \mathbb{H}_\infty(\mathsf{X}) = \lim_{\alpha\to\infty} R_\alpha(\mathsf{X}),$$

and, using l'Hôpital's rule,

$$\mathbb{H}(\mathsf{X}) = \lim_{\alpha\to 1} R_\alpha(\mathsf{X}).$$

Intuitively, the Shannon entropy, the collision entropy and the min-entropy of a random variable play the same role for discrete random variables as the arithmetic mean, geometric mean, and minimum value for a set of numbers.

Privacy amplification with hash functions

In this section, we introduce a generic privacy-amplification technique that exploits hash functions to distill a secret key. From the informal discussion at the beginning of Section 4.3.3, it should not be too surprising that hash functions play a role in privacy amplification. Hash functions are usually designed to produce significantly different outputs even when their inputs are quite similar, which is intuitively the sort of operation that we expect privacy amplification to perform. In what follows, we consider a specific class of hash functions called *universal families*.

Definition 4.18. *Given two finite sets $\mathcal{A}$ and $\mathcal{B}$, a family $\mathcal{G}$ of functions $g : \mathcal{A} \to \mathcal{B}$ is 2-universal (universal for short) if*

$$\forall\, x_1, x_2 \in \mathcal{A} \quad x_1 \neq x_2 \Rightarrow \mathbb{P}_{\mathsf{G}}[\mathsf{G}(x_1) = \mathsf{G}(x_2)] \leqslant \frac{1}{|\mathcal{B}|},$$

where G is the random variable that represents the choice of a function $g \in \mathcal{G}$ uniformly at random in $\mathcal{G}$.

Universal families of hash functions have been thoroughly studied and we provide two well-known examples of families without proving their universality.

Example 4.5. By identifying $\{0,1\}^n$ with $\mathrm{GF}(2)^n$, we associate with any binary matrix $\mathbf{M} \in \mathrm{GF}(2)^{k\times n}$ a function

$$h_{\mathbf{M}} : \mathrm{GF}(2)^n \to \mathrm{GF}(2)^k : \mathbf{x} \mapsto \mathbf{M}\mathbf{x}.$$

The family of hash functions $\mathcal{H}_1 = \left\{h_{\mathbf{M}} : \mathbf{M} \in \mathrm{GF}(2)^{k\times n}\right\}$ is universal.

Example 4.6. By identifying $\{0,1\}^n$ with $\mathrm{GF}(2^n)$, we associate with any element $y \in \mathrm{GF}(2^n)$ a function

$$h_y : \mathrm{GF}(2^n) \to \{0,1\}^k : x \mapsto k \text{ bits of the product } xy.$$

The k bits are fixed but their position can be chosen arbitrarily. The family of hash functions $\mathcal{H}_2 = \left\{h_y : y \in \mathrm{GF}(2^n)\right\}$ is universal.

Remark 4.9. *Identifying a function in the family $\mathcal{H}_1$ requires nk bits, while identifying a function in the family $\mathcal{H}_2$ requires only n bits. This difference does not affect the operation of privacy amplification, but, in practice, it is often desirable to limit the amount of communication; hence, the family $\mathcal{H}_2$ would be preferred to the family $\mathcal{H}_1$.*

The usefulness of universal families of hash functions for privacy amplification is justified in the following theorem.

Theorem 4.4 (Bennett *et al.*)**.** *Let $\mathsf{S} \in \{0,1\}^n$ be the random variable that represents the common sequence shared by Alice and Bob, and let E be the random variable that represents the total knowledge about S available to Eve. Let e be a particular realization of E. If Alice and Bob know the conditional collision entropy $\mathbb{H}_c(\mathsf{S}|\mathsf{E} = e)$ to be at least some constant c, and if they choose $\mathsf{K} = \mathsf{G}(\mathsf{S})$ as their secret key, where G is a hash function chosen uniformly at random from a universal family of hash functions $\mathcal{G} : \{0,1\}^n \to \{0,1\}^k$, then*

$$\mathbb{H}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e) \geqslant k - \frac{2^{k-c}}{\ln 2}.$$

Proof. Since $\mathbb{H}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e) \geqslant \mathbb{H}_{\mathrm{c}}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e)$ by Proposition 4.6, it suffices to establish a lower bound for the collision entropy $\mathbb{H}_{\mathrm{c}}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e)$. Note that

$$\begin{aligned}\mathbb{H}_{\mathrm{c}}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e) &= \sum_{g \in \mathcal{G}} p_{\mathsf{G}}(g) \mathbb{H}_{\mathrm{c}}(\mathsf{K}|\mathsf{G} = g, \mathsf{E} = e) \\ &= \sum_{g \in \mathcal{G}} p_{\mathsf{G}}(g) \left(-\log \mathbb{E}_{\mathsf{K}|\mathsf{G}=g, \mathsf{E}=e}\left[p_{\mathsf{K}|\mathsf{G}\mathsf{E}}(\mathsf{K}|g, e)\right]\right) \\ &\geqslant -\log \left(\sum_{g \in \mathcal{G}} p_{\mathsf{G}}(g) \mathbb{E}_{\mathsf{K}|\mathsf{G}=g, \mathsf{E}=e}\left[p_{\mathsf{K}|\mathsf{G}\mathsf{E}}(\mathsf{K}|g, e)\right] \right), \end{aligned} \tag{4.45}$$

where the last inequality follows from the convexity of the function $x \mapsto -\log x$ and Jensen's inequality. Now, let $\mathsf{S}_1 \in \{0, 1\}^n$ and $\mathsf{S}_2 \in \{0, 1\}^n$ be two random variables that are independent of each other and independent of G, which are distributed according to $p_{\mathsf{S}|\mathsf{E}=e}$. Then,

$$\begin{aligned}\mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2)|\mathsf{G} = g] &= \sum_{k \in \{0,1\}^k} p_{\mathsf{G}(\mathsf{S})|\mathsf{G}\mathsf{E}}(k|g, e) p_{\mathsf{G}(\mathsf{S})|\mathsf{G}\mathsf{E}}(k|g, e) \\ &= \mathbb{E}_{\mathsf{K}|\mathsf{G}=g, \mathsf{E}=e}\left[p_{\mathsf{K}|\mathsf{G}\mathsf{E}}(\mathsf{K}|g, e)\right], \end{aligned}$$

and we can rewrite inequality (4.45) as

$$\mathbb{H}_{\mathrm{c}}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e) \geqslant -\log \mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2)]. \tag{4.46}$$

We now develop an upper bound for $\mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2)]$. By the law of total probability,

$$\begin{aligned}\mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2)] &= \mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2), \mathsf{S}_1 = \mathsf{S}_2]\mathbb{P}[\mathsf{S}_1 = \mathsf{S}_2] \\ &\quad + \mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2), \mathsf{S}_1 \neq \mathsf{S}_2]\mathbb{P}[\mathsf{S}_1 \neq \mathsf{S}_2]. \end{aligned} \tag{4.47}$$

Note that

$$\mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2)|\mathsf{E} = e, \mathsf{S}_1 = \mathsf{S}_2] \leqslant 1 \quad \text{and} \quad \mathbb{P}[\mathsf{S}_1 \neq \mathsf{S}_2|\mathsf{E} = e] \leqslant 1.$$

In addition, by virtue of the definition of the collision entropy,

$$\mathbb{P}[\mathsf{S}_1 = \mathsf{S}_2] = \sum_{s \in \{0,1\}^n} p_{\mathsf{S}|\mathsf{E}=e}(s|e)^2 = 2^{-\mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{E}=e)}.$$

Finally, because the hash function $\mathcal{G}$ is chosen in a universal family, it holds that

$$\mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2)|\mathsf{S}_1 \neq \mathsf{S}_2] \leqslant 2^{-k}.$$

On substituting these inequalities into (4.47), we obtain

$$\mathbb{P}[\mathsf{G}(\mathsf{S}_1) = \mathsf{G}(\mathsf{S}_2)] \leqslant 2^{-\mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{E}=e)} + 2^{-k} \leqslant 2^{-k}(1 + 2^{k-c}), \tag{4.48}$$

where the last inequality follows from the assumption $\mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{E} = e) \geqslant c$. On substituting (4.48) into (4.46) and using the fact that $\ln(1 + x) \leqslant x$ for all $x > -1$, we obtain

$$\mathbb{H}_{\mathrm{c}}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e) \geqslant k - \frac{2^{k-c}}{\ln 2}. \qquad \square$$

Since $\mathbb{H}_c(\mathsf{K}|\mathsf{G}, \mathsf{E} = e) \leqslant k$ by definition, Theorem 4.4 states that it is possible to distill a secret key of k bits with hash functions, provided that their output size is small enough ($k \ll c$). The output sequence size depends directly on the a-priori uncertainty of the eavesdropper, which must be measured in terms of the collision entropy. Since this result is the essential tool that we use to analyze the achievable secret-key rates of sequential secret-key distillation strategies, it is worthwhile discussing it in detail. Note that, although the hash function G is chosen at random, the actual choice is known to Eve and this is reflected by the conditioning on G in the entropy; however, the theorem provides only a lower bound on $\mathbb{H}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e)$, which is an *average* over all possible choices of hash functions in $\mathcal{G}$. Consequently, for a specific choice of $g \in \mathcal{G}$, the entropy $\mathbb{H}(\mathsf{K}|\mathsf{G} = g, \mathsf{E} = e)$ might be significantly different from k, even if $k \ll c$; luckily, this happens with negligible probability. Most importantly, Theorem 4.4 provides an *explicit* privacy-amplification technique because it shows that it suffices to choose a hash function at random in a universal family. This is in sharp contrast with the random-coding argument used in Section 4.2.2, which guarantees only the *existence* of a suitable key-distillation function. Finally, we emphasize that Theorem 4.4 bounds the entropy $\mathbb{H}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e)$, which is a stronger result than a bound on the entropy rate $(1/n)\mathbb{H}(\mathsf{K}|\mathsf{G}, \mathsf{E} = e)$.

We now have all the tools to prove that sequential key-distillation strategies achieve strong secret-key rates.

Theorem 4.5. *Consider a source model with DMS* $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$ *and let* $\beta \in [0, 1]$. *All strong secret-key rates* R_s *that satisfy*

$$R_s < \beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \min(\mathbb{I}(\mathsf{X};\mathsf{Z}), \mathbb{I}(\mathsf{Y};\mathsf{Z}))$$

are achievable with sequential secret-key distillation strategies that consist of a reconciliation protocol with efficiency β *and privacy amplification with a universal family of hash functions. Additionally, these rates are achievable with one-way communication.*

Note that the secret-key rates given in Theorem 4.5 are achievable without advantage distillation. In addition, Theorem 4.5 shows that reconciliation efficiency acts as a penalty factor that reduces the information between Alice and Bob from $\mathbb{I}(\mathsf{X};\mathsf{Y})$ to $\beta\mathbb{I}(\mathsf{X};\mathsf{Z})$ but leaves the information leaked to the eavesdropper $\mathbb{I}(\mathsf{X};\mathsf{Z})$ or $\mathbb{I}(\mathsf{Y};\mathsf{Z})$ unchanged. This result is particularly relevant because any practical reconciliation protocol has an efficiency $\beta < 1$, which turns out to be one of the main limiting factors of secret-key rates. Although Proposition 4.5 guarantees the existence of reconciliation protocols with β arbitrarily close to unity, the design of efficient protocols can be challenging. The construction of practical yet efficient reconciliation protocols is discussed in greater detail in Chapter 6.

The proof of Theorem 4.5 is involved because we need to establish a lower bound for Eve's collision entropy about Alice and Bob's common sequence S before we can apply Theorem 4.4. There is no obvious bound since Eve's total knowledge consists not only of the observations of the source Z^n but also of the public messages A^r and B^r exchanged during the reconciliation phase.

Proof of Theorem 4.5. Fix an integer n and let $\epsilon > 0$. Let k be an integer to be determined later. We consider a sequential key-distillation strategy $\mathcal{S}_n$ that consists of

- a direct reconciliation protocol $\mathcal{R}_n$ with efficiency β, with which Alice sends messages A^r to Bob; Theorem 4.5 guarantees the existence of a protocol so that Bob's estimate $\hat{X}^n$ of X^n satisfies $\mathbb{P}[\hat{X}^n \neq X^n | \mathcal{R}_n] \leqslant \delta_\epsilon(n)$;
- privacy amplification based on a universal family of hash functions with output size k, at the end of which Alice computes her key $K = G(X^n)$, while Bob computes $\hat{K} = G(\hat{X}^n)$.

Note that $\mathbb{P}[\hat{K} \neq K | \mathcal{S}_n] \leqslant \mathbb{P}[\hat{X}^n \neq X^n | \mathcal{R}_n] \leqslant \delta_\epsilon(n)$ and that the strategy uses only one-way communication from Alice to Bob. The total information available to Eve after reconciliation consists of her observation Z^n of the DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$, the public messages A^r exchanged during the reconciliation protocol, and the hash function G chosen for privacy amplification. The strategy $\mathcal{S}_n$ is also known to Eve, but we omit the conditioning on $\mathcal{S}_n$ in order to simplify the notation. We show that, for a suitable choice of the output size k,

$$k \geqslant \mathbb{H}(K|Z^n A^r G) \geqslant k - \delta_\epsilon(n).$$

This result will follow from Theorem 4.4, provided that we establish a lower bound for the collision entropy $\mathbb{H}_c(X^n|Z^n = z^n, A^r = a^r)$. This is not straightforward because the collision entropy depends on the specific operation of the reconciliation protocol; nevertheless, we circumvent this difficulty in two steps.

First, we will relate $\mathbb{H}_c(X^n|Z^n = z^n, A^r = a^r)$ to $\mathbb{H}_c(X^n|Z^n = z^n)$ by means of the following lemma, whose proof is relegated to the appendix at the end of this chapter.

Lemma 4.3 (Cachin)**.** *Let $S \in \mathcal{S}$ and $U \in \mathcal{U}$ be two discrete random variables with joint distribution p_{SU}. For any $r > 0$, define the function $\chi : \mathcal{U} \to \{0, 1\}$ as*

$$\chi(u) \triangleq \begin{cases} 1 & \text{if } \mathbb{H}_c(S) - \mathbb{H}_c(S|u) \leqslant \log|\mathcal{U}| + 2r + 2, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\mathbb{P}_U[\chi(U) = 1] \geqslant 1 - 2^{-r}$.

Lemma 4.3 shows that, with high probability, the decrease in collision entropy caused by conditioning on U is bounded by $\log|\mathcal{U}| + 2r + 2$, which does not depend on the exact correlation between S and U.

Next, we will lower bound $\mathbb{H}_c(X^n|Z^n = z^n)$ by a term on the order of $n\mathbb{H}(X|Z)$. Essentially, this result follows from the fact that, for n large enough, the realizations of the random source (X^n, Z^n) are typical and almost uniformly distributed in the joint typical set. This idea is formalized in the following lemma, whose proof is again relegated to the appendix.

Lemma 4.4. *Consider a DMS $(\mathcal{X}\mathcal{Z}, p_{XZ})$ and define the random variable Θ as*

$$\Theta = \begin{cases} 1 & \text{if } \quad (X^n, Z^n) \in \mathcal{T}^n_{2\epsilon}(XZ) \quad \text{and} \quad Z^n \in \mathcal{T}^n_\epsilon(XZ), \\ 0 & \text{otherwise.} \end{cases}$$

Then, for n sufficiently large, $\mathbb{P}[\Theta = 1] \geqslant 1 - 2^{-\sqrt{n}}$. *Moreover, if* $z^n \in \mathcal{T}_\epsilon^n(\mathsf{Z})$,

$$\begin{aligned}\mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1) &\geqslant \mathbb{H}_\infty(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1)\\ &\geqslant n(\mathbb{H}(\mathsf{X}|\mathsf{Z}) - \delta(\epsilon)) - \delta_\epsilon(n).\end{aligned}$$

To leverage the results of Lemma 4.3 and Lemma 4.4, we start by defining the random variables Υ (a function of A^r) and Θ (a function of X^n and Z^n) as follows:

$$\Upsilon \triangleq \begin{cases} 1 & \text{if} \quad \mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n) - \mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \mathsf{A}^r) \leqslant \log|\mathcal{A}|^r + 2\sqrt{n} + 2,\\ 0 & \text{otherwise;}\end{cases}$$

$$\Theta \triangleq \begin{cases} 1 & \text{if} \quad (\mathsf{X}^n, \mathsf{Z}^n) \in \mathcal{T}_{2\epsilon}^n(\mathsf{XZ}) \quad \text{and} \quad \mathsf{Z}^n \in \mathcal{T}_\epsilon^n(\mathsf{XZ}),\\ 0 & \text{otherwise.}\end{cases}$$

Lemma 4.3 guarantees that $\mathbb{P}[\Upsilon = 1] \geqslant 1 - 2^{-\sqrt{n}}$ and Lemma 4.4 ensures that $\mathbb{P}[\Theta = 1] \geqslant 1 - 2^{-\sqrt{n}}$; therefore, by the union bound,

$$\mathbb{P}[\Theta = 1, \Upsilon = 1] \geqslant 1 - 2 \cdot 2^{-\sqrt{n}}.$$

Consequently, we can lower bound $\mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r)$ as

$$\begin{aligned}\mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r) &\geqslant \mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r\Theta\Upsilon)\\ &\geqslant \mathbb{P}[\Theta = 1, \Upsilon = 1]\mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r, \Theta = 1, \Upsilon = 1)\\ &\geqslant \left(1 - 2 \cdot 2^{-\sqrt{n}}\right) \mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r, \Theta = 1, \Upsilon = 1). \qquad (4.49)\end{aligned}$$

To bound $\mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r, \Theta = 1, \Upsilon = 1)$ with Theorem 4.4 it suffices to lower bound the collision entropy

$$\mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{G}, \mathsf{Z}^n = z^n, \mathsf{A}^r = a^r, \Theta = 1, \Upsilon = 1)$$

for any realization $z^n \in \mathcal{T}_\epsilon^n(\mathsf{Z})$ and a^r. By virtue of the definition of Υ,

$$\begin{aligned}&\mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \mathsf{A}^r = a^r, \Theta = 1, \Upsilon = 1)\\ &\qquad \geqslant \mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1) - \log|\mathcal{A}|^r - 2\sqrt{n} - 2. \qquad (4.50)\end{aligned}$$

The quantity $\log|\mathcal{A}|^r$ represents the number of bits required to describe the messages exchanged during the reconciliation phase, which we can express in terms of the reconciliation efficiency β as

$$\log|\mathcal{A}|^r = \mathbb{H}(\mathsf{X}^n) - n\beta\mathbb{I}(\mathsf{X};\mathsf{Y}).$$

Therefore, we can rewrite (4.50) as

$$\begin{aligned}&\mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \mathsf{A}^r = a^r, \Theta = 1, \Upsilon = 1)\\ &\qquad \geqslant \mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1) - \mathbb{H}(\mathsf{X}^n) + n\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - 2\sqrt{n} - 2. \qquad (4.51)\end{aligned}$$

Next, notice that Lemma 4.4 ensures

$$\mathbb{H}_\mathrm{c}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1) \geqslant n(\mathbb{H}(\mathsf{X}|\mathsf{Z}) - \delta(\epsilon)) - \delta_\epsilon(n). \qquad (4.52)$$

Combining (4.51) and (4.52) yields

$$\begin{aligned}\mathbb{H}_{\epsilon}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \mathsf{A}^r = a^r, \Theta = 1, \Upsilon = 1) \\ &\geqslant n(\mathbb{H}(\mathsf{X}|\mathsf{Z}) - \delta(\epsilon)) - \mathbb{H}(\mathsf{X}^n) + n\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - 2\sqrt{n} - 2 - \delta_{\epsilon}(n) \\ &= n(\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z}) - \delta(\epsilon)) - 2\sqrt{n} - 2 - \delta_{\epsilon}(n). \end{aligned} \tag{4.53}$$

Hence, if we set the output size of the hash function k to be less than the lower bound in (4.53) by $\sqrt{n}$, say

$$k \triangleq \lfloor n(\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z}) - \delta(\epsilon)) - 3\sqrt{n} - 2 - \delta_{\epsilon}(n)\rfloor, \tag{4.54}$$

then Theorem 4.4 ensures that

$$\begin{aligned}\mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n = z^n, \mathsf{A}^r = a^r, \Theta = 1, \Upsilon = 1) &\geqslant k - \frac{2^{-\sqrt{n}}}{\ln 2} \\ &= k - \delta(n). \end{aligned} \tag{4.55}$$

On substituting (4.55) back into (4.49), we finally obtain

$$\mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r) \geqslant \left(1 - 2\cdot 2^{-\sqrt{n}}\right)(k - \delta(n)) = k - \delta(n), \tag{4.56}$$

and, consequently,

$$\mathbb{I}(\mathsf{K};\mathsf{G}\mathsf{Z}^n\mathsf{A}^r) = \mathbb{H}(\mathsf{K}) - \mathbb{H}(\mathsf{K}|\mathsf{G}\mathsf{Z}^n\mathsf{A}^r) \leqslant \delta(n).$$

Notice that the corresponding secret-key rate is

$$R \triangleq \frac{k}{n} = \beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z}) - \delta(\epsilon) - \delta(n). \tag{4.57}$$

Hence, we have proved the existence of a $(2^{nR}, n)$ sequential key-distillation strategy $\mathcal{S}_n$ with rate given by (4.57) and based on a direct reconciliation protocol of efficiency β and privacy amplification with a universal family of hash functions such that

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \leqslant \delta_{\epsilon}(n), \quad \mathbf{L}(\mathcal{S}_n) \leqslant \delta_{\epsilon}(n), \quad \text{and} \quad \mathbf{U}(\mathcal{S}_n) \leqslant \delta_{\epsilon}(n).$$

Hence, $\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z}) - \delta(\epsilon)$ is an achievable strong secret-key rate. Since ϵ can be chosen arbitrarily small, all strong secret-key rates below $\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z})$ are achievable, as well.

The achievability of all strong secret-key rates below $\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z})$ follows from the same arguments on reversing the roles of Alice and Bob and considering a reverse reconciliation protocol. □

Although both direct and reverse reconciliation protocols can operate close to the reconciliation capacity R^{SM}, the secret-key rates obtained after privacy amplification are different. In the proof of Theorem 4.5, the secret-key rates below $\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z})$ are achievable with a direct reconciliation protocol, whereas the secret-key rates below $\beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z})$ are achievable with a reverse reconciliation protocol. Intuitively, a direct reconciliation protocol uses Alice's observations as the reference to distill the key and the information leaked to Eve is therefore $\mathbb{I}(\mathsf{X};\mathsf{Z})$. In contrast, a reverse reconciliation uses Bob's observations as the reference and the information leaked to Eve is then $\mathbb{I}(\mathsf{Y};\mathsf{Z})$.

Remark 4.10. *Privacy amplification with a universal family of hash functions is a special case of privacy amplification with an almost universal family of hash functions. For $\gamma \geqslant 1$, a γ-almost universal family $\mathcal{G}$ of hash functions $g{:}\mathcal{A} \to \mathcal{B}$ is such that*

$$\forall x_1, x_2 \in \mathcal{A} \quad x_1 \neq x_2 \Rightarrow \mathbb{P}_{\mathsf{G}}[\mathsf{G}(x_1) = \mathsf{G}(x_2)] \leqslant \frac{\gamma}{|\mathcal{B}|},$$

where G *is the random variable that represents the choice of a function g uniformly at random in $\mathcal{G}$. It is possible to show that sequential key-distillation strategies based on γ-almost universal families of hash functions can achieve all strong secret-key rates below*

$$\beta \mathbb{I}(\mathsf{X};\mathsf{Y}) - \min(\mathbb{I}(\mathsf{X};\mathsf{Z}), \mathbb{I}(\mathsf{Y};\mathsf{Z})) - \log \gamma,$$

which is lower than the rate given in Theorem 4.5 if $\gamma > 1$. Nevertheless, almost-universal families of hash functions might be preferred to universal families of hash functions in practice because of their greater flexibility and lower complexity. A more detailed discussion of privacy amplification based on almost-universal families of hash functions can be found in the textbook of Van Assche [38].

Corollary 4.2. *The strong secret-key capacity $\overline{C}_{\mathrm{s}}^{\mathsf{SM}}$ of a source model with DMS $(\mathcal{XYZ}, p_{\mathsf{XYZ}})$ is equal to its weak secret-key capacity $C_{\mathrm{s}}^{\mathsf{SM}}$. In addition, all strong secret-key rates below $\overline{C}_{\mathrm{s}}^{\mathsf{SM}}$ are achievable with sequential key-distillation strategies.*

Proof. This result follows directly from Proposition 4.2 and Theorem 4.5. According to Proposition 4.2, an advantage-distillation protocol $\mathcal{D}_n$ can be used to transform n realizations of the original DMS $(\mathcal{XYZ}, p_{\mathsf{XYZ}})$ into a single realization of a new DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$ for which

$$\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}') > 0 \quad \text{or} \quad \mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{Y}';\mathsf{Z}') > 0.$$

For this new source and any $\epsilon > 0$, Theorem 4.5 ensures that the strong secret-key rate

$$R'_{\mathrm{s}} = \max\big(\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}'), \mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{Y}';\mathsf{Z}')\big) - \delta(\epsilon)$$

is achievable, where R'_{s} is expressed in bits per observation of the DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$. The rate in bits per observation of the original DMS $(\mathcal{XYZ}, p_{\mathsf{XYZ}})$ is then

$$\begin{aligned} R_{\mathrm{s}} &= \frac{1}{n} \max\big(\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}'), \mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{Y}';\mathsf{Z}')\big) - \delta(\epsilon) \\ &= \mathbf{R}(\mathcal{D}_n) - \delta(\epsilon), \end{aligned}$$

which can be made as close as desired to the advantage-distillation capacity D^{SM}. By Proposition 4.2, $D^{\mathsf{SM}} = C_{\mathrm{s}}^{\mathsf{SM}}$ and, because R_{s} is a strong secret-key rate, we obtain $\overline{C}_{\mathrm{s}}^{\mathsf{SM}} \geqslant C_{\mathrm{s}}^{\mathsf{SM}}$ and hence $\overline{C}_{\mathrm{s}}^{\mathsf{SM}} = C_{\mathrm{s}}^{\mathsf{SM}}$. □

Theorem 4.5 and Corollary 4.2 have far-reaching implications. First, the fact that the strong secret-key capacity $\overline{C}_{\mathrm{s}}^{\mathsf{SM}}$ is equal to the weak secret-key capacity $C_{\mathrm{s}}^{\mathsf{SM}}$ is reassuring because it suggests that the fundamental limit of secret-key generation is fairly robust with respect to the actual choice of secrecy condition. Second, since sequential-key distillation strategies can achieve all rates below $C_{\mathrm{s}}^{\mathsf{SM}}$, we know that there is no loss of

optimality in handling the reliability and secrecy requirements independently. Whether sequential key-distillation strategies can achieve secret-key rates close to $C_{\mathrm{s}}^{\mathrm{SM}}$ depends on our ability to design advantage-distillation protocols, but at least we have an explicit procedure for privacy amplification, and we will see in Chapter 6 how to design efficient reconciliation protocols. Finally, the proof of Corollary 4.2 highlights the importance of two-way communications for secret-key agreement as a means to distill an advantage over the eavesdropper. Two-way communication is required for advantage distillation, although reconciliation and privacy amplification can be implemented with one-way communication only.

Privacy amplification with extractors

We conclude our study of privacy amplification by discussing an alternative to universal families of hash functions and by analyzing privacy amplification with a class of functions called *extractors*. In essence, the analysis and the results are identical to those obtained earlier, but they exploit the min-entropy in place of the collision entropy. The only difference between privacy amplification with hash functions and privacy amplification with extractors is the amount of communication over the public channel.

Definition 4.19. *A function $g : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ is called a (γ, ϵ)-extractor if, for any random variable $\mathsf{S} \in \{0,1\}^n$ with min-entropy $\mathbb{H}_\infty(\mathsf{S}) \geqslant \gamma n$ and a random variable U_d uniformly distributed over $\{0,1\}^d$, the variational distance between the random variable $(\mathsf{U}_d, g(\mathsf{S}, \mathsf{U}_d)) \in \{0,1\}^{d+k}$ and the random variable U_{d+k} with uniform distribution over $\{0,1\}^{d+k}$ satisfies*

$$\mathbb{V}((\mathsf{U}_d, g(\mathsf{S}, \mathsf{U}_d)), \mathsf{U}_{d+k}) \leqslant \epsilon.$$

In other words, an extractor is a function that converts a sequence of n bits with arbitrary distribution into a sequence of k bits with almost uniform distribution using d bits of randomness as a catalyst. If $d \leqslant k$, that is the extractor outputs more uniform randomness than is used at the input, this operation can be thought of as a way of "extracting" uniform randomness from the random variable S. In practice, an extractor is useful if $d \ll k$, which means that little extra randomness is necessary for the extraction. The existence of such extractors is guaranteed by the following proposition, which we state without proof.

Proposition 4.8 (Vadhan)**.** *For any $\epsilon > 0$, $\gamma \in (0,1)$, there exists a (γ, ϵ)-extractor $g : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ with*

$$k = \gamma n - 2\log\left(\frac{1}{\epsilon}\right) - O(1)$$

and

$$d = O\left(\left(\log\left(\frac{n}{\epsilon}\right)\right)^2 \log(\gamma n)\right).$$

In other words, for n large enough, there exist extractors that extract almost the entire min-entropy of the input S and require a comparatively *negligible* amount of uniform

randomness. The existence of such extractors allows us to prove the counterpart of Theorem 4.4 with extractors in place of hash functions.

Theorem 4.6 (Maurer and Wolf)**.** *Let* $\mathsf{S} \in \{0,1\}^n$ *be the random variable that represents the common sequence shared by Alice and Bob, and let* E *be the random variable that represents the total knowledge about* S *available to Eve. Let* e *be a particular realization of* E*. If Alice and Bob know the conditional min-entropy* $\mathbb{H}_\infty(\mathsf{S}|\mathsf{E}=e)$ *to be at least* γn *for some* $\gamma \in (0,1)$*, then there exists a function* $g : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ *with*

$$d \leqslant n\delta(n) \quad \text{and} \quad k \geqslant n(\gamma - \delta(n)),$$

such that, if U_d *is a random variable with uniform distribution on* $\{0,1\}^d$ *and Alice and Bob choose* $\mathsf{K} = g(\mathsf{S}, \mathsf{U}_d)$ *as their secret key, then*

$$\mathbb{H}(\mathsf{K}|\mathsf{U}_d, \mathsf{E}=e) \geqslant k - \delta(n).$$

Proof. Let $\epsilon \triangleq 2^{-\sqrt{n}/\log n}$ and let $\mathsf{K}_e \in \{0,1\}^k$ be the random variable with distribution

$$p_{\mathsf{K}_e} \triangleq p_{g(\mathsf{S},\mathsf{U}_d)|\mathsf{E}=e}.$$

According to Proposition 4.8, there exists a (γ, ϵ) extractor $g : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ with

$$d = O\left(\left(\log\left(\frac{n}{\epsilon}\right)\right)^2 \log(\gamma n)\right) = n\delta(n),$$

$$k = \gamma n - 2\log\left(\frac{1}{\epsilon}\right) - O(1) = n(\gamma - \delta(n)),$$

and such that $\mathbb{V}((\mathsf{U}_d, \mathsf{K}_e), \mathsf{U}_{d+k}) \leqslant 2^{-\sqrt{n}/\log n}$. Note that we can write the uniform distribution $p_{\mathsf{U}_{d+k}}$ over $\{0,1\}^{d+k}$ as the product of the uniform distribution p_{U_d} over $\{0,1\}^d$ with the uniform distribution p_{U_k} over $\{0,1\}^k$; hence,

$$\begin{aligned}
\mathbb{V}((\mathsf{U}_d, \mathsf{K}_e), \mathsf{U}_{d+k}) &= \sum_{u,s} \left| p_{\mathsf{U}_d}(u) p_{\mathsf{K}_e|\mathsf{U}_d}(s|u) - p_{\mathsf{U}_d}(u) p_{\mathsf{U}_k}(s) \right| \\
&= \sum_{u,s} p_{\mathsf{U}_d}(u) \left| p_{\mathsf{K}_e|\mathsf{U}_d}(s|u) - 2^{-k} \right| \\
&= \mathbb{E}_{\mathsf{U}_d}\left[\sum_s \left| p_{\mathsf{K}_e|\mathsf{U}_d}(s|\mathsf{U}_d) - 2^{-k} \right| \right].
\end{aligned}$$

By Markov's inequality,

$$\begin{aligned}
\mathbb{P}_{\mathsf{U}_d}\left[\sum_s \left| p_{\mathsf{K}_e|\mathsf{U}_d}(s|\mathsf{U}_d) - 2^{-k} \right| \geqslant 2^{-\sqrt{n}/(2\log n)} \right] &\leqslant \frac{\mathbb{E}_{\mathsf{U}_d}\left[\sum_s \left| p_{\mathsf{K}_e|\mathsf{U}_d}(s|\mathsf{U}_d) - 2^{-k} \right| \right]}{2^{-\sqrt{n}/(2\log n)}} \\
&\leqslant 2^{-\sqrt{n}/(2\log n)}. \qquad (4.58)
\end{aligned}$$

In other words, with high probability, the realization u_d of U_d is such that the variational distance between $p_{\mathsf{K}_e|\mathsf{U}_d}$ and a uniform distribution over $\{0,1\}^k$ is small. Formally, if we

define the random variable Θ function of U_d as

$$\Theta = \begin{cases} 1 & \text{if } \mathbb{V}((\mathsf{U}_d, \mathsf{K}_e), \mathsf{U}_{d+k}) \leqslant 2^{-\sqrt{n}/(2\log n)}, \\ 0 & \text{otherwise.} \end{cases}$$

then, by (4.58),

$$\mathbb{P}[\Theta = 1] \geqslant 1 - 2^{-\sqrt{n}/\log n}. \tag{4.59}$$

We now lower bound $\mathbb{H}(\mathsf{K}|\mathsf{U}_d, \mathsf{E} = e)$ as

$$\begin{aligned} \mathbb{H}(\mathsf{K}|\mathsf{U}_d, \mathsf{E} = e) &\geqslant \mathbb{H}(\mathsf{K}|\mathsf{U}_d, \Theta, \mathsf{E} = e) \\ &\geqslant \mathbb{P}[\Theta = 1]\mathbb{H}(\mathsf{K}_e|\mathsf{U}_d, \Theta = 1). \end{aligned} \tag{4.60}$$

Given $\Theta = 1$, the variational distance $\mathbb{V}(\mathsf{U}_d, \mathsf{K}_e; \mathsf{U}_{d+k})$ is less than $2^{-\sqrt{n}/\log n}$ and note that the function $x \mapsto x \log(2^k/x)$ is increasing for $x \in (0, 2^{k-1})$; therefore, for n large enough, we have, according to Proposition 2.1,

$$\begin{aligned} |\mathbb{H}(\mathsf{K}_e|\mathsf{U}_d, \Theta = 1) - k| &\leqslant 2^{-\sqrt{n}/\log n} \log\left(\frac{2^k}{2^{-\sqrt{n}/\log n}}\right) \\ &= 2^{-\sqrt{n}/\log n}\left(k + \frac{\sqrt{n}}{\log n}\right) \\ &= \delta(n). \end{aligned} \tag{4.61}$$

On combining (4.59) and (4.61) in (4.60), we finally obtain

$$\mathbb{H}(\mathsf{K}|\mathsf{U}_d, \mathsf{E} = e) \geqslant k - \delta(n),$$

which is the desired result. □

We can now establish the counterpart of Theorem 4.5.

Theorem 4.7. *Consider a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$ and let $\beta \in [0, 1]$. All strong secret-key rates R_{s} that satisfy*

$$R_{\mathrm{s}} < \beta\mathbb{I}(\mathsf{X};\mathsf{Y}) - \min(\mathbb{I}(\mathsf{X};\mathsf{Z}), \mathbb{I}(\mathsf{Y};\mathsf{Z}))$$

are achievable with sequential secret-key distillation strategies that consist of a reconciliation protocol with efficiency β and privacy amplification with extractors. Additionally, these rates are achievable with one-way communication.

Sketch of proof. The proof of Theorem 4.7 follows exactly that of Theorem 4.5. The only difference is the use of extractors instead of hash functions, which requires the use of the min-entropy instead of the collision entropy. Notice that Lemma 4.4 already establishes a lower bound for the min-entropy, hence we need just the counterpart of Lemma 4.3 for the min-entropy. As shown in the appendix to this chapter, the following result holds.

Lemma 4.5. *Let $\mathsf{S} \in \mathcal{S}$ and $\mathsf{U} \in \mathcal{U}$ be two discrete random variables with joint distribution p_{SU}. For any $r > 0$, define the function $\chi : \mathcal{U} \to \{0, 1\}$ as*

$$\chi(u) \triangleq \begin{cases} 1 & \text{if } \mathbb{H}_\infty(\mathsf{S}) - \mathbb{H}_\infty(\mathsf{S}|u) \leqslant \log|\mathcal{U}| + r, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\mathbb{P}_{\mathsf{U}}[\chi(\mathsf{U}) = 1] \geqslant 1 - 2^{-r}$.

Essentially, Lemma 4.5 states that, with high probability, conditioning on U reduces the min-entropy by at most on the order of $\log|\mathcal{U}|$. This allows us to relate the min-entropy before and after reconciliation independently of the exact operation of the reconciliation protocol. □

4.4 Secret-key capacity of the channel model

In this section, we turn our attention to the channel model for secret-key agreement and study the secret-key capacity of a channel model. The secret-key capacity of a channel model is sometimes called the *secrecy capacity with public discussion*, because a channel model can be viewed as a WTC enhanced by a public channel between Alice and Bob. However, this denomination is slightly misleading because, in contrast to the secrecy capacity, the secrecy capacity with public discussion characterizes a secret-key rate, not a secure message rate. In this book, we restrict ourselves to the term secret-key capacity and it will be clear from the context whether this refers to a source model or a channel model.

Definition 4.20. *The weak secret-key capacity of a channel model with DMC* $(\mathcal{X}, p_{\mathsf{YZ}|\mathsf{X}}, \mathcal{Y}, \mathcal{Z})$ *is*

$$C_{\mathrm{s}}^{\mathrm{CM}} \triangleq \sup\{R : R \text{ is an achievable weak secret-key rate}\}.$$

Similarly, the strong secret-key capacity is

$$\overline{C}_{\mathrm{s}}^{\mathrm{CM}} \triangleq \sup\{R : R \text{ is an achievable strong secret-key rate}\}.$$

It follows from the definition of achievable rates that $C_{\mathrm{s}}^{\mathrm{CM}} \leqslant \overline{C}_{\mathrm{s}}^{\mathrm{CM}}$. We prove in Section 4.5 that $C_{\mathrm{s}}^{\mathrm{CM}} = \overline{C}_{\mathrm{s}}^{\mathrm{CM}}$, but until then we focus on the weak secret-key capacity. Note that the definition of key-distillation strategies for the channel model is so broad that an exact characterization of the secret-key capacity for a general channel model seems out of reach. Nevertheless, it is possible to develop upper and lower bounds similar to those developed in Theorem 4.1 for the secret-key capacity.

Theorem 4.8 (Ahlswede and Csiszár)**.** *The secret-key capacity* $C_{\mathrm{s}}^{\mathrm{CM}}$ *of a channel model satisfies*

$$\max\left(\max_{p_{\mathsf{X}}}(\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z})), \max_{p_{\mathsf{X}}}(\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z}))\right) \leqslant C_{\mathrm{s}}^{\mathrm{CM}} \leqslant \max_{p_{\mathsf{X}}} \min\left(\mathbb{I}(\mathsf{X};\mathsf{Y}), \mathbb{I}(\mathsf{X};\mathsf{Y}|\mathsf{Z})\right).$$

Proof. We derive the lower bound by considering a specific key-distillation strategy, in which the the input X^n is chosen i.i.d. according to an arbitrary distribution p_{X}. For this choice of input, the channel model becomes a source model with DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{\mathsf{XYZ}})$

whose secret-key capacity is $C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{XYZ}})$. Therefore,

$$\begin{aligned} C_{\mathrm{s}}^{\mathsf{CM}} &\geqslant \max_{p_{\mathsf{X}}} C_{\mathrm{s}}^{\mathsf{SM}}(p_{\mathsf{XYZ}}) \\ &\geqslant \max\left(\max_{p_{\mathsf{X}}} (\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{X};\mathsf{Z})), \max_{p_{\mathsf{X}}} (\mathbb{I}(\mathsf{X};\mathsf{Y}) - \mathbb{I}(\mathsf{Y};\mathsf{Z}))\right), \end{aligned}$$

where the second inequality follows from Theorem 4.1.

We establish the upper bound with a converse argument similar to that used in the proof of Theorem 4.1; the proof is more technical because the inputs of the channel depend on previously exchanged messages and past inputs. Let R be an achievable secret-key rate and let $\epsilon > 0$. For n sufficiently large, there exists a $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ such that

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \leqslant \delta(\epsilon), \quad \frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \delta(\epsilon), \quad \text{and} \quad \frac{1}{n}\mathbf{U}(\mathcal{S}_n) \leqslant \delta(\epsilon). \tag{4.62}$$

In the following, we omit conditioning on $\mathcal{S}_n$ in order to simplify the notation. Fano's inequality combined with (4.62) ensures that

$$\frac{1}{n}\mathbb{H}\big(\mathsf{K}|\hat{\mathsf{K}}\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n\big) \leqslant \frac{1}{n}\mathbb{H}\big(\mathsf{K}|\hat{\mathsf{K}}\big) \leqslant \delta(\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n)) \leqslant \delta(\epsilon).$$

We also introduce the random variables Λ_j for $j \in [\![0, n]\!]$, which represent all public messages exchanged *after* Y_j and Z_j have been received but *before* Y_{j+1} and Z_{j+1} are received:

$$\begin{aligned} \Lambda_0 &\triangleq \big(\mathsf{A}_1, \ldots, \mathsf{A}_{i_1-1}, \mathsf{B}_1, \ldots, \mathsf{B}_{i_1-1}\big), \\ \Lambda_j &\triangleq \big(\mathsf{A}_{i_j+1}, \ldots, \mathsf{A}_{i_{j+1}-1}, \mathsf{B}_{i_j+1}, \ldots, \mathsf{B}_{i_{j+1}-1}\big) \quad \text{for } j \in [\![1, n]\!]. \end{aligned}$$

With this definition, note that $\mathsf{A}^r\mathsf{B}^r = \Lambda_0\Lambda^n$. Following the steps used to prove (4.20) for the source model, we can show that

$$\begin{aligned} R &\leqslant \frac{1}{n}\mathbb{I}(\mathsf{X}^n\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}|\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n) + \delta(\epsilon) \\ &= \frac{1}{n}\mathbb{I}(\mathsf{X}^n\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}|\Lambda_0\Lambda^n\mathsf{Z}^n) + \delta(\epsilon). \end{aligned} \tag{4.63}$$

However, in contrast to the case of the source model, $\mathsf{X}_j = h_j(\mathsf{B}^{i_j-1}, \mathsf{R}_{\mathsf{X}})$ is a function of B^{i_j-1} and R_{X}; therefore, the inequality simplifies to

$$R \leqslant \frac{1}{n}\mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}|\Lambda_0\Lambda^n\mathsf{Z}^n) + \delta(\epsilon).$$

Next, we expand $\mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}|\Lambda_0\Lambda^n\mathsf{Z}^n)$ as

$$\begin{aligned} &\mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}|\Lambda_0\Lambda^n\mathsf{Z}^n) \\ &\quad= \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{R}_{\mathsf{Y}}\mathsf{Z}^n\Lambda_0\Lambda^n) - \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Z}^n\Lambda_0\Lambda^n) \\ &\quad= \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{R}_{\mathsf{Y}}\Lambda_0) - \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \Lambda_0) + \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{Z}^n\Lambda^n|\mathsf{R}_{\mathsf{Y}}\Lambda_0) - \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Z}^n\Lambda^n|\Lambda_0), \\ &\quad= \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{R}_{\mathsf{Y}}|\Lambda_0) + \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Y}^n\mathsf{Z}^n\Lambda^n|\mathsf{R}_{\mathsf{Y}}\Lambda_0) - \mathbb{I}(\mathsf{R}_{\mathsf{X}}; \mathsf{Z}^n\Lambda^n|\Lambda_0) \end{aligned} \tag{4.64}$$

and study each term separately.

Since $\Lambda_0 = A^{i_1-1}B^{i_1-1}$, we can apply Lemma 4.2 to $\mathbb{I}(R_X; R_Y|\Lambda_0)$ with $S \triangleq R_X$, $T \triangleq R_Y$, $U \triangleq 0$, $r \triangleq i_1 - 1$, $V_i \triangleq A_i$, and $W_i \triangleq B_i$ to obtain

$$\mathbb{I}(R_X; R_Y|\Lambda_0) \leqslant \mathbb{I}(R_X; R_Y) = 0. \tag{4.65}$$

Next, notice that

$$\begin{aligned}\mathbb{I}(R_X; Y^n Z^n \Lambda^n|R_Y\Lambda_0) &= \sum_{j=1}^{n} \mathbb{I}\big(R_X; Z_j Y_j \Lambda_j|R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) \\ &= \sum_{j=1}^{n} \big(\mathbb{I}\big(R_X; Z_j Y_j|R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) \\ &\quad + \mathbb{I}\big(R_X; \Lambda_j|R_Y\Lambda_0 Z^j Y^j \Lambda^{j-1}\big)\big),\end{aligned} \tag{4.66}$$

and, similarly,

$$\mathbb{I}(R_X; Z^n\Lambda^n|\Lambda_0) = \sum_{j=1}^{n} \big(\mathbb{I}\big(R_X; Z_j|\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) + \mathbb{I}\big(R_X; \Lambda_j|\Lambda_0 Z^j\Lambda^{j-1}\big)\big). \tag{4.67}$$

On substituting (4.65), (4.66), and (4.67) into (4.64), we obtain

$$\begin{aligned}&\mathbb{I}(R_X; Y^n R_Y|\Lambda_0\Lambda^n Z^n) \\ &\quad\leqslant \sum_{j=1}^{n} \big(\mathbb{I}\big(R_X; Z_j Y_j|R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) - \mathbb{I}\big(R_X; Z_j|\Lambda_0 Z^{j-1}\Lambda^{j-1}\big)\big) \\ &\qquad + \sum_{j=1}^{n} \big(\mathbb{I}\big(R_X; \Lambda_j|R_Y\Lambda_0 Z^j Y^j\Lambda^{j-1}\big) - \mathbb{I}\big(R_X; \Lambda_j|\Lambda_0 Z^j\Lambda^{j-1}\big)\big).\end{aligned} \tag{4.68}$$

We proceed to bound the terms in each sum separately. First,

$$\begin{aligned}&\mathbb{I}\big(R_X; Z_j Y_j|R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) - \mathbb{I}\big(R_X; Z_j|\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) \\ &\quad= \mathbb{H}\big(Z_j Y_j|R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) - \mathbb{H}\big(Z_j Y_j|R_X R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) \\ &\qquad- \mathbb{H}\big(Z_j|\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) + \mathbb{H}\big(Z_j|R_X\Lambda_0 Z^{j-1}\Lambda^{j-1}\big).\end{aligned} \tag{4.69}$$

Since conditioning does not increase entropy, we have

$$\begin{aligned}&\mathbb{H}\big(Z_j Y_j|R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) - \mathbb{H}\big(Z_j|\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) \\ &\qquad\leqslant \mathbb{H}\big(Z_j Y_j|\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) - \mathbb{H}\big(Z_j|\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) \\ &\qquad= \mathbb{H}\big(Y_j|Z_j\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) \\ &\qquad\leqslant \mathbb{H}\big(Y_j|Z_j\big).\end{aligned} \tag{4.70}$$

In addition, since $R_X R_Y \Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1} \to X_j \to Y_j Z_j$ forms a Markov chain and $X_j = h_j(B^{i_j-1}, R_X)$, we have

$$\mathbb{H}\big(Z_j Y_j|R_X R_Y\Lambda_0 Z^{j-1}Y^{j-1}\Lambda^{j-1}\big) = \mathbb{H}\big(Z_j Y_j|X_j\big) \tag{4.71}$$

and

$$\mathbb{H}\big(Z_j|R_X\Lambda_0 Z^{j-1}\Lambda^{j-1}\big) = \mathbb{H}\big(Z_j|X_j\big). \tag{4.72}$$

On substituting (4.70), (4.71), and (4.72) into (4.69), we obtain

$$\begin{aligned}\mathbb{I}\big(\mathrm{R_X};\mathrm{Z}_j\mathrm{Y}_j|\mathrm{R_Y}\Lambda_0\mathrm{Z}^{j-1}\mathrm{Y}^{j-1}\Lambda^{j-1}\big)-\mathbb{I}\big(\mathrm{R_X};\mathrm{Z}_j|\Lambda_0\mathrm{Z}^{j-1}\Lambda^{j-1}\big)&\\ \leqslant \mathbb{H}\big(\mathrm{Y}_j|\mathrm{Z}_j\big)-\mathbb{H}\big(\mathrm{Z}_j\mathrm{Y}_j|\mathrm{X}_j\big)+\mathbb{H}\big(\mathrm{Z}_j|\mathrm{X}_j\big)&\\ =\mathbb{I}\big(\mathrm{X}_j;\mathrm{Y}_j|\mathrm{Z}_j\big).&\end{aligned} \tag{4.73}$$

We now turn our attention to the terms in the second sum of (4.68):

$$\begin{aligned}&\mathbb{I}\big(\mathrm{R_X};\Lambda_j|\mathrm{R_Y}\Lambda_0\mathrm{Z}^j\mathrm{Y}^j\Lambda^{j-1}\big)-\mathbb{I}\big(\mathrm{R_X};\Lambda_j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)\\ &\quad=\mathbb{I}\big(\mathrm{R_X};\Lambda_j\mathrm{R_Y}\mathrm{Y}^j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)-\mathbb{I}\big(\mathrm{R_X};\mathrm{R_Y}\mathrm{Y}^j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)-\mathbb{I}\big(\mathrm{R_X};\Lambda_j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)\\ &\quad=\mathbb{I}\big(\mathrm{R_X};\Lambda_j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)+\mathbb{I}\big(\mathrm{R_X};\mathrm{R_Y}\mathrm{Y}^j|\Lambda_0\mathrm{Z}^j\Lambda^j\big)-\mathbb{I}\big(\mathrm{R_X};\mathrm{R_Y}\mathrm{Y}^j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)\\ &\qquad-\mathbb{I}\big(\mathrm{R_X};\Lambda_j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)\\ &\quad=\mathbb{I}\big(\mathrm{R_X};\mathrm{R_Y}\mathrm{Y}^j|\Lambda_0\mathrm{Z}^j\Lambda^j\big)-\mathbb{I}\big(\mathrm{R_X};\mathrm{R_Y}\mathrm{Y}^j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big).\end{aligned}$$

By applying again Lemma 4.2 with $\mathrm{S}\triangleq\mathrm{R_X}$, $\mathrm{T}\triangleq\mathrm{R_Y}\mathrm{Y}^j$, $\mathrm{U}\triangleq\Lambda_0\mathrm{Z}^j$, $r\triangleq j$, $\mathrm{V}_i\triangleq\mathrm{A}_i$, and $W_i\triangleq\mathrm{B}_i$, we obtain

$$\mathbb{I}\big(\mathrm{R_X};\mathrm{R_Y}\mathrm{Y}_j|\Lambda_0\mathrm{Z}^j\Lambda^j\big)\leqslant\mathbb{I}\big(\mathrm{R_X};\mathrm{R_Y}\mathrm{Y}_j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)$$

and, therefore,

$$\mathbb{I}\big(\mathrm{R_X};\Lambda_j|\mathrm{R_Y}\Lambda_0\mathrm{Z}^j\mathrm{Y}^j\Lambda^{j-1}\big)-\mathbb{I}\big(\mathrm{R_X};\Lambda_j|\Lambda_0\mathrm{Z}^j\Lambda^{j-1}\big)\leqslant 0. \tag{4.74}$$

On substituting (4.73) and (4.74) into (4.68), we finally obtain

$$\mathbb{I}(\mathrm{R_X};\mathrm{Y}^n\mathrm{R_Y}|\Lambda^n\mathrm{Z}^n)\leqslant\sum_{j=1}^{n}\mathbb{I}\big(\mathrm{X}_j;\mathrm{Y}_j|\mathrm{Z}_j\big)\leqslant n\max_{p_\mathrm{X}}\mathbb{I}(\mathrm{X};\mathrm{Y}|\mathrm{Z}). \tag{4.75}$$

Hence,

$$R\leqslant\sum_{j=1}^{n}\mathbb{I}\big(\mathrm{X}_j;\mathrm{Y}_j|\mathrm{Z}_j\big)+\delta(\epsilon).$$

It remains to show that $R\leqslant\sum_{j=1}^{n}\mathbb{I}\big(\mathrm{X}_j;\mathrm{Y}_j\big)+\delta(\epsilon)$. Since

$$\frac{1}{n}\mathbb{H}\big(\mathrm{K}|\hat{\mathrm{K}}\mathrm{A}^r\mathrm{B}^r\big)\leqslant\frac{1}{n}\mathbb{H}\big(\mathrm{K}|\hat{\mathrm{K}}\big)\leqslant\delta(\epsilon),$$

it suffices to reiterate all the steps leading to (4.75) without the conditioning on Z^n to obtain the desired result. □

Even though Theorem 4.8 does not characterize the secret-key capacity exactly, it provides simple bounds that do not depend on auxiliary random variables.

Example 4.7. For some channel models, the secret-key capacity is even equal to the secrecy capacity. The simplest example of such a situation is a channel model in which the DMC $(\mathcal{X}, p_{\mathrm{YZ}|\mathrm{X}}, \mathcal{Y}, \mathcal{Z})$ is physically degraded. In this case $\mathrm{X}\to\mathrm{Y}\to\mathrm{Z}$ forms a Markov chain and

$$C_\mathrm{s}^{\mathrm{CM}}=\max_{p_\mathrm{X}}\mathbb{I}(\mathrm{X};\mathrm{Y}|\mathrm{Z})=C_\mathrm{s}.$$

In other words, using a wiretap code is an optimal key-distillation strategy for a channel model in which the eavesdropper's channel is physically degraded with respect to the main channel.

Remark 4.11. *The converse technique used in the proof of Theorem 4.8 can also be used to derive simple bounds for secure rates over WTCs; in fact, the secret-key capacity is always an upper bound for the secrecy capacity since a wiretap code is a special case of key-distillation strategy for a channel model. As an application, we revisit the WTC with confidential rate-limited feedback, for which we computed achievable rates in Section 3.6.2. Following the steps used to establish* (4.20)*, we can show that the secret-key capacity with public discussion of the WTC with confidential rate-limited feedback satisfies*

$$\frac{1}{n}\mathbb{H}(\mathsf{K}) \leqslant \frac{1}{n}\mathbb{I}(\mathsf{R}_X\mathsf{F}^n; \mathsf{R}_Y\mathsf{Y}^n|\mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r) + \delta(\epsilon).$$

The only difference from (4.63) *is the presence of the term* F^n *representing the messages sent over the confidential feedback channel. Nevertheless, we have*

$$\begin{aligned}\frac{1}{n}\mathbb{I}(\mathsf{R}_X\mathsf{F}^n; \mathsf{R}_Y\mathsf{Y}^n|\mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r) &= \frac{1}{n}\mathbb{I}(\mathsf{F}^n; \mathsf{R}_Y\mathsf{Y}^n|\mathsf{A}^r\mathsf{B}^r\mathsf{Z}^n) + \frac{1}{n}\mathbb{I}(\mathsf{R}_X; \mathsf{R}_Y\mathsf{Y}^n|\mathsf{A}^r\mathsf{B}^r\mathsf{F}^n\mathsf{Z}^n)\\ &\leqslant \frac{1}{n}\mathbb{H}(\mathsf{F}^n) + \frac{1}{n}\mathbb{I}(\mathsf{R}_X; \mathsf{R}_Y\mathsf{Y}^n|\mathsf{A}^r\mathsf{B}^r\mathsf{F}^n\mathsf{Z}^n).\end{aligned}$$

The second term on the right-hand side is similar to (4.63)*, since the messages sent over the secure feedback channel appear along* A^r *and* B^r *and can now be interpreted as public messages. Following the same steps and using the fact that* $(1/n)\mathbb{H}(\mathsf{F}^n) \leqslant R_{\mathrm{f}}$*, we obtain the upper bound*

$$\frac{1}{n}\mathbb{H}(\mathsf{K}) \leqslant R_{\mathrm{f}} + \sum_{j=1}^{n}\mathbb{I}\big(\mathsf{X}_j; \mathsf{Y}_j|\mathsf{Z}_j\big).$$

This outer bound coincides with the achievable rate obtained in Proposition 3.9 if the channel is physically degraded.

4.5 Strong secrecy from weak secrecy

In this section, we develop a generic mathematical procedure by which to construct a scheme (wiretap code or key-distillation strategy) that guarantees a strong secrecy condition from a scheme that guarantees a weak secrecy condition. We show that this procedure entails no rate loss, which allows us to prove that $C_{\mathrm{s}}^{\mathsf{CM}} = \overline{C}_{\mathrm{s}}^{\mathsf{CM}}$ for a channel model and $C_{\mathrm{s}} = \overline{C}_{\mathrm{s}}$ for a wiretap channel and justifies a posteriori the use of a weak secrecy condition in previous chapters.

Proposition 4.9. *The strong secret-key capacity* $\overline{C}_{\mathrm{s}}^{\mathsf{CM}}$ *of a channel model with DMC* $(\mathcal{X}, p_{\mathsf{YZ}|\mathsf{X}}, \mathcal{Y}, \mathcal{Z})$ *is equal to its weak secret-key capacity* $C_{\mathrm{s}}^{\mathsf{CM}}$*.*

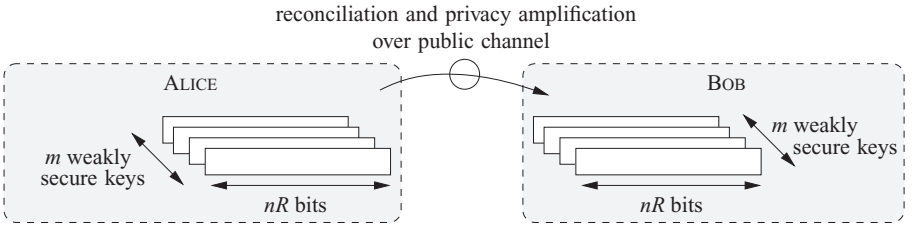

Figure 4.8 From weak to strong secrecy.

Proof. Consider a channel model with DMC $(\mathcal{X}, p_{\mathsf{YZ}|\mathsf{X}}, \mathcal{Y}, \mathcal{Z})$ and a $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ that achieves weak secret-key rates. As illustrated in Figure 4.8, we construct a new strategy by

- using the key-distillation strategy $\mathcal{S}_n$ m times to generate m weakly secure keys;
- treating the weakly secure keys as the realizations of a DMS and distilling strong secret keys by means of information reconciliation and privacy amplification with extractors.

Note that the post-processing of the weakly secure keys is possible because keys are not meant to carry any information by themselves and do not need to be known ahead of time.

Formally, let $\epsilon > 0$. By definition, there exists a $(2^{nR}, n)$ key-distillation strategy $\mathcal{S}_n$ with rate $R \geqslant C_{\mathrm{s}}^{\mathsf{CM}} - \epsilon$ such that

$$\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n) \leqslant \delta(\epsilon) \quad \text{and} \quad \frac{1}{n}\mathbf{L}(\mathcal{S}_n) \leqslant \delta(\epsilon).$$

Alice runs $\mathcal{S}_n$ m times to generate m independent keys. In each run $i \in [\![1, m]\!]$, Alice obtains a key K_i, Bob obtains a key $\hat{\mathsf{K}}_i$, and Eve obtains the observations Z_i^n together with public messages A_i^r and B_i^r. Effectively, the situation is as if Alice, Bob, and Eve observed m realizations of a DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$ with

$$\mathsf{X}' \triangleq \mathsf{K}, \quad \mathsf{Y}' \triangleq \hat{\mathsf{K}}, \quad \text{and} \quad \mathsf{Z}' \triangleq \mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r.$$

According to Theorem 4.7, Alice and Bob can distill a strong secret key $\overline{\mathsf{K}}$ of length

$$k \triangleq m(\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}') - \delta(\epsilon))$$

and such that

$$\mathbb{H}(\overline{\mathsf{K}}|\mathsf{Z}'^m\mathcal{S}_n) \geqslant k - \delta_\epsilon(m)$$

by means of a one-way direct reconciliation protocol and privacy amplification with extractors. Note that

$$\begin{aligned}
\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}') &= \mathbb{I}(\mathsf{K};\hat{\mathsf{K}}|\mathcal{S}_n) - \mathbb{I}(\mathsf{K};\mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r|\mathcal{S}_n) \\
&= \mathbb{H}(\mathsf{K}) - \mathbb{H}(\mathsf{K}|\hat{\mathsf{K}}\mathcal{S}_n) - \mathbb{I}(\mathsf{K};\mathsf{Z}^n\mathsf{A}^r\mathsf{B}^r|\mathcal{S}_n) \\
&\geqslant nR - n\delta(\mathbf{P}_{\mathrm{e}}(\mathcal{S}_n)) - n\mathbf{L}(\mathcal{S}_n) \\
&\geqslant nC_{\mathrm{s}}^{\mathsf{CM}} - n\delta(\epsilon).
\end{aligned}$$

Therefore, the rate (in bits per use of the channel $(\mathcal{X}, p_{\mathsf{YZ}|\mathsf{X}}, \mathcal{Y}, \mathcal{Z})$) at which the strong secret-key $\overline{\mathsf{K}}$ is generated is

$$\frac{k}{mn} \geqslant C_{\mathrm{s}}^{\mathsf{CM}} - \delta(\epsilon).$$

Since ϵ can be chosen arbitrarily small, we conclude that $\overline{C}_{\mathrm{s}}^{\mathsf{CM}} \geqslant C_{\mathrm{s}}^{\mathsf{CM}}$ and, therefore, $\overline{C}_{\mathrm{s}}^{\mathsf{CM}} = C_{\mathrm{s}}^{\mathsf{CM}}$. □

Proposition 4.10. *The strong secrecy capacity $\overline{C}_{\mathrm{s}}^{\mathsf{WT}}$ of a WTC $(\mathcal{X}, p_{\mathsf{YZ}|\mathsf{X}}, \mathcal{Y}, \mathcal{Z})$ is equal to its weak secrecy capacity $C_{\mathrm{s}}^{\mathsf{WT}}$.*

Proof. The proof is similar to that of Proposition 4.9 but the objective is now to transmit messages (instead of generating keys) without relying on a public channel. Let $\epsilon > 0$. By definition, there exists a $(2^{nR}, n)$ code $\mathcal{C}_n$ with rate $R \geqslant C_{\mathrm{s}}^{\mathsf{WT}} - \epsilon$ such that

$$\mathbf{P}_{\mathrm{e}}(\mathcal{C}_n) \leqslant \delta(\epsilon) \quad \text{and} \quad \frac{1}{n}\mathbf{L}(\mathcal{C}_n) \leqslant \delta(\epsilon).$$

Alice uses the code $\mathcal{C}_n$ m times to transmit m independent messages. In each run $i \in [\![1, m]\!]$, Alice transmits a message M_i, Bob obtains a message $\hat{\mathsf{M}}_i$, and Eve obtains the observations Z_i^n. Again, the situation is as if Alice, Bob, and Eve observed m realizations of a DMS $(\mathcal{X}'\mathcal{Y}'\mathcal{Z}', p_{\mathsf{X}'\mathsf{Y}'\mathsf{Z}'})$ with

$$\mathsf{X}' \triangleq \mathsf{M}, \quad \mathsf{Y}' \triangleq \hat{\mathsf{M}}, \quad \text{and} \quad \mathsf{Z}' \triangleq \mathsf{Z}^n.$$

According to Theorem 4.7, Alice and Bob can distill a strong secret key $\overline{\mathsf{K}}$ of length

$$k \triangleq m(\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}') - \delta(\epsilon))$$

and such that

$$\mathbb{H}(\overline{\mathsf{K}}|\mathsf{Z}'^m\mathcal{C}_n) \geqslant k - \delta_\epsilon(m)$$

by means of a one-way direct reconciliation protocol and privacy amplification with extractors. The reader can check that

$$\mathbb{I}(\mathsf{X}';\mathsf{Y}') - \mathbb{I}(\mathsf{X}';\mathsf{Z}') \geqslant nC_{\mathrm{s}}^{\mathsf{WT}} - n\delta(\epsilon).$$

However, because there is no public channel, the messages required in order to perform reconciliation and privacy amplification must be transmitted over the channel; we must also account for these additional channel uses in the calculation of the final key rate.

By Proposition 4.5, there exists a one-way reconciliation protocol that exchanges $m(\mathbb{H}(\mathsf{X}'|\mathsf{Y}') + \delta(\epsilon))$ bits of public messages and that guarantees $\mathbb{P}[\mathsf{X}' \neq \hat{\mathsf{X}}'] \leqslant \delta_\epsilon(m)$. By Theorem 4.6, there exists an extractor that requires the transmission of $m\delta(m)$ bits of uniform randomness in order to distill the key $\overline{\mathsf{K}}$. Alice can transmit these bits to Bob over the main channel $(\mathcal{X}, p_{\mathsf{Y}|\mathsf{X}}, \mathcal{Y})$ with an error-correcting code of length m. Let C_{m} denote the capacity of the channel $(\mathcal{X}, p_{\mathsf{Y}|\mathsf{X}}, \mathcal{Y})$. From Shannon's channel coding theorem, we know that there exists a code of rate $C_{\mathrm{m}} - \delta(\epsilon)$ that guarantees an average probability of error less than $\delta_\epsilon(m)$. Therefore, the transmission of these additional bits

requires

$$m' \triangleq \left\lceil \frac{m(\mathbb{H}(X'|Y') + \delta(\epsilon)) + m\delta(m)}{C_{\mathrm{m}} - \delta(\epsilon)} \right\rceil$$

channel uses. By virtue of Fano's inequality,

$$\mathbb{H}\big(\mathsf{X}'|\mathsf{Y}'\big) = \mathbb{H}\big(\mathsf{M}|\hat{\mathsf{M}}\big) \leqslant \mathbb{H}_{\mathrm{b}}(\mathbf{P}_{\mathrm{e}}(\mathcal{C}_n)) + \mathbf{P}_{\mathrm{e}}(\mathcal{C}_n) n C_{\mathrm{s}}^{\mathsf{WT}} = n\delta_\epsilon(n).$$

All in all, the strong secret key $\overline{\mathsf{K}}$ can be generated at a rate

$$\begin{aligned}\frac{k}{mn + m'} &\geqslant \frac{m(nC_{\mathrm{s}}^{\mathsf{WT}} - n\delta(\epsilon))}{mn + \dfrac{m(n\delta_\epsilon(n) + \delta(\epsilon)) + m\delta(m)}{C_{\mathrm{m}} - \delta(\epsilon)} + 1} \\ &= C_{\mathrm{s}}^{\mathsf{WT}} - \delta(\epsilon) - \delta_\epsilon(n, m).\end{aligned}$$

In other words, for n and m large enough, the transmission of reconciliation and privacy-amplification messages over the channel incurs a *negligible* rate penalty.

To conclude, it remains to show that the key $\overline{\mathsf{K}}$ can be interpreted as a message so that the key rate is a message rate. Notice that all communications over the channel are one-way; therefore, in principle, Alice could choose the final key $\overline{\mathsf{K}}$ ahead of time, "invert" the privacy-amplification process, and artificially split the transmission over mn channel uses. Hence, the final strong secret-key $\overline{\mathsf{K}}$ can be treated as a message M that satisfies a strong secrecy condition. Since ϵ can be arbitrarily small, we obtain $\overline{C}_{\mathrm{s}}^{\mathsf{WT}} \geqslant C_{\mathrm{s}}^{\mathsf{WT}}$ and thus $\overline{C}_{\mathrm{s}}^{\mathsf{WT}} = C_{\mathrm{s}}^{\mathsf{WT}}$. □

Remark 4.12. *Privacy amplification with extractors is critical to obtaining strong secrecy with a negligible rate penalty. It is possible to show that the minimum size of a universal family of hash functions $\mathcal{G} : \{0,1\}^{mn} \to \{0,1\}^k$ is 2^{mn-k}; therefore, the minimum number of bits to describe a randomly chosen hash function in the family is $mn - k$ and the number of channel uses required to transmit this choice would be*

$$\frac{mn - k}{C_{\mathrm{m}} - \delta(\epsilon)},$$

which incurs a non-negligible rate penalty.

Remark 4.13. *The mathematical procedure used to convert weakly secure codes in strongly secure codes is not really practical. Although the "inversion" of the process is conceptually feasible, it becomes quickly intractable as n and m become large. Hence, this procedure does not replace the construction of wiretap codes, which we discuss in Chapter 6.*

4.6 Conclusions and lessons learned

The results obtained in this chapter allow us to draw several crucial conclusions. First and foremost, the analysis of the secret-key capacity for source and channel models shows that *feedback improves secrecy*. We already knew from Section 3.6.2 that secure

feedback is beneficial for secrecy, but this statement remains true even if the feedback is known to the eavesdropper. This suggests that the need for an advantage over the eavesdropper at the physical layer highlighted in Chapter 3 was largely the consequence of the restrictions imposed on the coding schemes.

Second, secret-key agreement seems much more practical than wiretap coding at this point. The sequential key-distillation strategies based on advantage distillation, information reconciliation, and privacy amplification described in Section 4.3 handle the reliability and secrecy requirements *independently*, which leads to effective ways of distilling secret keys. Nevertheless, note that the fundamental limits of secret-key agreement from source models or channel models are not as well understood as those of secure communication over wiretap channels. This state of affairs is partially explained by the fact that two-way communications seem to be an essential ingredient of key-distillation strategies, and they are significantly harder to analyze than one-way communications. One should also note the tight connection between key distillation and source coding with side information.

Finally, the study of secret-key agreement allows us to develop a generic mathematical procedure by which to strengthen secrecy results, which justifies a posteriori the relevance of the fundamental limits derived with a weak secrecy criterion in Chapter 3 and the first sections of this chapter. In some sense, strong secrecy comes "for free" but the reader should keep in mind that, although the fundamental limits of secure communication and secret-key generation remain the same, the coding schemes achieving strong secrecy might be quite different from those achieving weak secrecy.

4.7 Appendix

Proof of Lemma 4.1

Proof. The result will follow from Chebyshev's inequality, provided that we first establish an upper bound for $\mathrm{Var}\left(p_{\mathsf{K}_{\mathsf{S}_n}}(1)\right) = \mathbb{E}_{\mathsf{S}_n}\left[(p_{\mathsf{K}_{\mathsf{S}_n}}(1))^2\right] - \mathbb{E}_{\mathsf{S}_n}\left[(p_{\mathsf{K}_{\mathsf{S}_n}}(1))\right]^2$. By virtue of the definition of $\mathsf{K}_{\mathcal{S}_n}$, we can write $p_{\mathsf{K}_{\mathcal{S}_n}}(1)$ as

$$p_{\mathsf{K}_{\mathcal{S}_n}}(1) = \sum_{x^n \in \mathcal{T}_\epsilon^n(\mathsf{X})} \frac{p_{\mathsf{X}^n}(x^n)}{\mathbb{P}[\Xi = 1]} \mathbb{1}(\kappa_{\mathrm{a}}(x^n) = 1),$$

where κ_{a} is the key-distillation function used by Alice in the strategy $\mathcal{S}_n$. Hence,

$$\begin{aligned}
\mathbb{E}_{\mathsf{S}_n}\left[p_{\mathsf{K}_{\mathsf{S}_n}}(1)\right] &= \sum_{x^n \in \mathcal{T}_\epsilon^n(\mathsf{X})} \frac{p_{\mathsf{X}^n}(x^n)}{\mathbb{P}[\Xi = 1]} \mathbb{E}_{\mathsf{S}_n}[\mathbb{1}(\kappa_{\mathrm{a}}(x^n) = 1)] \\
&= \sum_{x^n \in \mathcal{T}_\epsilon^n(\mathsf{X})} \frac{p_{\mathsf{X}^n}(x^n)}{\mathbb{P}[\Xi = 1]} \frac{1}{\lceil 2^{nR} \rceil} \\
&= \frac{1}{\lceil 2^{nR} \rceil}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}_{S_n}\left[(p_{K|_{S_n}}(1))^2\right] \\
&\quad = \mathbb{E}_{S_n}\left[\left(\sum_{x^n\in\mathcal{T}_\epsilon^n(X)} \frac{p_{X^n}(x^n)}{\mathbb{P}[\Xi=1]}\mathbb{1}(\kappa_a(x^n)=1)\right)^2\right] \\
&\quad \leqslant \mathbb{E}_{S_n}\left[\sum_{x^n\in\mathcal{T}_\epsilon^n(X)}\left(\frac{p_{X^n}(x^n)}{\mathbb{P}[\Xi=1]}\right)^2 \mathbb{1}(\kappa_a(x^n)=1)\right. \\
&\qquad\qquad \left. + \sum_{x^n\in\mathcal{T}_\epsilon^n(X)}\sum_{x'^n\in\mathcal{T}_\epsilon^n(X)} \frac{p_{X^n}(x^n)\,p_{X^n}(x'^n)}{\mathbb{P}[\Xi=1]^2}\mathbb{1}(\kappa_a(x^n)=1)\,\mathbb{1}\left(\kappa_a(x'^n)=1\right)\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{E}_{S_n}\left[(p_{K_{S_n}}(1))^2\right] \\
&\quad \leqslant \sum_{x^n\in\mathcal{T}_\epsilon^n(X)}\left(\frac{p_{X^n}(x^n)}{\mathbb{P}[\Xi=1]}\right)^2 \mathbb{E}_{S_n}[\mathbb{1}(\kappa_a(x^n)=1)] \\
&\qquad + \sum_{x^n\in\mathcal{T}_\epsilon^n(X)}\sum_{x'^n\in\mathcal{T}_\epsilon^n(X)} \frac{p_{X^n}(x^n)p_{X^n}(x'^n)}{\mathbb{P}[\Xi=1]^2}\mathbb{E}_{S_n}\left[\mathbb{1}(\kappa_a(x^n)=1)\,\mathbb{1}\left(\kappa_a(x'^n)=1\right)\right].
\end{aligned}$$

Using the AEP, we bound the first term on the right-hand side as

$$\sum_{x^n\in\mathcal{T}_\epsilon^n(X)}\left(\frac{p_{X^n}(x^n)}{\mathbb{P}[\Xi=1]}\right)^2 \mathbb{E}_{S_n}[\mathbb{1}(\kappa_a(x^n)=1)] \leqslant \frac{2^{-n(\mathbb{H}(X)-\delta(\epsilon))}}{\lceil 2^{nR}\rceil(1-\delta_\epsilon(n))}.$$

Similarly, we bound the second term on the right-hand side as

$$\sum_{x^n\in\mathcal{T}_\epsilon^n(X)}\sum_{x'^n\in\mathcal{T}_\epsilon^n(X)} \frac{p_{X^n}(x^n)p_{X^n}(x'^n)}{\mathbb{P}[\Xi=1]^2}\mathbb{E}_{S_n}\left[\mathbb{1}(\kappa_a(x^n)=1)\,\mathbb{1}\left(\kappa_a(x'^n)=1\right)\right] = \left(\frac{1}{\lceil 2^{nR}\rceil}\right)^2.$$

Therefore, we obtain the following bound for $\mathrm{Var}\left(p_{K_{S_n}}(1)\right)$:

$$\mathrm{Var}\left(p_{K_{S_n}}(1)\right) = \mathbb{E}_{S_n}\left[(p_{K_{S_n}}(1))^2\right] - \mathbb{E}_{S_n}\left[(p_{K_{S_n}}(1))\right]^2 \leqslant \frac{1}{2^{nR}}\frac{2^{-n(\mathbb{H}(X)-\delta(\epsilon))}}{1-\delta_\epsilon(n)}.$$

By virtue of Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}_{S_n}\left[\left|p_{K_{S_n}}(1) - 2^{-nR}\right| > \epsilon 2^{-nR}\right] &\leqslant \frac{\mathrm{Var}\left(p_{K_{S_n}}(1)\right)}{\epsilon^2 2^{-2nR}} \\
&\leqslant \frac{1}{\epsilon^2}\frac{2^{-n(\mathbb{H}(X)-R-\delta(\epsilon))}}{1-\delta_\epsilon(n)}.
\end{aligned}$$

If $R < \mathbb{H}(X) - \delta(\epsilon)$ then $\mathbb{P}_{S_n}\left[\left|p_{K_{S_n}}(1) - 2^{-nR}\right| > \epsilon 2^{-nR}\right] \leqslant \delta_\epsilon(n)$. □

Proof of Lemma 4.3

Proof. Let $t \triangleq r + 1$. We show that $\mathbb{P}_{\mathsf{U}}[\chi(\mathsf{U}) = 0] \leqslant 2^{-r}$ by developing an upper bound as follows:

$$\begin{aligned}\mathbb{P}_{\mathsf{U}}[\chi(\mathsf{U}) = 0] &= \mathbb{P}_{\mathsf{U}}[\mathbb{H}_{\mathrm{c}}(\mathsf{S}) - \log|\mathcal{U}| - 2t - \mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{U}) > 0] \\ &= \mathbb{P}_{\mathsf{U}}[\mathbb{H}_{\mathrm{c}}(\mathsf{S}) + \log p_{\mathsf{U}}(u) - \mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{U}) - t - \log p_{\mathsf{U}}(u) - \log|\mathcal{U}|] \\ &\quad - t > 0 \\ &\leqslant \mathbb{P}_{\mathsf{U}}[\mathbb{H}_{\mathrm{c}}(\mathsf{S}) + \log p_{\mathsf{U}}(u) - \mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{U}) - t \geqslant 0] \\ &\quad + \mathbb{P}_{\mathsf{U}}[-\log p_{\mathsf{U}}(u) - \log|\mathcal{U}| - t \geqslant 0].\end{aligned}$$

We introduce the random variables

$$\mathsf{X}_{\mathsf{U}} \triangleq 2^{\log p_{\mathsf{U}}(\mathsf{U}) - \mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{U}) + \mathbb{H}_{\mathrm{c}}(\mathsf{S})} \quad \text{and} \quad \mathsf{Y}_{\mathsf{U}} \triangleq -\log|\mathcal{U}| - \log p_{\mathsf{U}}(\mathsf{U})$$

so that we obtain

$$\mathbb{P}_{\mathsf{U}}[\chi(\mathsf{U}) = 0] \leqslant \mathbb{P}\left[\mathsf{X}_{\mathsf{U}} \geqslant 2^{t}\right] + \mathbb{P}[\mathsf{Y}_{\mathsf{U}} \geqslant t], \tag{4.76}$$

and we upper bound each term on the right-hand side of (4.76) separately.

First, note that

$$\mathbb{P}[\mathsf{Y}_{\mathsf{U}} \geqslant t] = \mathbb{P}\left[p_{\mathsf{U}}(\mathsf{U}) < \frac{2^{-t}}{|\mathcal{U}|}\right] = \sum_{u \in \mathcal{U}: p_{\mathsf{U}}(u) < 2^{-t}/|\mathcal{U}|} p_{\mathsf{U}}(u) \leqslant 2^{-t}. \tag{4.77}$$

Next, we develop an upper bound $\mathbb{P}\left[\mathsf{X}_{\mathsf{U}} \geqslant 2^{t}\right]$. Since X_{U} is positive, we establish it by upper bounding $\mathbb{E}_{\mathsf{U}}[\mathsf{X}_{\mathsf{U}}]$ and using Markov's inequality. We start by writing the collision entropy $\mathbb{H}_{\mathrm{c}}(\mathsf{SU})$ as

$$\begin{aligned}\mathbb{H}_{\mathrm{c}}(\mathsf{SU}) &= -\log\left(\sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{U}} (p_{\mathsf{SU}}(s,u))^2\right) \\ &= -\log\left(\sum_{u \in \mathcal{U}} (p_{\mathsf{U}}(u))^2 \sum_{s \in \mathcal{S}} \left(p_{\mathsf{S}|\mathsf{U}}(s|u)\right)^2\right).\end{aligned}$$

Note that

$$\sum_{s \in \mathcal{S}} \left(p_{\mathsf{S}|\mathsf{U}}(s|u)\right)^2 = 2^{-\mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{U}=u)} \quad \text{and} \quad p_{\mathsf{U}}(u) = 2^{\log p_{\mathsf{U}}(u)};$$

therefore,

$$\mathbb{H}_{\mathrm{c}}(\mathsf{SU}) = -\log\left(\sum_{u \in \mathcal{U}} p_{\mathsf{U}}(u) 2^{\log p_{\mathsf{U}}(u) - \mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{U}=u)}\right),$$

which we can rewrite as

$$2^{-\mathbb{H}_{\mathrm{c}}(\mathsf{SU})} = \mathbb{E}_{\mathsf{U}}\left[2^{\log p_{\mathsf{U}}(\mathsf{U}) - \mathbb{H}_{\mathrm{c}}(\mathsf{S}|\mathsf{U})}\right].$$

Hence,

$$\mathbb{E}_{\mathsf{U}}[\mathsf{X}_{\mathsf{U}}] = 2^{-\mathbb{H}_{\mathrm{c}}(\mathsf{SU})+\mathbb{H}_{\mathrm{c}}(\mathsf{S})}.$$

The reader can easily check that $\mathbb{H}_{\mathrm{c}}(\mathsf{S}) \leqslant \mathbb{H}_{\mathrm{c}}(\mathsf{SU})$ and, therefore, $\mathbb{E}_{\mathsf{U}}[\mathsf{X}_{\mathsf{U}}] \leqslant 1$. Hence, by virtue of Markov's inequality,

$$\mathbb{P}_{\mathsf{U}}\left[\mathsf{X}_{\mathsf{U}} \geqslant 2^{t}\right] \leqslant \frac{\mathbb{E}_{\mathsf{U}}[\mathsf{U}]}{2^{t}} \leqslant 2^{-t}. \tag{4.78}$$

On substituting (4.77) and (4.78) into (4.76), we finally obtain

$$\mathbb{P}_{\mathsf{U}}[\chi(\mathsf{U}) = 0] \leqslant 2^{-t} + 2^{-t} = 2^{-(t-1)} = 2^{-r}. \qquad \square$$

Proof of Lemma 4.4

Proof. Using a strengthened version of Theorem 2.1 and Corollary 2.1 [6] we obtain

$$\mathbb{P}\left[\mathsf{Z}^n \in \mathcal{T}_{\epsilon}^{n}(\mathsf{Z})\right] \geqslant 1 - \frac{2^{-\sqrt{n}}}{2} \quad \text{and} \quad \mathbb{P}\left[(\mathsf{X}^n, \mathsf{Z}^n) \in \mathcal{T}_{2\epsilon}^{n}(\mathsf{XZ})\right] \geqslant 1 - \frac{2^{-\sqrt{n}}}{2}$$

for n sufficiently large. Hence, by the union bound, $\mathbb{P}[\Theta = 1] \geqslant 1 - 2^{-\sqrt{n}}$. In addition, for $z^n \in \mathcal{T}_{\epsilon}^{n}(\mathsf{Z})$, Theorem 2.2 guarantees that $\mathbb{P}_{\mathsf{X}^n|\mathsf{Z}^n}\left[\mathsf{X}^n \in \mathcal{T}_{2\epsilon}^{n}(\mathsf{XZ}|z^n)|\mathsf{Z}^n = z^n\right] \geqslant 1 - 2^{-\sqrt{n}}$.

To obtain the remaining part of the lemma, it suffices to establish a lower bound for the min-entropy $\mathbb{H}_{\infty}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1)$ with $z^n \in \mathcal{T}_{\epsilon}^{n}(\mathsf{Z})$ because Proposition 4.7 already proves that

$$\mathbb{H}_{\mathrm{c}}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1) \geqslant \mathbb{H}_{\infty}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1).$$

By definition,

$$\mathbb{H}_{\infty}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1) = -\log \max_{x^n} p_{\mathsf{X}^n|\mathsf{Z}^n\Theta}(x^n|z^n, 1).$$

By Bayes' rule, we obtain for all $x^n \in \mathcal{X}^n$

$$\begin{aligned} p_{\mathsf{X}^n|\mathsf{Z}^n\Theta}(x^n|z^n, 1) &= \frac{\mathbb{P}[\Theta = 1|\mathsf{X}^n = x^n, \mathsf{Z}^n = z^n] p_{\mathsf{X}^n|\mathsf{Z}^n}(x^n|z^n)}{\mathbb{P}[\Theta = 1|\mathsf{Z}^n = z^n]} \\ &= \frac{\mathbb{P}[\Theta = 1|\mathsf{X}^n = x^n, \mathsf{Z}^n = z^n] p_{\mathsf{X}^n|\mathsf{Z}^n}(x^n|z^n)}{\mathbb{P}\left[\mathsf{X}^n \in \mathcal{T}_{2\epsilon}^{n}(\mathsf{XZ}|z^n)|\mathsf{Z}^n = z^n\right]} \\ &\leqslant \frac{2^{-n(\mathbb{H}(\mathsf{X}|\mathsf{Z})-\delta(\epsilon))}}{1-\delta_{\epsilon}(n)}, \end{aligned}$$

where the last inequality follows from $p_{\mathsf{X}^n|\mathsf{Z}^n}(x^n|z^n) \leqslant 2^{-n(\mathbb{H}(\mathsf{X}|\mathsf{Z})-\delta(\epsilon))}$ by Theorem 2.1 if $\mathbb{P}[\Theta = 1|\mathsf{X}^n = x^n, \mathsf{Z}^n = z^n] > 0$. Hence,

$$\begin{aligned} \mathbb{H}_{\infty}(\mathsf{X}^n|\mathsf{Z}^n = z^n, \Theta = 1) &\geqslant n(\mathbb{H}(\mathsf{X}|\mathsf{Z}) - \delta(\epsilon)) + \log(1 - \delta_{\epsilon}(n)) \\ &= n(\mathbb{H}(\mathsf{X}|\mathsf{Z}) - \delta(\epsilon)) - \delta_{\epsilon}(n). \end{aligned}$$

$\square$

Proof of Lemma 4.5

Proof. We show that $\mathbb{P}_{\mathsf{U}}[\chi(\mathsf{U}) = 0] \leqslant 2^{-r}$. Note that $\mathbb{H}_\infty(\mathsf{S}|\mathsf{U} = u)$ satisfies

$$\begin{aligned}\mathbb{H}_\infty(\mathsf{S}|\mathsf{U} = u) &= -\log \max_s p_{\mathsf{S}|\mathsf{U}}(s|u) \\ &= -\log \max_s p_{\mathsf{S}\mathsf{U}}(s,u) + \log p_{\mathsf{U}}(u) \\ &\geqslant -\log \max_s p_{\mathsf{S}}(s) + \log p_{\mathsf{U}}(u) \\ &= \mathbb{H}_\infty(\mathsf{S}) + \log p_{\mathsf{U}}(u),\end{aligned}$$

where the inequality follows because $\forall (s,u) \in \mathcal{S} \times \mathcal{U}\ p_{\mathsf{S}\mathsf{U}}(s,u) \leqslant p_{\mathsf{S}}(s)$. Hence,

$$\begin{aligned}\mathbb{P}_{\mathsf{U}}[\chi(\mathsf{U}) = 0] &= \mathbb{P}_{\mathsf{U}}[\mathbb{H}_\infty(\mathsf{S}) > \mathbb{H}_\infty(\mathsf{S}|\mathsf{U} = u) + r + \log|\mathcal{U}|] \\ &\leqslant \mathbb{P}[\log p_{\mathsf{U}}(\mathsf{U}) < -r - \log|\mathcal{U}|] \\ &= \mathbb{P}\left[p_{\mathsf{U}}(\mathsf{U}) < \frac{2^{-r}}{|\mathcal{U}|}\right] \\ &= \sum_{u \in \mathcal{U}: p_{\mathsf{U}}(u) < 2^{-r}/|\mathcal{U}|} p_{\mathsf{U}}(u) \\ &\leqslant 2^{-r}.\end{aligned}$$

□

4.8 Bibliographical notes

Since the design of sequential key-distillation strategies was largely motivated by applications to quantum key distribution, several survey papers and textbooks discuss secret-key agreement in this context. A high-level presentation of quantum cryptography can be found in the textbook of Nielsen and Chuang [39]. The textbook of Van Assche [38] provides a comprehensive discussion of secret-key distillation and its applications to continuous-variable quantum key distribution. The survey paper of Gisin, Ribordy, Tittel, and Zbinden [40] also discusses secret-key agreement for quantum key distribution and focuses on practical system implementation. A more recent survey of secret-key agreement by Maurer, Renner, and Wolf is also available [41].

The fundamental limits of secret-key agreement by public discussion from common randomness were first investigated by Maurer [42] and Ahlswede and Csiszár [43]. The bounds for secret-key capacity presented in this chapter are the simplest known bounds for a general source or channel model; nevertheless, alternative bounds exist, such as those derived by Gohari and Anantharam [44, 45]. The alternative upper bounds for the secret-key capacity presented in this chapter are those obtained by Maurer and Wolf [46] and Renner and Wolf [47].

When restrictions are placed on public communication, it is sometimes possible to characterize the secret-key capacity exactly. For instance, Ahlswede and Csiszár characterized the forward secret-key capacity of source and channel models, defined as the secret-key capacity when public communication is limited to a single transmission

from Alice to Bob [43]. The secret-key capacity of a source model with rate-limited and one-way communication was characterized by Csiszár and Narayan [48] as the consequence of more general results on secret-key agreement with a helper; these results were used subsequently by Khisti, Diggavi, and Wornell [49] as well as Prabhakaran and Ramchandran [50] to analyze secret-key agreement without a public channel. A closed-form expression for the secret-key capacity of a Gaussian source model with rate-limited public communication was established by Watanabe and Oohama [51].

The concept of advantage distillation and the "satellite" source model were introduced by Maurer [42] to show that key-distillation strategies with two-way communication are, in general, more powerful than key-distillation strategies with one-way communication. Maurer's repetition protocol suffices to illustrate the principle of advantage distillation but it achieves a low advantage-distillation rate. Alternative protocols with higher advantage-distillation rates have been studied, such as the "bit-pair iteration" protocol of Gander and Maurer [52]. Liu, Van Tilborg, and Van Dijk also extended and optimized the bit-pair iteration protocol to perform advantage distillation and reconciliation jointly and further increase the advantage-distillation rate [53]. Although the satellite source model is based on a binary DMS $(\mathcal{X}\mathcal{Y}\mathcal{Z}, p_{XYZ})$, the ideas behind the design of advantage protocols generalize to continuous sources. For instance, Naito, Watanabe, Matsumoto, and Uyematsu proposed an advantage-distillation protocol based on a post-selection procedure for a Gaussian satellite model [54]. The notion of advantage-distillation capacity and its relation to secret-key capacity were studied by Muramatsu, Yoshimura, and Davis [55].

Information-reconciliation protocols for secret-key agreement were first introduced by Brassard and Salvail [56] for discrete random variables. In essence, the original reconciliation protocols are simple error-correcting codes that rely on two-way communication to exchange parity checks and identify error locations. Examples of such protocols include CASCADE, proposed by Brassard and Salvail [56], and WINNOW, developed by Buttler, Lamoreaux, Torgerson, Nickel, Donahue, and Peterson [57]. More efficient protocols based on powerful error-control codes can be designed at the cost of increased complexity. For instance, Elkouss, Leverrier, Alléaume, and Boutros optimized a reconciliation protocol based on low-density parity-check codes [58]. Reconciliation protocols for continuous random variables were developed susbsequently by Van Assche, Cardinal, and Cerf [59], Nguyen, Van Assche, and Cerf [60], Bloch, Thangaraj, McLaughlin, and Merolla [61], and Ye, Reznik, and Shah [62].

The Rényi entropy, from which the collision entropy and the min-entropy are derived, was introduced and studied by Rényi [63]. Privacy amplification with universal families of hash functions was analyzed by Bennett, Brassard, Crépeau, and Maurer [64]. The examples of universal families provided in this chapter are due to Carter and Wegman [65, 66], who introduced these functions in the context of authentication. Stinson generalized the work of Carter and Wegman and proved that the minimum size of a universal family of hash functions $\mathcal{G} : \{0, 1\}^n \to \{0, 1\}^k$ is 2^{n-k} [67]. A thorough discussion of the implementation and properties of privacy amplification with almost universal families of hash functions can be found in the textbook of Van Assche [38]. Privacy amplification with extractors was studied by Maurer and Wolf [68], on the basis of the existence

of extractors requiring only a small amount of input randomness as established by Vadhan [69]. The link between reconciliation and privacy amplification was developed by Cachin and Maurer [70, 71]. Although privacy amplification is often described explicitly in terms of universal families of hash functions or extractors, other classes of functions may, in principle, perform the same function. For instance, Muramatsu proved the existence of key-distillation strategies based on low-density parity-check codes [72].

The generic procedure to create strongly secure keys from weakly secure ones presented in this chapter is due to Maurer and Wolf [68]. An alternative (perhaps more direct) approach to strong secrecy was developed by Csiszár [24] for discrete random variables, and was extended to continuous correlations by Nitinawarat [73].

The authentication of the public channel is a crucial assumption used to derive all of the results in this chapter. The problem of secret-key agreement over unauthenticated public channels is significantly more involved; nevertheless, Maurer and Wolf showed that the secret-key capacity remains the same, provided that the DMS of the source model satisfies a condition called the *simulatability* condition [74]. In general, deciding whether a source satisfies the simulatability condition is not straightforward; however, Maurer and Wolf developed an efficient algorithm to check a necessary condition for simulatability [75] and studied privacy amplification protocols for key distillation over unauthenticated channels [76].

There have been a few experimental demonstrations of key-distillation strategies. For instance, Imai, Kobara, and Morozov gathered experimental data to assess the possibility of key agreement with a directional antenna [77]. Ye, Mathur, Reznik, Shah, Trappe, and Mandayam [78] and Chen and Jensen [79] also implemented several key-distillation strategies relying on the reciprocity of fading channels. Conceptually, the key-distillation strategies for wireless channels are similar to those implemented for continuous-variable quantum key distribution, see for instance the quantum key-distribution systems proposed by Grosshans, Van Assche, Wenger, Tualle-Brouri, Cerf, and Grangier [80] and Lodewyck, Bloch, García-Patrón, Fossier, Karpov, Diamanti, Debuisschert, Cerf, Tualle-Brouri, McLaughlin, and Grangier [81]. Chabanne and Fumaroli also analyzed a low-cost key-distillation strategy for RFID tags [82].