Information Theoretic Cryptography for Information Theorists (Notes for a tutorial at ISIT 2017)

Himanshu Tyagi[†] Shun Watanabe[‡]

25th June, 2017

1 Randomness extraction

Randomness extraction refers to generating almost uniform bits as a function of a given random variable X. This will play a central role in all our applications, in particular, in the *privacy amplification* step of secret key agreement. The form of functions that can enable randomness extraction depends on the underlying class of distributions of X for which we want the extraction to work. We begin by reviewing various classes of interest.

1.1 Source Distribution

Definition 1 (Randomness Extractor)**.** For a given source $X \sim \mathrm{P}_X$, an *ε-extractor* of length $\log |\mathcal{K}|$ for P_X consists of a random mapping $F : \mathcal{X} \to \mathcal{K}$, selected using a distribution P_F from a set of mappings $\mathcal{F}$, such that $K = F(X)$ satisfies

$$d\left(\mathrm{P}_{KF}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_F\right) \leq \varepsilon,$$

where $\mathrm{P}_{\mathtt{unif}}$ is the uniform distribution on $\mathcal{K}$ and $d\left(\mathrm{P}, \mathrm{Q}\right)$ is the variational distance given by

$$d\left(\mathrm{P}, \mathrm{Q}\right) := \frac{1}{2} \sum_x \left|\mathrm{P}(x) - \mathrm{Q}(x)\right|.$$

If F constitutes an ε-extractor for every distribution P in a family $\mathcal{P}$, we say that F is an ε-extractor for $\mathcal{P}$.

While the class that has received the most attention in the information theory literature is the class of i.i.d. distributions, we need results for the more general class of *k-sources* introduced by Chor and Goldreich.

Definition 2 (Min-Entropy)**.** The min-entropy of $X \sim \mathrm{P}_X$ is defined as

$$H_{\min}(X) := \min_{x \in \mathcal{X}} \log \frac{1}{\mathrm{P}_X\left(x\right)}.$$

[†]Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India. Email: htyagi@ece.iisc.ernet.in

[‡]Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan. Email: shunwata@cc.tuat.ac.jp

Definition 3 (k-source)**.** A distribution P on $\mathcal{X}$ constitutes a *k-source* if

$$H_{\min}(\mathrm{P}) \geq k.$$

The class of all k-sources on $\mathcal{X}$ is denoted by $\mathcal{P}_k(\mathcal{X})$

Heuristically, a k-source can be regarded as a distribution with at least k-bits of randomness. However, it is not possible to extract even one bit of uniform randomness from $\mathcal{P}_k(\mathcal{X})$ if we allow only deterministic extractors. Indeed, given a mapping $f : \{0,1\}^n \to \{0,1\}$ such that $|f^{-1}(0)| \geq |f^{-1}(1)|$, the uniform distribution on $f^{-1}(0)$ is an $(n-1)$-source but $\mathrm{P}\left(f(X) = 0\right) = 1$. Thus, we must take recourse to randomized extractors. The *leftover hash* lemma given in the next section shows the existence of a randomized ε-extractor for $\mathcal{P}_k$.

1.2 The leftover hash lemma

Definition 4 (2-Universal hash family)**.** A class of functions $\mathcal{F}$ from $\mathcal{X}$ to $\{0,1\}^l$ constitutes a 2-*universal hash family* (2-UHF) if

$$\mathrm{P}\left(F(x) = F(x')\right) \leq \frac{1}{2^l}, \quad \forall x, x' \in \mathcal{X} \text{ s.t. } x \neq x',$$

where F is uniformly distributed over $\mathcal{F}$.

A simple example of such a family of length l over $\{0,1\}^n$ is the family of linear maps $f(x) = Ax$ where A is a binary $n \times l$ matrix and operations are modulo 2.

Definition 5 (Collision entropy or Rényi Entropy of order 2)**.** The *Rényi entropy of order* 2 for a given source $X \sim \mathrm{P}_X$ is defined as

$$H_2(X) := -\log\left(\sum_x \mathrm{P}_X\left(x\right)^2\right).$$

Theorem 1 (Leftover hash lemma)**.** *For a mapping F chosen uniformly at random from a 2-UHF $\mathcal{F}$, $K = F(X)$ satisfies*

$$d\left(\mathrm{P}_{KF}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_F\right) \leq \frac{1}{2}\sqrt{2^{l - H_2(X)}} \tag{1}$$

$$\leq \frac{1}{2}\sqrt{2^{l - H_{\min}(X)}}. \tag{2}$$

Proof. Since $H_2(X) \geq H_{\min}(X)$, it suffices to show (1). For each $F = f$, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned} d(\mathrm{P}_{f(X)}, \mathrm{P}_{\mathtt{unif}}) &= \frac{1}{2}\sum_k |\mathrm{P}_{f(X)}\left(k\right) - \mathrm{P}_{\mathtt{unif}}\left(k\right)| \\ &\leq \frac{1}{2}\sqrt{2^l \sum_k \left(\mathrm{P}_{f(X)}\left(k\right) - \mathrm{P}_{\mathtt{unif}}\left(k\right)\right)^2}. \end{aligned}$$

The term under $\sqrt{\cdot}$ can be evaluated as

$$\sum_k \left(\mathrm{P}_{f(X)}\left(k\right) - \mathrm{P}_{\mathtt{unif}}\left(k\right)\right)^2$$

$$
\begin{aligned}
&= \sum_k \mathrm{P}_{f(X)}(k)^2 - 2\sum_k \mathrm{P}_{f(X)}(k)\,\mathrm{P}_{\mathtt{unif}}(k) + \sum_k \mathrm{P}_{\mathtt{unif}}(k)^2 \\
&= \sum_k \mathrm{P}_{f(X)}(k)^2 - \frac{1}{2^l} \\
&= \sum_{x,x'} \mathrm{P}_X(x)\,\mathrm{P}_X(x')\left(\mathbf{1}[f(x)=f(x')] - \frac{1}{2^l}\right) \\
&= \sum_x \mathrm{P}_X(x)^2\left(1 - \frac{1}{2^l}\right) + \sum_{x\neq x'} \mathrm{P}_X(x)\,\mathrm{P}_X(x')\left(\mathbf{1}[f(x)=f(x')] - \frac{1}{2^l}\right) \\
&\le 2^{-H_2(X)} + \sum_{x\neq x'} \mathrm{P}_X(x)\,\mathrm{P}_X(x')\left(\mathbf{1}[f(x)=f(x')] - \frac{1}{2^l}\right).
\end{aligned}
$$

By taking the average with respect to F, the second term is bounded as

$$
\begin{aligned}
&\sum_f \mathrm{P}_F(f) \sum_{x\neq x'} \mathrm{P}_X(x)\,\mathrm{P}_X(x')\left(\mathbf{1}[f(x)=f(x')] - \frac{1}{2^l}\right) \\
&= \sum_{x\neq x'} \mathrm{P}_X(x)\,\mathrm{P}_X(x')\left(\mathrm{P}\left(F(x)=F(x')\right) - \frac{1}{2^l}\right) \\
&\le 0,
\end{aligned}
$$

where the inequality follows from the property of the 2-UHF. Thus, concavity of $\sqrt{\cdot}$ implies

$$
\begin{aligned}
d(\mathrm{P}_{KF}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_F) &= \sum_f \mathrm{P}_F(f)\, d(\mathrm{P}_{f(X)}, \mathrm{P}_{\mathtt{unif}}) \\
&\le \frac{1}{2}\sqrt{2^l \sum_f \mathrm{P}_F(f) \sum_k \left(\mathrm{P}_{f(X)}(k) - \mathrm{P}_{\mathtt{unif}}(k)\right)^2} \\
&\le \frac{1}{2}\sqrt{2^{l-H_2(X)}}.
\end{aligned}
$$

■

Note that the bound in Theorem 1 holds for any source. Therefore, for every $\mathrm{P}_X \in \mathcal{P}_k(\mathcal{X})$,

$$
d\left(\mathrm{P}_{KF}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_F\right) \le \varepsilon
$$

as long as

$$
l \le k - 2\log\frac{1}{2\varepsilon}.
$$

Thus, F constitutes an ε-extractor of length $k - 2\log(1/2\varepsilon)$ for $\mathcal{P}_k(\mathcal{X})$.

1.3 Leftover hash lemma with side-information

In applications, we need a variant of the leftover hash lemma where we seek almost independence of K not only from F but also jointly from an additional side information Z. In this context, min-entropy is replaced by the corresponding *conditional min-entropy.*

Definition 6 (Conditional min-entropy)**.** For distributions P_{XZ} and Q_Z, the *conditional min-entropy* of P_{XZ} given Q_Z is defined as

$$H_{\min}(\mathrm{P}_{XZ}|\mathrm{Q}_Z) := \min_{x\in\mathcal{X},z\in\mathtt{supp}(\mathrm{Q}_Z)} \log \frac{\mathrm{Q}_Z(z)}{\mathrm{P}_{XZ}(x,z)}.$$

Then, the conditional min-entropy of P_{XZ} given Z is defined as

$$H_{\min}(\mathrm{P}_{XZ}|Z) := \max_{\mathrm{Q}_Z} H_{\min}(\mathrm{P}_{XZ}|\mathrm{Q}_Z). \tag{3}$$

Definition 7 (Conditional collision entropy)**.** For distributions P_{XZ} and Q_Z, the conditional collision entropy of P_{XZ} given Q_Z is defined as

$$H_2(\mathrm{P}_{XZ}|\mathrm{Q}_Z) := -\log \sum_{x\in\mathcal{X},z\in\mathtt{supp}(\mathrm{Q}_Z)} \frac{\mathrm{P}_{XZ}(x,z)^2}{\mathrm{Q}_Z(z)}.$$

Theorem 2. *Given a distribution* P_{XZ} *on* $\mathcal{X}\times\mathcal{Z}$*, for a mapping* F *chosen uniformly at random from a* 2*-UHF* $\mathcal{F}$*,* $K = F(X)$ *satisfies*

$$\begin{aligned} d\left(\mathrm{P}_{KZF}, \mathrm{P}_{\mathtt{unif}}\times\mathrm{P}_Z\times\mathrm{P}_F\right) &\le \frac{1}{2}\sqrt{2^{l-H_2(\mathrm{P}_{XZ}|Z)}} && (4)\\ &\le \frac{1}{2}\sqrt{2^{l-H_{\min}(\mathrm{P}_{XZ}|Z)}}. && (5)\end{aligned}$$

1.4 Smoothing

The notion of conditional min-entropy introduced in the previous section, while fundamental, is not easy to evaluate for the interesting case of product distributions. An alternative quantity which can be evaluated easily is a *smooth version* of conditional min-entropy, which is defined as the maximum of conditional min-entropy over all distributions which are close to the distribution P_{XZ}. In fact, in the argument of Section 1.3, P_{XZ} need not be normalized to 1. Likewise, it is possible to consider smoothing over all subnormalized distributions close to P_{XZ}.

Definition 8 (Smooth conditional min-entropy)**.** For distributions P_{XZ} and Q_Z, and smoothing parameter $0\le\varepsilon<1$, define

$$H^{\varepsilon}_{\min}(\mathrm{P}_{XZ}|\mathrm{Q}_Z) := \max_{\tilde{\mathrm{P}}_{XZ}\in\mathcal{B}_\varepsilon(\mathrm{P}_{XZ})} H_{\min}(\tilde{\mathrm{P}}_{XZ}|\mathrm{Q}_Z),$$

where

$$\mathcal{B}_\varepsilon(\mathrm{P}_{XZ}) := \left\{\tilde{\mathrm{P}}_{XZ}\in\mathcal{P}_{\mathtt{sub}}(\mathcal{X}\times\mathcal{Z}) : d(\tilde{\mathrm{P}}_{XZ},\mathrm{P}_{XZ})\le\varepsilon\right\},$$

and $\mathcal{P}_{\mathtt{sub}}(\mathcal{X}\times\mathcal{Z})$ is the set of all subnormalized distributions on $\mathcal{X}\times\mathcal{Z}$. Then, the smooth conditional min-entropy of P_{XZ} given Z is defined as

$$H^{\varepsilon}_{\min}(\mathrm{P}_{XZ}|Z) := \max_{\mathrm{Q}_Z} H^{\varepsilon}_{\min}(\mathrm{P}_{XZ}|\mathrm{Q}_Z). \tag{6}$$

The smooth version of conditional Rényi entropy of order 2, $H^{\varepsilon}_2(\mathrm{P}_{XZ}|Z)$, is defined similarly.

The next corollary follows from Theorem 2 by the triangle inequality

$$d\left(\mathrm{P}_{KZF},\mathrm{P}_{\mathtt{unif}}\times\mathrm{P}_Z\times\mathrm{P}_F\right) \le d\left(\mathrm{P}_{KZF},\tilde{\mathrm{P}}_{KZF}\right) + d\left(\mathrm{P}_Z,\tilde{\mathrm{P}}_Z\right) + d\left(\tilde{\mathrm{P}}_{KZF},\mathrm{P}_{\mathtt{unif}}\times\tilde{\mathrm{P}}_Z\times\mathrm{P}_F\right).$$

Corollary 3 (Leftover Hash Lemma with Smoothing). *For a mapping F chosen uniformly at random from a 2-UHF $\mathcal{F}$, $K = F(X)$ satisfies*

$$d\left(\mathrm{P}_{KZF}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_Z \times \mathrm{P}_F\right) \leq 2\varepsilon + \frac{1}{2}\sqrt{2^{l - H_2^{\varepsilon}(\mathrm{P}_{XZ}|Z)}}. \tag{7}$$

$$\leq 2\varepsilon + \frac{1}{2}\sqrt{2^{l - H_{\min}^{\varepsilon}(\mathrm{P}_{XZ}|Z)}}. \tag{8}$$

Thus, $K = F(X)$ and F satisfy

$$d(\mathrm{P}_{KZF}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_Z \times \mathrm{P}_F) \leq \varepsilon$$

if, for $0 < \eta \leq \varepsilon$,

$$l \geq H_{\min}^{(\varepsilon-\eta)/2}(\mathrm{P}_{XZ}|\mathrm{Q}_Z) - \log(1/4\eta^2) - 1. \tag{9}$$

A common smoothing technique is to consider a set $\mathcal{B}_\varepsilon(\mathrm{P}_{XZ})$ comprimising the subnormalized distribution $\tilde{\mathrm{P}}_{XZ}$ obtained by removing the set of "nontypical" sequences, namely,

$$\tilde{\mathrm{P}}_{XZ}(x,z) = \mathrm{P}_{XZ}\left(x,z\right) \mathbf{1}\left[\log \frac{1}{\mathrm{P}_{X|Z}\left(x|z\right)} > r\right].$$

For this choice, we have

$$H_{\min}^{\varepsilon/2}(\mathrm{P}_{XZ}|\mathrm{P}_Z) \geq \sup\left\{r : \mathrm{P}\left(\log \frac{1}{\mathrm{P}_{X|Z}\left(X|Z\right)} \leq r\right) \leq \varepsilon\right\}. \tag{10}$$

When P_{XZ}^n is an i.i.d. distribution, (9), (10), together with the law of large number imply

$$l \geq nH(X|Z) - o(n).$$

Furthermore, by applying the central limit theorem, we have

$$l \geq nH(X|Z) - \sqrt{nV}\mathsf{Q}^{-1}(\varepsilon) - \mathcal{O}(\log n), \tag{11}$$

which can be shown to be optimal up to a second-order term.

Remark 1. The min-entropy bound (8) suffices to derive the asymptotic limit up to the second-order term (11). However, when security parameter ε is very small, such as the large deviation regime, it is known that the collision entropy bound (7) provides a much tighter bound.

Finally, we review a useful variant of the leftover hash lemma. Suppose that the side-information comprises V on $\mathcal{V}$ and Z on $\mathcal{Z}$, and consider the joint distribution P_{XVZ}.

Theorem 4. *For a mapping F chosen uniformly at random from a 2-UHF $\mathcal{F}$, $K = F(X)$ satisfies*

$$d(\mathrm{P}_{KVZF}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{VZ} \times \mathrm{P}_F) \leq 2\varepsilon + \frac{1}{2}\sqrt{|\mathcal{V}|2^{l - H_{\min}^{\varepsilon}(\mathrm{P}_{XZ}|Z)}}. \tag{12}$$

In other word, if an extra l-bit side information V is revealed, the extracted randomness K reduces by at most l bits.

2 Secret key agreement

2.1 Problem description

We consider the problem of secret key agreement using interactive public communication by two (trusted) parties $\mathcal{P}_1$ and $\mathcal{P}_2$ observing, respectively, random variables X and Y taking values in countable sets $\mathcal{X}$ and $\mathcal{Y}$. Upon making these observations, the parties communicate interactively over a public communication channel that is accessible by an eavesdropper. We assume that the communication channel is error-free and authenticated. Specifically, the communication is sent over multiple rounds of interaction using an interactive communication protocol π. We restrict to *tree protocols* represented by a labeled binary-tree with each node labeled by one of the parties, which communicates when the protocol reaches that node. The protocol starts at the root and, at each node, moves to the left- or right-child based on the 1-bit communicated by the party corresponding to that node. Communication at each node is a function of the observation of the party corresponding to that node and a *locally generated* randomness denoted by[1] U_x and U_y. The overall interactive communication for fixed values of (X, Y, U_x, U_y) is called a *transcript* of the protocol and is a random variable denoted denoted by Π. The transcript Π of the protocol is available to the eavesdropper. In addition, the eavesdropper observes a random variable Z taking values in a countable set $\mathcal{Z}$. In this section, we assume that the joint distribution P_{XYZ} is known to the parties as well as the eavesdropper.

Definition 9 (Secret keys)**.** A random variable K with range $\mathcal{K}$ constitutes an (ε, δ)-*secret key* ((ε, δ)-SK) of length $\log |\mathcal{K}|$ if there exist an interactive communication protocol π and functions K_x and K_y of (U_x, X, Π) and (U_y, Y, Π), respectively, such that the following two conditions are satisfied

$$\Pr\left(K_x = K_y = K\right) \geq 1 - \varepsilon, \tag{13}$$

$$d\left(\mathrm{P}_{K\Pi Z}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{\Pi Z}\right) \leq \delta, \tag{14}$$

where $\mathrm{P}_{\mathtt{unif}}$ is the uniform distribution on $\mathcal{K}$.

The first condition above represents the *reliability* of the secret key and the second condition guarantees *secrecy*.

Definition 10. Given $\varepsilon, \delta \in [0, 1)$, the supremum over the lengths $\log |\mathcal{K}|$ of an (ε, δ)-SK is denoted by $S_{\varepsilon,\delta}(X, Y|Z)$.

2.2 Secret key agreement protocols

A secret key agreement protocol typically consists of two steps: An *information reconciliation step* where the parties engage in public communication to convert their correlated observations into shared random bits, termed *common randomness*; and the *privacy amplification step* where a secure randomness almost independent of the public communication is extracted from the common randomness. Below we illustrate a standard information reconciliation technique and a privacy amplification technique. A common tool used in both the steps is a 2-UHF (see Definition 4).

Information reconciliation At a high level, our two-party information reconciliation procedure entails the two parties agreeing on X and is described as follows:

[1] The random variables U_x and U_y are mutually independent and independent jointly of (X, Y).

1. $\mathcal{P}_2$ upon observing $Y = y$ forms a list $\mathcal{L}_y \subset \mathcal{X}$ of guesses for X.

2. $\mathcal{P}_1$ chooses a random member F of a 2-UHF of length t over $\mathcal{X}$ and sends $(F, F(X))$ to $\mathcal{P}_2$.

3. $\mathcal{P}_2$ finds a unique $\hat{x} \in \mathcal{L}_y$ such that $F(y) = F(\hat{x})$.

There are two possible error events for this process:

$$\mathcal{E}_1 = \{X \notin \mathcal{L}_Y\} \quad \text{or} \quad \mathcal{E}_2 = \{\exists\, \hat{x} \neq X \text{ s.t. } \hat{x} \in \mathcal{L}_Y \text{ and } F(\hat{x}) = F(X)\}.$$

By choosing $\mathcal{L}_y$ and l such that $\Pr(X \in \mathcal{L}_Y) \geq 1-\varepsilon$ and $l \geq \max_y \log|\mathcal{L}_y| + \gamma$, the 2-UHF property and union bound yield that the probability of error is bounded above by $\varepsilon + 2^{-\gamma}$. A standard choice for the list $\mathcal{L}_y$ consists of those x which have significant likelihood given y, namely

$$\mathcal{L}_y = \{x : -\log \mathrm{P}_{X|Y}(x|y) \leq \lambda\}; \tag{15}$$

for this choice,

$$|\mathcal{L}_y| \leq 2^{\lambda}, \qquad \forall y \in \mathcal{Y}. \tag{16}$$

Note that for the case of i.i.d. observations, this choice with $\lambda = n(H(X|Y) + \eta)$ coincides with a standard *typical set*

$$\left\{\mathbf{x} : -\frac{1}{n}\sum_{i=1}^{n} \log \mathrm{P}_{X|Y}(x_i|y_i) \leq H(X|Y) + \eta\right\}.$$

Privacy amplification Heuristically, a privacy amplification protocol applies a random function to a random variable U such that the output of the function is almost independent of another random variable V. To do this formally, we take recourse to the final form of the leftover hash lemma given in Theorem 4.

We now describe our complete secret key agreement protocol.

1. Use the information reconciliation protocol of the previous section which communicates t bits.

2. *Privacy amplification step:* Let $\hat{X}$ be the estimate of X by $\mathcal{P}_2$.

   (a) $\mathcal{P}_1$ chooses F' uniformly over a 2-UHF of length l over $\mathcal{X}$ and sends F' to $\mathcal{P}_2$.

   (b) $\mathcal{P}_1$ generates $K_x = F'(X)$ and $\mathcal{P}_2$ generates $K_y = F'(\hat{X})$.

The scheme above involves two 2-UHFs: A family $\mathcal{F}$ of length t used in the information reconciliation step and another, say $\mathcal{F}'$, of length l used in the privacy amplification step. It follows by Theorem 4 that the secret key above satisfies (14) if

$$l \leq H_{\min}^{\eta}(\mathrm{P}_{XZ}|Z) - t - 2\log\frac{1}{\delta - \eta}. \tag{17}$$

2.3 Information theoretic limits: Achievability

Using the development of the previous section, we can find a lower bound $S_{\varepsilon,\delta}(X, Y|Z)$ by appropriately choosing t and then fixing l to satisfy (17). Specifically, let the lists $\mathcal{L}_y$ be given by (15). Then, by (16) and (17), on choosing $t = \lambda + \gamma$ we get an (ε, δ)-SK of length

$$H_{\min}^{\eta}(\mathrm{P}_{XZ}|Z) - \lambda - \gamma - 2\log\frac{1}{\delta - \eta},$$

where $\lambda > 0$ satisfies

$$\mathrm{P}_{XY}\left(\{(x,y) : -\log \mathrm{P}_{X|Y}(x|y) \leq \lambda\}\right) \geq 1 - \varepsilon + 2^{-\gamma}.$$

For the i.i.d. case, we can evaluate an explicit bound for $H_{\min}^{\eta}(\mathrm{P}_{XZ}|Z)$ and λ above using, say, the Chebyshev's inequality to obtain the following result.

Theorem 5. *For an i.i.d. distribution* $\mathrm{P}_{X^nY^nZ^n}$ *and* $0 \leq \varepsilon, \delta < 1$

$$S_{\varepsilon,\delta}(X^n, Y^n|Z^n) \geq n(H(X|Z) - H(X|Y)) + o(n).$$

3 Converse techniques

In this section, we present techniques for deriving converse bounds (impossibility results) for the secret key agreement problem.

3.1 A basic converse bound

We start with a basic converse bound based on Fano's inequality. The following simple, but fundamental, property of interactive communication protocol will be used throughout.

Lemma 6. *For any protocol* Π*,*

$$I(X \wedge Y|Z,\Pi) \leq I(X \wedge Y|Z).$$

In particular, if $\mathrm{P}_{XYZ} = \mathrm{P}_{X|Z}\mathrm{P}_{Y|Z}\mathrm{P}_Z$*, then* $\mathrm{P}_{XYZ\Pi} = \mathrm{P}_{X|Z\Pi}\mathrm{P}_{Y|Z\Pi}\mathrm{P}_{Z\Pi}$*.*

By combining Lemma 6 with Fano's inequality, we can derive the next result.

Theorem 7. *For every* $0 \leq \varepsilon, \delta < 1$ *with* $0 \leq \varepsilon + \delta < 1$*, it holds that*

$$S_{\varepsilon,\delta}(X,Y|Z) \leq \frac{I(X \wedge Y|Z) + h(\varepsilon) + h(\delta)}{1 - \varepsilon - \delta}$$

Proof outline By Fano's inequality and by the continuity of Shannon's entropy, an (ε,δ)-SK with estimates K_1, K_2 for $\mathcal{P}_1, \mathcal{P}_2$, respectively, satisfies

$$\begin{aligned} H(K_1|K_2) &\leq h(\varepsilon) + \varepsilon \log|\mathcal{K}|, \\ \log|\mathcal{K}| - H(K_1|Z,\Pi) &\leq h(\delta) + \delta \log|\mathcal{K}|. \end{aligned}$$

Thus,

$$\begin{aligned} \log|\mathcal{K}| &\leq H(K_1|Z,\Pi) + h(\delta) + \delta \log|\mathcal{K}| \\ &= I(K_1 \wedge K_2|Z,\Pi) + H(K_1|K_2,Z,\Pi) + h(\delta) + \delta \log|\mathcal{K}| \\ &\leq I(K_1 \wedge K_2|Z,\Pi) + h(\varepsilon) + h(\delta) + (\varepsilon+\delta)\log|\mathcal{K}| \\ &\leq I(X \wedge Y|Z) + h(\varepsilon) + h(\delta) + (\varepsilon+\delta)\log|\mathcal{K}|, \end{aligned}$$

where the final inequality uses the data processing inequality and Lemma 6. ■

3.2 Conditional independence testing bound

The next bound we cover relates the secret key agreement problem to binary hypothesis testing. We begin with a review of binary hypothesis testing.

For distributions P and Q on $\mathcal{X}$, a test is described by a (stochastic) mapping $\mathrm{T} : \mathcal{X} \to \{0, 1\}$. Let

$$\beta_\varepsilon(\mathrm{P}, \mathrm{Q}) := \inf_{\mathrm{T}:\mathrm{P}[\mathrm{T}]\geq 1-\varepsilon} \mathrm{Q}[\mathrm{T}],$$

where

$$\mathrm{P}[\mathrm{T}] = \sum_x \mathrm{P}(x)\mathrm{T}(0|x),$$
$$\mathrm{Q}[\mathrm{T}] = \sum_x \mathrm{Q}(x)\mathrm{T}(0|x).$$

When P^n and Q^n are i.i.d. distributions, Stein's lemma says

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta_\varepsilon(\mathrm{P}^n, \mathrm{Q}^n) = D(\mathrm{P}\|\mathrm{Q}), \ \forall 0 < \varepsilon < 1. \tag{18}$$

The following theorem gives a bound for secret key length in terms of $\beta_\varepsilon(\cdot, \cdot)$.

Theorem 8. *Given $0 \leq \varepsilon, \delta < 1$ and $0 < \eta < 1 - \varepsilon - \delta$, it holds that*

$$S_{\varepsilon,\delta}(X, Y|Z) \leq -\log \beta_{\varepsilon+\delta+\eta}(\mathrm{P}_{XYZ}, \mathrm{Q}_{XYZ}) + 2\log(1/\eta)$$

for any $\mathrm{Q}_{XYZ} = \mathrm{Q}_{X|Z}\mathrm{Q}_{Y|Z}\mathrm{Q}_Z$.

Proof Outline The first key observation is that (ε, δ)-SK implies the following combined security criterion.

Lemma 9. *An (ε, δ)-SK with estimates K_1, K_2 for $\mathcal{P}_1, \mathcal{P}_2$, respectively, satisfies the following combined security criterion:*

$$d(\mathrm{P}_{K_1K_2Z\Pi}, \mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi}) \leq \varepsilon + \delta, \tag{19}$$

where

$$\mathrm{P}^{(2)}_{\mathtt{unif}}(k_1, k_2) := \frac{\mathbf{1}[k_1 = k_2]}{|\mathcal{K}|}.$$

The core of the proof of Theorem 8 is the following lemma.

Lemma 10. *For any K_1, K_2 taking values in $\mathcal{K}$ and satisfying (19) and any $\mathrm{Q}_{K_1K_2Z\Pi}$ of the form $\mathrm{Q}_{K_1|Z\Pi}\mathrm{Q}_{K_2|Z\Pi}\mathrm{Q}_{Z\Pi}$, it holds that*

$$\log |\mathcal{K}| \leq -\log \beta_{\varepsilon+\delta+\eta}(\mathrm{P}_{K_1K_2Z\Pi}, \mathrm{Q}_{K_1K_2Z\Pi}) + 2\log(1/\eta).$$

For a given protocol generating an (ε, δ)-SK from P_{XYZ}, let $\mathrm{Q}_{K_1K_2Z\Pi}$ be the joint distribution obtained by running the same protocol for $\mathrm{Q}_{XYZ} = \mathrm{Q}_{X|Z}\mathrm{Q}_{Y|Z}\mathrm{Q}_Z$. By Lemma 6, $\mathrm{Q}_{K_1K_2Z\Pi}$ satisfies the assumption of Lemma 10. Thus, by applying the data processing inequality with respect to $\beta_\varepsilon(\cdot, \cdot)$, Theorem 8 follows from Lemma 10.

To prove Lemma 10, we construct a likelihood ratio test using the secret key agreement protocol. The key idea is to consider the likelihood ratio between $\mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi}$ and $\mathrm{Q}_{K_1K_2Z\Pi}$, instead of the likelihood ratio test between $\mathrm{P}_{K_1K_2Z\Pi}$ and $\mathrm{Q}_{K_1K_2Z\Pi}$. Indeed, consider the test with acceptance region defined by

$$\mathcal{A} := \left\{(k_1, k_2, z, \tau) : \log \frac{\mathrm{P}^{(2)}_{\mathtt{unif}}(k_1, k_2)}{\mathrm{Q}_{K_1K_2|Z\Pi}\left(k_1, k_2|z, \tau\right)} \geq \lambda\right\},$$

where

$$\lambda = \log|\mathcal{K}| - 2\log(1/\eta).$$

Then, a change-of-measure argument yields the following bound for the type II error probability:

$$\begin{aligned}
\mathrm{Q}_{K_1K_2Z\Pi}\left(\mathcal{A}\right) &= \sum_{z,\tau} \mathrm{Q}_{Z\Pi}\left(z, \tau\right) \sum_{(k_1,k_2):(k_1,k_2,z,\tau)\in\mathcal{A}} \mathrm{Q}_{K_1K_2|Z\Pi}\left(k_1, k_2|z, \tau\right) \\
&\leq 2^{-\lambda} \sum_{z,\tau} \mathrm{Q}_{Z\Pi}\left(z, \tau\right) \sum_{(k_1,k_2)} \mathrm{P}^{(2)}_{\mathtt{unif}}(k_1, k_2) \\
&= \frac{1}{|\mathcal{K}|\eta^2}. \qquad (20)
\end{aligned}$$

On the other hand, the security condition (19) yields the following bound for the type I error probability:

$$\begin{aligned}
\mathrm{P}_{K_1K_2Z\Pi}\left(\mathcal{A}^c\right) &\leq d(\mathrm{P}_{K_1K_2Z\Pi}, \mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi}) + \mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi}(\mathcal{A}^c) \\
&\leq \varepsilon + \delta + \mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi}(\mathcal{A}^c), \qquad (21)
\end{aligned}$$

where the first inequality follows from the definition of the variational distance. Furthermore, the last term above can be expressed as

$$\begin{aligned}
\mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi}(\mathcal{A}^c) &= \sum_{z,\tau} \mathrm{P}_{Z\Pi}\left(z, \tau\right) \frac{1}{|\mathcal{K}|} \sum_{k} \mathbf{1}[(k, k, z, \tau) \in \mathcal{A}^c] \\
&= \sum_{z,\tau} \mathrm{P}_{Z\Pi}\left(z, \tau\right) \frac{1}{|\mathcal{K}|} \sum_{k} \mathbf{1}[\mathrm{Q}_{K_1K_2|Z\Pi}\left(k, k|z, \tau\right) |\mathcal{K}|^2\eta^2 > 1].
\end{aligned}$$

The inner sum can be bounded further as

$$\begin{aligned}
&\sum_{k} \mathbf{1}[\mathrm{Q}_{K_1K_2|Z\Pi}\left(k, k|z, \tau\right) |\mathcal{K}|^2\eta^2 > 1] \\
&\leq \sum_{k} \left(\mathrm{Q}_{K_1K_2|Z\Pi}\left(k, k|z, \tau\right) |\mathcal{K}|^2\eta^2\right)^{\frac{1}{2}} \\
&= |\mathcal{K}|\eta \sum_{k} \mathrm{Q}_{K_1K_2|Z\Pi}\left(k, k|z, \tau\right)^{\frac{1}{2}} \\
&= |\mathcal{K}|\eta \sum_{k} \mathrm{Q}_{K_1|Z\Pi}\left(k|z, \tau\right)^{\frac{1}{2}} \mathrm{Q}_{K_2|Z\Pi}\left(k|z, \tau\right)^{\frac{1}{2}},
\end{aligned}$$

where the last equality uses Lemma 6. Next, an application of the Cauchy-Schwartz inequality yields

$$\sum_{k} \mathrm{Q}_{K_1|Z\Pi}\left(k|z, \tau\right)^{\frac{1}{2}} \mathrm{Q}_{K_2|Z\Pi}\left(k|z, \tau\right)^{\frac{1}{2}}$$

$$
\begin{aligned}
&\leq \Bigg(\sum_{k_1} \mathrm{Q}_{K_1|Z\Pi}\left(k_1|z,\tau\right) \Bigg)^{\frac{1}{2}} \Bigg(\sum_{k_2} \mathrm{Q}_{K_2|Z\Pi}\left(k_2|z,\tau\right) \Bigg)^{\frac{1}{2}} \\
&= 1.
\end{aligned}
$$

Combining the bounds above, we have

$$
\mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi}(\mathcal{A}^c) \leq \eta,
$$

which, together with (20) and (21), implies Lemma 10. ■

3.3 Bounds using monotones

The final bound that we describe is based on defining a "monotone," namely a function which increases or decreases monotonically as the protoocol proceeds. Heuristically, such monotones are formed by carefully examining the steps in the standard converse proofs and identifying the abstract properties that enable the proof.

Definition 11 (Monotone)**.** For a given joint distribution P_{XYZ}, a non-negative function $M_{\varepsilon,\delta}(X,Y|Z)$ of P_{XYZ} constitutes a monotone for secret key agreement if it satisfies the following properties:

1. $M_{\varepsilon,\delta}(X,Y|Z)$ does not increase by a local processing, i.e., for any X' satisfying $X' -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y,Z)$,

$$
M_{\varepsilon,\delta}(X,Y|Z) \geq M_{\varepsilon,\delta}(X',Y|Z);
$$

similarly, for any $Y' -\!\!\circ\!\!- Y -\!\!\circ\!\!- (X,Z)$,

$$
M_{\varepsilon,\delta}(X,Y|Z) \geq M_{\varepsilon,\delta}(X,Y'|Z).
$$

2. $M_{\varepsilon,\delta}(X,Y|Z)$ does not increase by any interactive communication, i.e., for any protocol π,

$$
M_{\varepsilon,\delta}(X,Y|Z) \geq M_{\varepsilon,\delta}((X,\Pi),(Y,\Pi)|(Z,\Pi)).
$$

3. For any (ε,δ)-SK (K_1,K_2) generated by protocol π,

$$
\log|\mathcal{K}| \leq M_{\varepsilon,\delta}((K_1,\Pi),(K_2,\Pi)|(Z,\Pi)) + \Delta(\varepsilon,\delta) \tag{22}
$$

for a suitable $\Delta(\varepsilon,\delta) \geq 0$.

Proposition 11. *For $0 \leq \varepsilon,\delta < 1$ and a monotone $M_{\varepsilon,\delta}(X,Y|Z)$ satisfying the properties in Definition 11, it holds that*

$$
S_{\varepsilon,\delta}(X,Y|Z) \leq M_{\varepsilon,\delta}(X,Y|Z) + \Delta(\varepsilon,\delta).
$$

Proof Outline For any (ε,δ), Propertiess 3, 1, and 2 imply

$$
\begin{aligned}
\log|\mathcal{K}| &\leq M_{\varepsilon,\delta}((K_1,\Pi),(K_2,\Pi)|(Z,\Pi)) + \Delta(\varepsilon,\delta) \\
&\leq M_{\varepsilon,\delta}((X,\Pi),(Y,\Pi)|(Z,\Pi)) + \Delta(\varepsilon,\delta) \\
&\leq M_{\varepsilon,\delta}(X,Y|Z) + \Delta(\varepsilon,\delta).
\end{aligned}
$$

■

Example 1 (Conditional Mutual Information)**.** For

$$M_{\varepsilon,\delta}(X,Y|Z) = \frac{1}{1-\varepsilon-\delta} I(X \wedge Y|Z),$$

in the manner of the proof of Theorem 7, we can verify that $M_{\varepsilon,\delta}(X,Y|Z)$ satisfies the properties in Definition 11 with $\Delta(\varepsilon,\delta) = \frac{h(\varepsilon)+h(\delta)}{1-\varepsilon-\delta}$ for $0 \le \varepsilon+\delta < 1$.

Example 2 (Intrinsic Information)**.** For

$$M_{\varepsilon,\delta}(X,Y|Z) = \frac{1}{1-\varepsilon-\delta} \inf_{Z' \text{-}\!\circ\!\text{-} Z \text{-}\!\circ\!\text{-} (X,Y)} I(X \wedge Y|Z'),$$

noting that

$$H(K_1|Z,\Pi) \le H(K_1|Z',\Pi), \tag{23}$$

we can verify that $M_{\varepsilon,\delta}(X,Y|Z)$ satisfies the properties in Definition 11 with $\Delta(\varepsilon,\delta) = \frac{h(\varepsilon)+h(\delta)}{1-\varepsilon-\delta}$ for $0 \le \varepsilon+\delta < 1$.

Example 3 (Conditional Independence Testing)**.** Let

$$M_{\varepsilon,\delta}(X,Y|Z) = \inf_{\mathrm{Q}_{XYZ} \in \mathcal{Q}_{\mathtt{CI}}} \big[-\log \beta_{\varepsilon+\delta+\eta}(\mathrm{P}_{XYZ}, \mathrm{Q}_{XYZ}) \big], \tag{24}$$

where $\mathcal{Q}_{\mathtt{CI}}$ is the set of all conditionally independent distributions. Then, we can verify that $M_{\varepsilon,\delta}(X,Y|Z)$ satisfies[2] the properties in Definition 11 with $\Delta(\varepsilon,\delta) = 2\log(1/\eta)$ for $0 < \eta < 1-\varepsilon-\delta$.

4 Applications of the conditional independence testing bound

4.1 Strong converse for secret key agreement

For the case of i.i.d. observations, it is of interest to characterize the maximum possible rate of a secret key, namely the *secret key capacity.* Specifically, for an n-length i.i.d. sequence (X^n, Y^n, Z^n), the (ε,δ)-SK capacity $C_{\varepsilon,\delta}(X,Y|Z)$ is defined as

$$C_{\varepsilon,\delta}(X,Y|Z) = \liminf_{n\to\infty} \frac{1}{n} S_{\varepsilon,\delta}(X^n, Y^n|Z^n).$$

Our achievability result in Theorem 5 shows that

$$C_{\varepsilon,\delta}(X,Y|Z) \ge H(X|Z) - H(X|Y). \tag{25}$$

On the other hand, by using the conditional independence testing bound with an appropriate choice of Q_{XYZ}, it follows that

$$C_{\varepsilon,\delta}(X,Y|Z) \le \liminf_{n\to\infty} -\frac{1}{n} \log \beta_{\varepsilon+\delta+\eta}(\mathrm{P}_{X^nY^nZ^n}, \mathrm{P}_{X^n|Z^n}\mathrm{P}_{Y^n|Z^n}\mathrm{P}_{Z^n}).$$

Thus, by an application of the Stein's lemma (cf. (18)) and by noting $D(\mathrm{P}_{XYZ} \| \mathrm{P}_{X|Z}\mathrm{P}_{Y|Z}\mathrm{P}_Z) = I(X \wedge Y|Z)$, we have

$$C_{\varepsilon,\delta}(X,Y|Z) \le I(X \wedge Y|Z), \tag{26}$$

whenever $\varepsilon+\delta < 1$. The following theorem is obtained by combining (25) and (26).

[2] More specifically, Property 1 follows from the data processing inequality with respect to $\beta_\varepsilon(\cdot,\cdot)$, Property 2 follows from Lemma 6, and Property 3 follows from Lemma 10.

Theorem 12 (Secret key capacity, with strong converse). *For X, Y, Z such that $\mathrm{P}_{XYZ} = \mathrm{P}_{XZ}\mathrm{P}_{Y|Z}$ and $0 < \varepsilon, \delta$ such that $\varepsilon + \delta < 1$,*

$$C_{\varepsilon,\delta}(X, Y|Z) = I(X \wedge Y|Z).$$

For the case when $\varepsilon + \delta > 1$, it is easy to see that $C_{\varepsilon,\delta}(X, Y|Z) = \infty$.

4.2 Converse for oblivious transfer

Next, we describe the *oblivious transfer* (OT) problem. Suppose that $\mathcal{P}_1$ generates K_0 and K_1, distributed uniformly over $\{0,1\}^l$, and $\mathcal{P}_2$ generates B, distributed uniformly over $\{0,1\}$, as inputs to an OT protocol. The random variables K_0, K_1, and B are assumed to be mutually independent. The goal of an OT protocol is for $\mathcal{P}_2$ to obtain K_B in such a manner that B is concealed from $\mathcal{P}_1$ and $K_{\overline{B}}$ is concealed from $\mathcal{P}_2$, where $\overline{B} = 1 \oplus B$. Furthermore, $\mathcal{P}_1$ and $\mathcal{P}_2$ observe, respectively, X_1 and X_2, as a resource to implement an OT protocol, where (X_1, X_2) are independent jointly of (K_0, K_1, B). During the protocol, the parties are allowed to communicate interactively. In general, the parties are allowed to use local randomization; for simplicity of presentation, we restrict ourselves to protocols without local randomization. However, our results remain valid even when local randomization is allowed.

Definition 12 (Oblivious transfer). An execution of a protocol realizing an $(\varepsilon, \delta_1, \delta_2)$-OT (for a passive adversary) of length l consists of an interactive communication protocol π and an estimate $\hat{K} = \hat{K}(X_2, B, \Pi)$ by $\mathcal{P}_2$ such that the following conditions are satisfied:

$$\begin{aligned} \Pr\left(K_B \neq \hat{K}\right) &\leq \varepsilon, \\ d\left(\mathrm{P}_{K_{\overline{B}}X_2B\Pi}, \mathrm{P}_{K_{\overline{B}}} \times \mathrm{P}_{X_2B\Pi}\right) &\leq \delta_1, \\ d\left(\mathrm{P}_{BK_0K_1X_1\Pi}, \mathrm{P}_B \times \mathrm{P}_{K_0K_1X_1\Pi}\right) &\leq \delta_2, \end{aligned}$$

where $\overline{B} = 1 \oplus B$. The first condition above denotes the reliability of OT, while the second and the third conditions ensure secrecy for party 1 and 2, respectively. Denote by $L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2)$ the largest l such that a protocol realizing an $(\varepsilon, \delta_1, \delta_2)$-OT of length l exists.

When the underlying observations X_1, X_2 consist of n-length i.i.d. sequences X_1^n, X_2^n with common distribution $\mathrm{P}_{X_1X_2}$, it is known that $L_{\varepsilon,\delta_1,\delta_2}(X_1^n, X_2^n)$ may grow linearly with n; the largest rate of growth is termed the *OT capacity*.

Definition 13 (OT capacity). For $0 < \varepsilon < 1$, the ε-OT capacity of (X_1, X_2) is defined[3] as

$$C_\varepsilon(X_1, X_2) = \lim_{\delta_1,\delta_2 \to 0} \liminf_{n \to \infty} \frac{1}{n} L_{\varepsilon,\delta_1,\delta_1}(X_1^n, X_2^n).$$

We derive an upper bound for $L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2)$ which in turn yields an upper bound for $C_\varepsilon(X_1, X_2)$ for every $0 < \varepsilon < 1$.

Theorem 13 (Single-shot bound for OT length). *For random variables X_1, X_2, the following inequalities hold:*

$$L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) \leq -\log \beta_\eta\left(\mathrm{P}_{X_1X_2}, \mathrm{P}_{X_1}\mathrm{P}_{X_2}\right) + 2\log(1/\xi), \tag{27}$$

$$L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) \leq -\log \beta_\eta\left(\mathrm{P}_{X_1X_1X_2}, \mathrm{P}_{X_1|X_2}\mathrm{P}_{X_1|X_2}\mathrm{P}_{X_2}\right) + 2\log(1/\xi), \tag{28}$$

for all $\xi > 0$ with $\eta = \varepsilon + \delta_1 + 2\delta_2 + \xi < 1$.

[3] For brevity, we use the same notation for SK capacity and OT capacity; the meaning will be clear from the context. Similarly, the notation L, used here to denote the optimal OT length, is also used to denote the optimal BC length in the next section.

Corollary 14 (Strong bound for OT capacity)**.** *For $0 < \varepsilon < 1$, the ε-OT capacity of (X_1, X_2) satisfies*

$$C_\varepsilon(X_1, X_2) \le \min\{I(X_1 \wedge X_2),\, H(X_1|X_2)\}.$$

The proof of Theorem 13 entails reducing two SK agreement problems to OT. The bound (27) is obtained by recovering K_B as a SK, while (28) is obtained by recovering $K_{\overline{B}}$ as a SK; we note these two reductions as separate lemmas below.

Lemma 15 (Reduction 1 of SK agreement to OT)**.** *Consider SK agreement for two parties observing X_1 and X_2. Given a protocol realizing an $(\varepsilon, \delta_1, \delta_2)$-OT of length l, there exists a protocol for generating an $(\varepsilon, \delta_1 + 2\delta_2)$-SK of length l. In particular,*

$$L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) \le S_{\varepsilon,\delta_1+2\delta_2}(X_1, X_2).$$

Proof sketch. Let $\hat{K}$ be the estimate of K_B formed by $\mathcal{P}_2$. The following protocol generates an $(\varepsilon, \delta_1 + 2\delta_2)$-SK of length l.

(i) $\mathcal{P}_1$ generates two random strings K_0 and K_1 of length l, and $\mathcal{P}_2$ generates a random bit B. Two parties run the OT protocol, and $\mathcal{P}_2$ obtains an estimate $\hat{K}$ of K_B.

(ii) $\mathcal{P}_2$ sends B over the public channel.

(iii) Using B, $\mathcal{P}_1$ computes K_B.

We show that K_B constitutes an $(\varepsilon, \delta_1 + 2\delta_2)$-SK. The reliability is guaranteed since both parties agree on K_B with probability greater than $1 - \varepsilon$. For establishing secrecy, note that if $\mathcal{P}_2$ sends $\overline{B}$ instead of B, the eavesdropper cannot determine K_B from $(\overline{B}, \Pi)$ by the secrecy condition for $\mathcal{P}_1$. On the other hand, by the secrecy condition for $\mathcal{P}_2$, the overall observation (K_0, K_1, X_1, Π) of $\mathcal{P}_1$ has roughly the same distribution even when B is replaced by $\overline{B}$. Thus, the eavesdropper cannot determine K_B from (B, Π) as well. ■

Lemma 16 (Reduction 2 of SK agreement to OT)**.** *Consider two party SK agreement where $\mathcal{P}_1$ observes X_1, $\mathcal{P}_2$ observes (X_1, X_2) and the eavesdropper observes X_2. Given a protocol realizing an $(\varepsilon, \delta_1, \delta_2)$-OT of length l, there exists a protocol for generating an $(\varepsilon, \delta_1 + 2\delta_2)$-SK of length l. In particular,*

$$L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) \le S_{\varepsilon,\delta_1+2\delta_2}(X_1, (X_1, X_2)|X_2).$$

Proof sketch. The following protocol generates an $(\varepsilon, \delta_1 + 2\delta_2)$-SK of length l.

(i) $\mathcal{P}_1$ generates two random strings K_0 and K_1 of length l, and $\mathcal{P}_2$ generates a random bit B. Two parties run the OT protocol.

(ii) Upon observing Π, $\mathcal{P}_2$ samples $\tilde{X}_2$ according to the distribution $\mathrm{P}_{X_2|V_1B\Pi}\left(\cdot|V_1, \overline{B}, \Pi\right)$.

(iii) $\mathcal{P}_2$ sends B over the public channel.

(iv) $\mathcal{P}_1$ computes $K_{\overline{B}}$ and $\mathcal{P}_2$ computes $\tilde{K} = \hat{K}(\tilde{X}_2, \overline{B}, \Pi)$.

We show that $K_{\overline{B}}$ constitutes an $(\varepsilon, \delta_1 + 2\delta_2)$-SK with estimate $\tilde{K}$ available to $\mathcal{P}_2$. This protocol entails $\mathcal{P}_2$ emulating $\tilde{X}_2$, pretending that the protocol was executed for $\overline{B}$ instead of B. Since the communication of $\mathcal{P}_1$ is oblivious of the value of B, plugging $\tilde{X}_2$ into $\hat{K}$ will lead to an estimate of $K_{\overline{B}}$ provided that the emulated $\tilde{X}_2$ preserves the joint distribution. ■

4.3 Converse for bit commitment

Two parties observing correlated observations X_1 and X_2 want to implement information theoretically secure *bit commitment* (BC) using interactive public communication. A BC protocol consists of two phases: the *commit phase* and the *reveal phase.* In the commit phase, $\mathcal{P}_1$ generates a random string K, distributed uniformly over $\{0,1\}^l$ and independent jointly of (X_1, X_2). Furthermore, the two parties communicate interactively using a protocol π. In the reveal phase, $\mathcal{P}_1$ "reveals" its data, *i.e.*, it sends X_1' and K', claiming these were its initial choices of X_1 and K, respectively. Subsequently, $\mathcal{P}_2$ applies a (randomized) test function $T = T(K', X_1', X_2, \mathbf{F})$, where $T = 0$ and $T = 1$, respectively, indicate $K' = K$ and $K' \neq K$.

Definition 14 (Bit commitment)**.** An $(\varepsilon, \delta_1, \delta_2)$-BC of length l consists of an interactive communication protocol π used to communicate during the commit phase and a $\{0,1\}$-valued randomized test function T to be used in the reveal phase such that the following conditions are satisfied:

$$\begin{aligned} \Pr\left(T(K, X_1, X_2, \Pi) \neq 0\right) &\leq \varepsilon, \\ d\left(\mathrm{P}_{KX_2\Pi}, \mathrm{P}_K \times \mathrm{P}_{X_2\Pi}\right) &\leq \delta_1, \\ \Pr\left(T(K', X_1', X_2, \Pi) = 0, K' \neq K\right) &\leq \delta_2, \end{aligned}$$

for any choice of random variables K' and X_1' that have the same range-sets as K and X_1, respectively, and satisfy

$$(K', X_1') — (K, X_1, \Pi) — X_2.$$

The first condition above is the *soundness condition*, which captures the reliability of BC when $\mathcal{P}_1$ is honest. The next condition is the *hiding condition*, which ensures that $\mathcal{P}_2$ cannot ascertain the secret in the commit phase. The final *binding condition* restricts the probability with which $\mathcal{P}_1$ can cheat in the reveal phase. Denote by $L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2)$ the largest l such that a protocol realizing an $(\varepsilon, \delta_1, \delta_2)$-BC of length l exists.

For n-length i.i.d. sequences X_1^n, X_2^n generated from $\mathrm{P}_{X_1X_2}$, the largest rate of $L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2)$ is called the *BC capacity.*

Definition 15 (BC capacity)**.** For $0 < \varepsilon, \delta_1, \delta_2 < 1$, the $(\varepsilon, \delta_1, \delta_2)$-BC capacity of (X_1, X_2) is defined as

$$C_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) = \liminf_{n\to\infty} \frac{1}{n} L_{\varepsilon,\delta_1,\delta_2}(X_1^n, X_2^n).$$

The characterization of $C_{\varepsilon,\delta_1,\delta_2}(X_1, X_2)$ entails the notion of *minimum sufficient statistic.*

Definition 16 (Minimum Sufficient Satistics)**.** A *sufficient statistic* for X_2 given X_1 is a random variable U which equals a function $g(X_1)$ with probability 1 and for which the Markov chain $X_1—U—X_2$ holds. The minimum sufficient statistics for X_2 given X_1, denoted by $\mathrm{mss}(X_2|X_1)$, is a sufficient statistics for X_2 given X_1 such that it is a function of every sufficient statistic U for X_2 given X_1, i.e., $H(\mathrm{mss}(X_2|X_1)|U) = 0$.

Theorem 17 (Single-shot bound for BC length)**.** *Given* $0 < \varepsilon, \delta_1, \delta_2$, $\varepsilon + \delta_1 + \delta_2 < 1$, *for random variables* X_1, X_2 *and* $V_1 = \mathrm{mss}(X_2|X_1)$, *the following inequality holds:*

$$L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) \leq -\log \beta_\eta\left(\mathrm{P}_{V_1V_1X_2}, \mathrm{P}_{V_1|X_2}\mathrm{P}_{V_1|X_2}\mathrm{P}_{X_2}\right) + 2\log(1/\xi),$$

for all ξ *with* $\eta = \varepsilon + \delta_1 + \delta_2 + \xi$.

We remark that even in (28) X_1 can be replaced by V_1.

Corollary 18 (BC capacity, with strong converse)**.** *For* $0 < \varepsilon, \delta_1, \delta_2$*,* $\varepsilon + \delta_1 + \delta_2 < 1$*, the* $(\varepsilon, \delta_1, \delta_2)$*-BC capacity is given by*

$$C_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) = H(V_1|X_2),$$

where $V_1 = \mathrm{mss}(X_2|X_1)$*.*

We omit the discussion on the scheme that achieves the capacity above.

Theorem 17 is obtained by a reduction of secret key agreement to BC; the following lemma captures the resulting bound.

Lemma 19 (Reduction of SK to BC)**.** *For* $0 < \varepsilon, \delta_1, \delta_2$*,* $\varepsilon + \delta_1 + \delta_2 < 1$*, it holds that*

$$L_{\varepsilon,\delta_1,\delta_2}(X_1, X_2) \le S_{\varepsilon+\delta_2,\delta_1}(X_1, (V_1, X_2)|X_2),$$

where $V_1 = \mathrm{mss}(X_2|X_1)$*.*

Proof sketch. Given an $(\varepsilon, \delta_1, \delta_2)$-BC of length l, consider secret key agreement by two parties observing X_1 and (V_1, X_2), respectively, with the eavesdropper observing X_2. To generate a SK, the parties run the commit phase of the BC protocol, *i.e.*, $\mathcal{P}_1$ generates $K \sim \texttt{unif}\{0,1\}^l$ and the parties send the interactive communication Π. We show that the committed secret K constitutes a $(\varepsilon + \delta_2, \delta_1)$-SK. Indeed, by the hiding condition, the secret key K satisfies the secrecy condition (14) with $\delta = \delta_1$. To establish the reliability of this secret key, we show that, roughly, K is the unique string which is compatible with (V_1, X_2, Π), namely that any other string will fail the test T, since otherwise a dishonest $\mathcal{P}_1$ can change the string in the reveal phase, contradicting the binding condition. Thus, $\mathcal{P}_2$ can obtain an estimate of K by finding the unique string that is compatible with (V_1, X_2, Π). ■

Example 4 (Reduction of BC to OT)**.** Suppose that two parties have at their disposal an OT of length n. Using this as a resource, what is the maximum length l of an $(\varepsilon, \delta_1, \delta_2)$-BC that can be constructed?

Denoting by K_0, K_1 the OT strings, and by B the OT bit of $\mathcal{P}_2$, let $X_1 = (K_0, K_1)$ and $X_2 = (B, K_B)$. Note that

$$D(\mathrm{P}_{X_1X_1X_2} \| \mathrm{P}_{X_1|X_2}\mathrm{P}_{X_1X_2}) = n.$$

Therefore, by Theorem 17, we get

$$l \le n + \log(1/(1 - \varepsilon - \delta_1 - \delta_2 - \eta)) + 2\log(1/\eta),$$

where $0 < \eta < 1 - \varepsilon - \delta_1 - \delta_2$.

5 Interactive secret key agreement

We now move on to interactive protocols for secret key agreement.

5.1 Second-order rate of secret key agreement

When (X, Y, Z) is a degraded source, i.e., $X \multimap Y \multimap Z$, the secrecy capacity is given by

$$C_{\varepsilon,\delta}(X, Y|Z) = I(X \wedge Y|Z),$$

for $\varepsilon + \delta < 1$. We noted that a simple protocol where each party communicates only once, based only on its local observation, attains $C_{\varepsilon,\delta}(X, Y|Z)$. However, this only provides the first order asymptotic optimality. To attain rates which are optimal even up to the second order asymptotic term, we propose a more sophisticated interactive protocol.

Theorem 20. *For $0 < \varepsilon, \delta < 1$ with $0 < \varepsilon + \delta < 1$, it holds that*

$$S_{\varepsilon,\delta}(X^n, Y^n|Z^n) = nI(X \wedge Y|Z) - \sqrt{nV}Q^{-1}(\varepsilon + \delta) + \mathcal{O}(\log n),$$

where

$$V = \mathrm{Var}\left[\log \frac{\mathrm{P}_{XY|Z}(X, Y|Z)}{\mathrm{P}_{X|Z}(X|Z)\,\mathrm{P}_{Y|Z}(Y|Z)}\right].$$

The converse proof of Theorem 20 follows from Theorem 8 by the central limit theoerm. The achievability part uses an interactive protocol, which we describe below. But before we do that, in the next section we examine if a simple non-interactive protocol can attain this bound.

5.2 Why doesn't a simple protocol work?

Recall that in the information reconciliation step of the simple protocol of Section 2.2, $\mathcal{P}_1$ sends $(F, F(X))$ to $\mathcal{P}_2$ observing y. Then, the receiver looks for a unique $\hat{x}$ in the list $\mathcal{L}_y$ comprising x such that (x, y) belong to the typical set

$$\mathcal{T}_{\mathrm{P}_{X|Y}} = \{(x, y) : h_{\mathrm{P}_{X|Y}}(x|y) \le t - \gamma\},$$

compatible with the hash value received. Here $h_{\mathrm{P}_{X|Y}}(x|y) = -\log \mathrm{P}_{X|Y}(x|y)$ is the conditional entropy density.

The error probability of this simple protocol is bounded as

$$\Pr\left(X \neq \hat{X}\right) \le \mathrm{P}_{XY}\left(\mathcal{T}^c_{\mathrm{P}_{X|Y}}\right) + 2^{-\gamma}.$$

Essentially, the result above says that Party 1 can send X to Party 2 with probability of error less than ε using roughly as many bits as the ε-tail of $h_{\mathrm{P}_{X|Y}}(X|Y)$, namely the infimum over t such that $\mathrm{P}_{XY}\left(h_{\mathrm{P}_{X|Y}}(X|Y) > t\right)$ is less than ε.

Since the bits revealed in the information reconciliation phase must be subtracted in the privacy amplification phase, the length of SK generated is roughly

$$\left[\delta\text{-tail of } h_{\mathrm{P}_{X|Z}}(X|Z)\right] - \left[\varepsilon\text{-tail of } h_{\mathrm{P}_{X|Y}}(X|Y)\right].$$

A drawback of the above scheme is that $\mathcal{P}_1$ always sends t bits even if, for the observed realization (x, y) of (X, Y), $h_{\mathrm{P}_{X|Y}}(x|y)$ is much smaller than the ε-tail of $h_{\mathrm{P}_{X|Y}}(X|Y)$. We show below that there exists a secret key agreement protocol such that the length of the secret key is given by roughly the

$$(\varepsilon + \delta)\text{-tail of } \left[h_{\mathrm{P}_{X|Z}}(X|Z) - h_{\mathrm{P}_{X|Y}}(X|Y)\right].$$

5.3 Interactive Slepian-Wolf coding

We rely on a "spectrum slicing" technique. Our protocol focuses on the "essential spectrum" of $h_{\mathrm{P}_{X|Y}}(X|Y)$, i.e., those values of (X, Y) for which $h_{\mathrm{P}_{X|Y}}(X|Y) \in (\lambda_{\min}, \lambda_{\max})$. For $\lambda_{\min}, \lambda_{\max}, \Delta > 0$ with $\lambda_{\max} > \lambda_{\min}$, let

$$N = \frac{\lambda_{\max} - \lambda_{\min}}{\Delta}, \tag{29}$$

and

$$\lambda_i = \lambda_{\min} + (i-1)\Delta, \quad 1 \le i \le N. \tag{30}$$

Further, let

$$\mathcal{T}_0 = \Big\{(x, y) : h_{\mathrm{P}_{X|Y}}(x|y) \ge \lambda_{\max} \text{ or } h_{\mathrm{P}_{X|Y}}(x|y) < \lambda_{\min}\Big\}, \tag{31}$$

and for $1 \le i \le N$, let $\mathcal{T}_i$ denote the ith slice of the spectrum given by

$$\mathcal{T}_i = \Big\{(x, y) : \lambda_i \le h_{\mathrm{P}_{X|Y}}(x|y) < \lambda_i + \Delta\Big\}. \tag{32}$$

Note that $\mathcal{T}_0$ corresponds to the complement of the "typical set." Finally, let $\mathcal{H}_l(\mathcal{X})$ denote the set of all mappings $h : \mathcal{X} \to \{0, 1\}^l$.

Our protocol for transmitting X to an observer of Y is described in Protocol 1. The lemma below bounds the probability of error for Protocol 1 when $(x, y) \in \mathcal{T}_i$, $1 \le i \le N$.

Lemma 21 (Performance of Protocol 1). *For $(x, y) \in \mathcal{T}_i$, $1 \le i \le N$, denoting by $\hat{X} = \hat{X}(x, y)$ the estimate of x at Party 2 at the end of the protocol (with the convention that $\hat{X} = \emptyset$ if an error is declared), Protocol 1 sends at most $(l + (i-1)\Delta + i)$ bits and has probability of error bounded above as follows:*

$$\Pr\Big(\hat{X} \ne x \mid X = x, Y = y\Big) \le i 2^{\lambda_{\min} + \Delta - l}.$$

Proof. Since $(x, y) \in \mathcal{T}_i$, an error occurs if there exists a $\hat{x} \ne x$ such that $(\hat{x}, y) \in \mathcal{T}_j$ and $\Pi_{2k-1} = h_{2k-1}(\hat{x})$ for $1 \le k \le j$ for some $j \le i$. Therefore, the probability of error is bounded above as

$$\begin{aligned}
&\Pr\Big(\hat{X} \ne x \mid X = x, Y = y\Big) \\
&\le \sum_{j=1}^{i} \sum_{\hat{x} \ne x} \Pr\left(h_{2k-1}(x) = h_{2k-1}(\hat{x}), \, \forall\, 1 \le k \le j\right) \mathbb{1}\big((\hat{x}, y) \in \mathcal{T}_j\big) \\
&\le \sum_{j=1}^{i} \sum_{\hat{x} \ne x} \frac{1}{2^{l + (j-1)\Delta}} \mathbb{1}\big((\hat{x}, y) \in \mathcal{T}_j\big) \\
&= \sum_{j=1}^{i} \sum_{\hat{x} \ne x} \frac{1}{2^{l + (j-1)\Delta}} |\{\hat{x} \mid (\hat{x}, y) \in \mathcal{T}_j\}| \\
&\le i 2^{\lambda_{\min} - l + \Delta},
\end{aligned}$$

where we have used the fact that $\log |\{\hat{x} \mid (\hat{x}, y) \in \mathcal{T}_j\}| \le \lambda_j + \Delta$. Note that the protocol sends l bits in the first transmission, and Δ bits and 1-bit feedback in every subsequent transmission. Therefore, no more than $(l + (i-1)\Delta + i)$ bits are sent. ■

Protocol 1: Interactive Slepian-Wolf compression

```
Input: Observations $X$ and $Y$, uniform public randomness $U$, and parameters $l, \Delta$
Output: Estimate $\hat{X}$ of $X$ at party 2
Both parties use $U$ to select $h_1$ uniformly from $\mathcal{H}_l(\mathcal{X})$
Party 1 sends $\Pi_1 = h_1(X)$
if Party 2 finds a unique $x \in \mathcal{T}_1$ with hash value $h_1(x) = \Pi_1$ then
    set $\hat{X} = x$
    send back $\Pi_2 = $ ACK
else
    send back $\Pi_2 = $ NACK
while $2 \le i \le N$ and party 2 did not send an ACK do
    Both parties use $U$ to select $h_i$ uniformly from $\mathcal{H}_\Delta(\mathcal{X})$, independent of $h_1, ..., h_{i-1}$
    Party 1 sends $\Pi_{2i-1} = h_i(X)$
    if Party 2 finds a unique $x \in \mathcal{T}_i$ with hash value $h_j(x) = \Pi_{2j-1}, \forall 1 \le j \le i$ then
        set $\hat{X} = x$
        send back $\Pi_{2i} = $ ACK
    else
        if More than one such $x$ found then
            protocol declares an error
        else
            send back $\Pi_{2i} = $ NACK
    Reset $i \to i + 1$
if No $\hat{X}$ found at party 2 then
    Protocol declares an error
```

Corollary 22 (Interactive Slepian-Wolf). *Protocol 1 with $l = \lambda_{\min} + \Delta + \eta$ sends at most $(h_{\mathrm{P}_{X|Y}}(X|Y) + \Delta + N + \eta)$ bits when the observations are $(X, Y) \notin \mathcal{T}_0$ and has probability of error less than*

$$\Pr\left(\hat{X} \neq X\right) \le \mathrm{P}_{XY}\left(\mathcal{T}_0\right) + N2^{-\eta}.$$

If we choose $\mathcal{T}_0$ appropriately, then we can make $\Pr\left(\hat{X} \neq X\right) \simeq 0$.

5.4 Interactive secret key agreement scheme

We construct a secret key agreement protocol generating (ε, δ)-SK for

$$\varepsilon \simeq 0,$$
$$\delta \lesssim \Pr\left(h_{\mathrm{P}_{X|Z}}(X|Z) - h_{\mathrm{P}_{X|Y}}(X|Y) \le \log|\mathcal{K}|\right).$$

From this protocol, by a coupling argument, we can construct a (ε, δ)-SK for any (ε, δ) satisfying

$$\varepsilon + \delta \le \Pr\left(h_{\mathrm{P}_{X|Z}}(X|Z) - h_{\mathrm{P}_{X|Y}}(X|Y) \le \log|\mathcal{K}|\right).$$

The protocol is described in Protocol 2.

```
Protocol 2: Secret key agreement protocol
  Input: Observations $X$ and $Y$
  Output: Secret key estimates $K_1$ and $K_2$
  Information reconciliation
  Use interactive Slepian-Wolf Coding
  if No ACK received then
      Protocol declares an error and aborts
  else
      Privacy amplification
      First party generates the random seed $S$ and sends it to the second party using public
      communication
      First party generates the secret key $K_1 = K = f_S(X)$
      The second party generates the estimate $K_2$ of $K$ as $K_2 = f_S(\hat{X})$
```

Outline of Security Analysis Let

$$\mathcal{E} = \big\{(x, y, z) : h_{\mathrm{P}_{X|Z}}(x|z) - h_{\mathrm{P}_{X|Y}}(x|y) \le \lambda + \Delta\big\},$$

and

$$J = \begin{cases} 0 & \text{if } (X,Y) \in \mathcal{T}_0 \text{ or } (X,Y,Z) \in \mathcal{E} \\ j & \text{if} (X,Y) \in \mathcal{T}_j \text{ and } (X,Y,Z) \in \mathcal{E}^c \end{cases}.$$

Then,

$$\begin{aligned}
& d(\mathrm{P}_{KZ\Pi S}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi S}) \\
& \le d(\mathrm{P}_{KZ\Pi SJ}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi SJ}) \\
& \le \mathrm{P}_{XYZ}\left(\mathcal{E}\right) + \mathrm{P}_{XY}\left(\mathcal{T}_0\right) + \sum_{j=1}^{N} d(\mathrm{P}_{KZ\Pi S|J=j}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi S|J=j}).
\end{aligned}$$

Conditioned on $J = j$, it can be shown that

- $\log \|\Pi\| \le \lambda_j + \Delta + \eta + \log N$,
- $H_{\min}(\mathrm{P}_{XZ|J=j}|\mathrm{P}_Z) \ge \lambda_j + \lambda + \Delta - 2\log N$.

Thus, by the leftover hash lemma,

$$d(\mathrm{P}_{KZ\Pi S|J=j}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{Z\Pi S|J=j}) \le \frac{1}{2}\sqrt{|\mathcal{K}|2^{-(\lambda-\eta-3\log N)}}$$

for each j. It follows that, when $J = j$,

$$\begin{aligned}
\log \|\Pi\| &\lesssim h_{\mathrm{P}_{X|Y}}(X|Y), \\
H_{\min}(\mathrm{P}_{XZ|J=j}|\mathrm{P}_Z) &\gtrsim \Big[\delta\text{-tail of } \big[h_{\mathrm{P}_{X|Z}}(X|Z) - h_{\mathrm{P}_{X|Y}}(X|Y)\big]\Big] + h_{\mathrm{P}_{X|Y}}(X|Y),
\end{aligned}$$

which enables us to take

$$\log |\mathcal{K}| \simeq \Big[\delta\text{-tail of } \big[h_{\mathrm{P}_{X|Z}}(X|Z) - h_{\mathrm{P}_{X|Y}}(X|Y)\big]\Big].$$

6 Multiparty and universal schemes

6.1 Problem description

In the multiparty version of the secret key agreement problem, m parties $\mathcal{P}_1, ..., \mathcal{P}_m$ observe, respectively, random variables $X_1, ..., X_m$ and seek to generate shared random bits which are almost independent of the communication used to generate them. Specifically, a $\mathcal{K}$-valued random variable K constitutes an (ε, δ)-SK of length $\log |\mathcal{K}|$ if there exist local randomness U_i available to $\mathcal{P}_i$, an interactive communication protocol π , and estimates $K_i = K_i(X_i^n, U_i, \Pi)$ such that the secrecy condition (14) holds together with the following recovery condition

$$\Pr\left(K_i = K, \forall\, i \in \mathcal{M}\right) \geq 1 - \varepsilon.$$

Denote by $S_{\varepsilon,\delta}(X_1, ..., X_m)$ the maximum length of an (ε, δ)-SK.

Generating such a secret key, as before, entails an information reconciliation and a privacy amplification step. We consider protocols where the information reconciliation entails each party recovering the data of every other party, namely, each party recovers $(X_1, ..., X_m)$ and the parties attain *omniscience.* Then, if t bits were communitated for attaining omniscience, a secret key of length roughly $H_{\min}^{\eta}(\mathrm{P}_{X_1,...,X_m}) - t$ can be extracted using the privacy amplification as before.

In fact, this scheme is asymptotically optimal. Specifically, for an n-length i.i.d. sequence $\{(X_{1i}, ..., X_{mi})\}_{i=1}^n$, the (ε, δ)-SK capacity is defined as

$$C_{\varepsilon,\delta}(X_1, ..., X_m) = \liminf_{n\to\infty} \frac{1}{n} S_{\varepsilon,\delta}(X_1^n, ..., X_m^n).$$

Theorem 23 (Multiparty SK capacity, with strong converse)**.** *For random variables $X_1, ..., X_m$ and $0 \leq \varepsilon, \delta$ such that $\varepsilon + \delta < 1$,*

$$C_{\varepsilon,\delta}\left(\mathrm{P}_{X_1,...,X_m}\right) = \min_{\sigma\in\Sigma(\mathcal{M})} \frac{1}{|\sigma| - 1} D\left(\mathrm{P}_{X_\mathcal{M}} \| \prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\right),$$

where $\Sigma(\mathcal{M})$ denotes the set of nontrivial partitions of $\mathcal{M}$.

The achievability part of the proof is as outlined above and the converse uses a multiparty form on the conditional independence testing bound. An important feature of the result above is that while interaction was allowed in the model, simple communication involving transmissions depending only on the local data of the parties can attain it.

6.2 A universal scheme and necessity of interaction

A universal secret key agreement protocol needs to operate without the knowledge of the distribution of the data, yet we hope for the performance which is the same as of that of the optimal protocol given the joint distribution $\mathrm{P}_{X_\mathcal{M}} = \mathrm{P}_{X_1,...,X_m}$, i.e., we seek to attain rates approaching

$$\begin{aligned} C(\mathrm{P}_{X_\mathcal{M}}) &:= \lim_{\varepsilon,\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} S_{\varepsilon,\delta}(X_1^n, ..., X_m^n) \\ &= \min_{\sigma\in\Sigma(\mathcal{M})} \frac{1}{|\sigma| - 1} D\left(\mathrm{P}_{X_\mathcal{M}} \| \prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\right). \end{aligned}$$

Such universally optimal protocols can only be interactive, as otherwise the communication sent in the information reconciliation phase may be too high to yield any positive rate secret key. The main result we shall cover is the following.

Theorem 24. *For $\Delta = \frac{1}{\sqrt{n}}$, $0 < \delta < 1$, and every distribution $\mathrm{P}_{X_{\mathcal{M}}}$, there is a universal protocol that generates a variable length (ε_n, δ)-SK with ε_n vanishing to 0 as $n \to \infty$ and average key length greater than*

$$nC(\mathrm{P}_{X_{\mathcal{M}}}) - \mathcal{O}(\sqrt{n \log n}). \tag{33}$$

We outline the universally rate optimal protocol for attaining omniscience, a universal *data exchange protocol*, that is used in the information reconciliation step of the promised universal secret key agreement protocol.

We begin with a formal description of the problem for i.i.d. observations. Specifically, parties in a set $\mathcal{M} = \{1, \ldots, m\}$ observe an i.i.d. sequence $X_{\mathcal{M}}^n = (X_{\mathcal{M}1}, \ldots, X_{\mathcal{M}n})$, with the ith party observing $\{X_{it}\}_{t=1}^n$ and $X_{\mathcal{M}t} = (X_{it} : i \in \mathcal{M}) \sim \mathrm{P}_{X_{\mathcal{M}}}$ denoting the collective data at the tth time instance. The parties have access to shared public randomness (public coins) U such that U is independent jointly of $X_{\mathcal{M}}^n$. Furthermore, the ith party, $i \in \mathcal{M}$, has access to private randomness (private coins) U_i such that $U_{\mathcal{M}}$, U, and $X_{\mathcal{M}}^n$ are mutually independent. Thus, the ith party observes (X_i^n, U_i, U).

A tree-protocol π for $\mathcal{M}$ consists of a binary tree, termed the *protocol-tree*, with the vertices labeled by the elements of $\mathcal{M}$. The protocol starts at the root and proceeds towards the leaves. When the protocol is at vertex v with label i_v, party i_v communicates a bit b_v based on its local observations $(X_{i_v}^n, U_{i_v}, U)$. The protocol proceeds to the left- or the right-child of v, respectively, if b_v is 0 or 1. The protocol terminates when it reaches a leaf, at which point each party produces an output based on its local observations and the bits communicated during the protocol, namely the transcript $\Pi = \pi(X_{\mathcal{M}}^n, U_{\mathcal{M}}, U)$.

The (worst-case) length $|\pi|$ of a protocol π is the maximum number of bits that are transmitted in any execution of the protocol and equals the depth of the protocol-tree.

Definition 17. A protocol π constitutes an ε-omniscience protocol if, at the end of the protocol, the ith party can output an estimate $\widehat{\mathbf{X}}_i = \widehat{\mathbf{X}}_i(X_i^n, U_i, U, \Pi) \in \mathcal{X}_{\mathcal{M}}^n$ such that

$$\Pr\left(\widehat{\mathbf{X}}_i = X_{\mathcal{M}}^n : i \in \mathcal{M}\right) \geq 1 - \varepsilon.$$

Definition 18 (Communication for omniscience)**.** Given IID observations with a common distribution $\mathrm{P}_{X_{\mathcal{M}}}$ as above, for $0 \leq \varepsilon < 1$, a rate $R \geq 0$ is an ε-achievable omniscience rate if there exists an ε-omniscience protocol π with length $|\pi|$ less than nR, for all n sufficiently large. The infimum over all ε-achievable omniscience rates is denoted by $R_\varepsilon(\mathrm{P}_{X_{\mathcal{M}}})$. The *minimum rate of communication for omniscience* $R(\mathrm{P}_{X_{\mathcal{M}}})$ is given by

$$R(\mathrm{P}_{X_{\mathcal{M}}}) = \lim_{\varepsilon \to 0} R_\varepsilon(\mathrm{P}_{X_{\mathcal{M}}}).$$

Theorem 25 (Minimum communication for omniscience)**.** *For a joint distribution $\mathrm{P}_{X_{\mathcal{M}}}$,*

$$R(\mathrm{P}_{X_{\mathcal{M}}}) = \min\left\{\sum_{i=1}^m R_i : \sum_{i \in B} R_i \geq H(X_B | X_{B^c}), \quad \forall\, B \subsetneq \mathcal{M}\right\}. \tag{34}$$

The collection of all rate vectors $\mathbf{R} = (R_1, \ldots, R_m)$ satisfying the constraints on the right-side of (34), termed the CO region, will be denoted by $\mathcal{R}_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$, and the minimum sum-rate by $R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$. Using duality of linear programming,

$$\min\left\{\sum_{i=1}^m R_i : \sum_{i \in B} R_i \geq H(X_B | X_{B^c}), \quad \forall\, B \subsetneq \mathcal{M}\right\} = \max_{\sigma \in \Sigma(\mathcal{M})} \frac{1}{|\sigma| - 1} \sum_{i=1}^{|\sigma|} H(X_{\mathcal{M}} | X_{\sigma_i}). \tag{35}$$

Thus,

$$\min_{\sigma \in \Sigma(\mathcal{M})} \frac{1}{|\sigma|-1} D\left(\mathrm{P}_{X_{\mathcal{M}}} \| \prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}} \right) = H(X_{\mathcal{M}}) - \mathcal{R}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right),$$

which shows the connection between secret key agreement and data exchange.

We give a universal protocol for omniscience, which, when a sequence $\mathbf{x}_{\mathcal{M}}$ is observed, will transmit communication of rate no more than $R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}\right)$, where $\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}$ denotes the joint type of $\mathbf{x}_{\mathcal{M}}$. The protocol directly achieves the right-side of (35). We present the protocol under ideal conditions.

(a) *Continuous rate assumption:* Communication-rate, defined as the total number of bits of communication up to a certain time divided by n, can be increased continuously in time; and

(b) *Ideal decoder assumption:* We assume the availability of an error-free, ideal decoder $\mathrm{DEC}_{\mathtt{id}}$ which correctly decodes a sequence once sufficient communication has been sent and declares a `NACK` otherwise.

Protocol 3 summarizes the ideal decoder $\mathrm{DEC}_{\mathtt{id}}$ we use. Note that a random hash denotes the output of a 2-UHF. With this ideal decoder at our disposal, under the continuous rates assumption, finding a universal protocol is tantamount to finding a policy for increasing the rates $(R_1, \ldots, R_m)$ such that when the rate vector enters $\mathcal{R}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}\right)$ for the first time, the sum-rate is $R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}\right)$, where

```
Protocol 3: Ideal decoder $\mathrm{DEC}_{\mathtt{id}}(j, \sigma, \mathbf{R})$
  Input: An index $1 \le j \le m$, a partition $\sigma \in \Sigma(\mathcal{M})$, a rate vector $\mathbf{R} = (R_1, \ldots, R_m)$.
  Output: An ACK message $(\mathtt{ACK}, A)$ or a NACK message
    1. For $\sigma_i$ such that $j \in \sigma_i$, search for the maximal set $A \subseteq \mathcal{M}$ such that $\sigma_i \subsetneq A$ and
       $(R_l : l \in A) \in \mathcal{R}_{\mathtt{CO}}(A \mid \mathrm{P}_{\mathbf{x}_A})$,
       and reveal $\mathbf{x}_A$ to party $j$.
    2. if If such an A was found in Step 1 then
         return $(\mathtt{ACK}, A)$.
       else
         return NACK.
```

$$\mathcal{R}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) := \left\{ (R_1, \ldots, R_m) : \sum_{i \in B} R_i \ge H(X_B|X_{B^c}), \quad \forall\, B \subsetneq \mathcal{M} \right\}$$

is the omniscience region for a given joint distribution $\mathrm{P}_{X_{\mathcal{M}}}$. Note that initially the marginal types $\mathrm{P}_{\mathbf{x}_i}$ are available to each party and can be transmitted using $\mathcal{O}(\log n)$ bits, since there are only polynomially many types. Also, if a subset A attains local omniscience during the execution of the protocol, any $j \in A$ upon recovering $\mathbf{x}_A$ can transmit $\mathrm{P}_{\mathbf{x}_A}$ in $\mathcal{O}(\log n)$ bits to all the parties, who in turn can use it to compute $H(\mathrm{P}_{\mathbf{x}_A})$.

As an illustration, consider the simple case when $m = 2$. Parties first share $\mathrm{P}_{\mathbf{x}_1}$ and $\mathrm{P}_{\mathbf{x}_2}$; suppose $H(\mathrm{P}_{\mathbf{x}_1}) \ge H(\mathrm{P}_{\mathbf{x}_2})$. Then, party 1 starts communicating and increases its rate R_1 at slope 1. When the rate R_1 reaches $H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_2})$, party 2 starts communicating at slope 1 as well.

Throughout the protocol, each party is trying to decode the other using the ideal decoder $\mathrm{DEC}_{\mathtt{id}}$ and they keep on communicating as long as the ideal decoders output NACKs. The parties will decode each other as soon as (R_1, R_2) enters $\mathcal{R}_{\mathtt{CO}}(\{1,2\}|\mathrm{P}_{\mathbf{x}_1,\mathbf{x}_2})$, *i.e.*, when

$$R_1 \geq H(\overline{X}_1|\overline{X}_2) \text{ and } R_2 \geq H(\overline{X}_2|\overline{X}_1),$$

where $(\overline{X}_1, \overline{X}_2) \sim \mathrm{P}_{\mathbf{x}_1,\mathbf{x}_2}$. Note that once both parties start communicating, the difference $R_1 - R_2$ is maintained as $H(\overline{X}_1) - H(\overline{X}_2)$. Thus, when (R_1, R_2) enters $\mathcal{R}_{\mathtt{CO}}(\{1,2\})$, it holds that

$$R_1 = H(\overline{X}_1|\overline{X}_2) \text{ and } R_2 = H(\overline{X}_2|\overline{X}_1).$$

RDE extends the idea above to a general m. We design RDE so that the first subset A which attains local omniscience does so by using communication only from the parties in A and of sum rate

$$R_A = \mathbb{H}_{\sigma_f(A)}(A|\mathrm{P}_{\mathbf{x}_A}) = \sum_{i \in A} R_i^*(A); \tag{36}$$

the second equality requires a proof. It can also be seen that for every A

$$R_i^*(A) - R_j^*(A) = H(\overline{X}_i) - H(\overline{X}_j). \tag{37}$$

A key point here is that for $\mathrm{P}_{\mathbf{x}_\mathcal{M}}$ this difference can be computed using only the marginal types $\mathrm{P}_{\mathbf{x}_i}$ and $\mathrm{P}_{\mathbf{x}_j}$. RDE ensures that for every pair (i, j) of communicating parties, the rate of communication

$$R_i - R_i^*(A) = R_j - R_j^*(A),$$

which by (37) in turn can be ensured if the *constant difference* property, namely

$$R_i - R_j = H(\overline{X}_i) - H(\overline{X}_j), \tag{38}$$

is maintained throughout the protocol for every pair of communicating parties. Thus, all communicating parties i reach the rate $R_i^*(A)$ at the same time. Specifically, we first arrange parties in decreasing order of the entropy of the empirical distribution of their local observations, which are shared in $\mathcal{O}(\log n)$-bits. Assuming $H(\mathrm{P}_{\mathbf{x}_1}) \geq H(\mathrm{P}_{\mathbf{x}_2}) \geq \cdots \geq H(\mathrm{P}_{\mathbf{x}_m})$, party 1 starts communicating, and the ith party starts communicating when $R_1 \geq H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_i})$. This ensures the constant difference property (38) for every pair (i, j) of communicating parties. For notational convenience, we assign -1 to R_i when the ith party has not started communicating; the rate vector $(0, -1, -1, \ldots, -1)$ indicates that party 1 starts communicating and every one else remains quiet. When a subset A attains local omniscience, we decrease the rate-slope for each party $i \in A$ to $1/|A|$, thereby ensuring that collectively parties in A increase the rate of communication R_A at slope 1. Note that since parties in A have recovered $\mathbf{x}_A$, any one party $i \in A$ can compute the type $\mathrm{P}_{\mathbf{x}_A}$ and transmit it using $\mathcal{O}(\log n)$ bits. Our main observation is that at this point the rates appear as if the parties in A were collocated to begin with and have been executing the protocol as a single party. In particular, $R_A - R_j = H(\overline{X}_A) - H(\overline{X}_j)$ for any communicating party j outside A. The second crucial observation is that for the first subset A which attains local omniscience, $(R_i^*(A) : i \in A) \in \mathcal{R}_{\mathtt{CO}}(A)$. Since by (36) $\sum_{i \in A} R_i^*(A)$ is a lower bound for $\mathcal{R}_{\mathtt{CO}}(A)$, the parties in A cannot attain local omniscience before they communicate at sum-rate $\sum_{i \in A} R_i^*(A)$. Further, RDE ensures that all parties in A reach the rate $R_i^*(A)$ at the same time. Thus, the parties in A must have communicated at sum-rate

$$R_A = \sum_{i \in A} R_i^*(A) = \mathbb{H}_{\sigma_f(A)}(A|\mathrm{P}_{\mathbf{x}_A}) \tag{39}$$

Protocol 4: $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R})$

```
Input: A partition $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$, an entropy estimate vector
       $\mathbf{H} = (H_{\sigma_i} : 1 \le i \le k)$, a rate vector $\mathbf{R} = (R_1, \ldots, R_m)$; we assume that $\mathbf{H}$ is sorted,
       i.e., $H_{\sigma_1} \ge H_{\sigma_2} \ge \cdots \ge H_{\sigma_k}$.
Output: A rate vector $\mathbf{R}^{\mathtt{out}}$, a family of subsets $\mathcal{O}$ that have attained omniscience.
  1. Initialize $s := \max\{i : R_{\sigma_i} \ge 0\}$.
  2. All parties $j$ such that $j \in \sigma_i$ for some $1 \le i \le s$ increase their rates $R_j$ at slope $1/|\sigma_i|$.
  3. if There exists $i > s$ such that $R_{\sigma_1} \ge H_{\sigma_1} - H_{\sigma_i}$ then
       set $R_j = 0$ for all $j \in \sigma_i$, and set $s = \max\{i : R_{\sigma_i} \ge 0\}$.
  4. For all $j$ such that $j \in \sigma_i$ for some $1 \le i \le s$, execute $\mathrm{DEC}_{\mathtt{id}}(j, \sigma, \mathbf{R})$, which outputs NACK or
     $(\mathtt{ACK}, A_j)$.
  5. if All parties send a NACK then
       return to Step 2.
     else
       Identify the omniscience family
           $\mathcal{O} = \{B \subset \mathcal{M} : \text{ all } j \in B \text{ returned } (\mathtt{ACK}, B)\}$.
       Set $\mathbf{R}^{\mathtt{out}} = \mathbf{R}$ and return $(\mathbf{R}, \mathcal{O})$.
```

when they attain local omniscience. As the protocol proceeds, subsets of parties keep attaining local omniscience and start behaving as a single party. Proceeding recursively, it follows that when all parties attain omniscience, the rate of communication must equal $\mathbb{H}_\sigma(\mathcal{M}|\mathrm{P}_{\mathbf{X}_\mathcal{M}})$ for some $\sigma \in \Sigma(\mathcal{M})$, which is no more than $R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{\mathbf{X}_\mathcal{M}})$ and must be optimal in the limit as $n \to \infty$.

We describe the one-step omniscience protocol $\mathrm{OMN}_{\mathtt{id}}$ in Protocol 4. The protocol takes as input a partition σ such that parties in any one part are behaving as collocated parties, a vector $\mathbf{H} = (H_{\sigma_i}, 1 \le i \le |\sigma|)$ consisting of estimates of entropy for marginal distribution of parties in any part of σ, and a rate vector $\mathbf{R} = (R_1, \ldots, R_m)$ of rates of communication sent by all the parties up to this point.

Definition 19. For $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$ and $\mathbf{H} = (H_{\sigma_1}, \ldots, H_{\sigma_k})$ with $H_{\sigma_1} \ge H_{\sigma_2} \ge \cdots \ge H_{\sigma_k}$, a rate vector $(R_1, \ldots, R_m)$ is $(\sigma, \mathbf{H})$-valid if

$$(R_j, j \in \sigma_i) \in \mathcal{R}_{\mathtt{CO}}(\sigma_i), \quad \forall i \text{ s.t. } |\sigma_i| \ge 2,$$

and $(R_{\sigma_i}, 1 \le i \le k)$ can be obtained by starting with $(0, -1, -1, \ldots, -1)$ and incrementing the rates as in Protocol 4 when the parties in each part σ_i are collocated, *i.e.*, each part σ_i starts increasing its rate at slope 1 once $R_{\sigma_1} \ge H_{\sigma_1} - H_{\sigma_i}$.

The result below shows a recursive property of $\mathrm{OMN}_{\mathtt{id}}$ that renders RDE universally rate-optimal. Specifically, it shows that if $\mathbf{R}$ is $(\sigma, \mathbf{H})$-valid then, when $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R})$ terminates, the output rate vector is $(\sigma^{\mathtt{out}}, \mathbf{H}^{\mathtt{out}})$-valid where $\sigma^{\mathtt{out}}$ is a sub-partition of σ which is obtained by combining the parts that have achieved local omniscience; $\mathbf{H}^{\mathtt{out}}$ is the corresponding estimate for entropies of the marginals of parts of $\sigma^{\mathtt{out}}$. Furthermore, for every set A that attains local omniscience, the sum-rate R_A at the end of $\mathrm{OMN}_{\mathtt{id}}$ is exactly $\mathbb{H}_{\sigma_f(A_\sigma)}(A_\sigma)$.

Theorem 26. *For $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$ and $\mathbf{H} = (H_{\sigma_1}, \ldots, H_{\sigma_k})$ with $H_{\sigma_1} \geq H_{\sigma_2} \geq \cdots H_{\sigma_k}$, let $\mathbf{R}^{\mathtt{in}} = (R_1^{\mathtt{in}}, \ldots, R_m^{\mathtt{in}})$ be $(\sigma, \mathbf{H})$-valid. Then, if $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R}^{\mathtt{in}})$ is executed, the final rates $\mathbf{R}^{\mathtt{out}}$ and the omniscience family $\mathcal{O}$ satisfy the following:*

1) Every $A \in \mathcal{O}$ consists of parts of σ, i.e.,

$$A = \bigcup_{l=1}^{c} \sigma_{i_l}$$

for some $\{i_1, \ldots, i_c\} \subseteq \{1, \ldots, |\sigma|\}$, and the sum-rate $R_A^{\mathtt{out}}$ satisfies

$$R_A^{\mathtt{out}} = \mathbb{H}_{\{\sigma_{i_1}|\cdots|\sigma_{i_c}\}}\left(A|\mathrm{P}_{\mathbf{x}_A}\right).$$

2) Let $\sigma^{\mathtt{out}} \in \Sigma(\mathcal{M})$ be the partition obtained by combining the parts in σ that belong to the same A in $\mathcal{O}$. Let $H_{\sigma_i^{\mathtt{out}}}$ denote the entropy of the type of $\mathbf{x}_{\sigma_i^{\mathtt{out}}}$. Then, with $\mathbf{H}^{\mathtt{out}} = \left(H_{\sigma_i^{\mathtt{out}}}, 1 \leq i \leq |\sigma^{\mathtt{out}}|\right)$, $\mathbf{R}^{\mathtt{out}}$ is $(\sigma^{\mathtt{out}}, \mathbf{H}^{\mathtt{out}})$-valid.

Thus, if we proceed by recursively calling $\mathrm{OMN}_{\mathtt{id}}$, each time with $(\sigma^{\mathtt{out}}, \mathbf{H}^{\mathtt{out}}, \mathbf{R}^{\mathtt{out}})$ obtained from the previous call, we shall ultimately attain omniscience using the sum-rate $\mathbb{H}_{\sigma}(\mathcal{M})$ for some partition σ. Since $\mathbb{H}_{\sigma}(\mathcal{M})$ is a lower bound for $\mathcal{R}_{\mathtt{CO}}(\mathcal{M})$ by (35), this rate must be optimal. We summarize the overall ideal protocol in Protocol 5.

Protocol 5: $\mathrm{RDE}_{\mathtt{id}}$: The recursive data exchange protocol under ideal conditions

```
1. Initialize $\sigma = \sigma_f(\mathcal{M})$, $\mathbf{R} = (0, -1, -1, \ldots, -1)$, $k = |\sigma|$.
2. while $k > 1$ do
    (i) For $1 \leq i \leq k$, a party $j \in \sigma_i$ computes $\mathrm{P}_{\mathbf{x}_{\sigma_i}}$
        and broadcasts it. Each party computes $H_{\sigma_i} = H\left(\mathrm{P}_{\mathbf{x}_{\sigma_i}}\right)$, $1 \leq i \leq k$.
    (ii) Let $\mathbf{H}$ be the sorted version of
        $(H_{\sigma_i} : 1 \leq i \leq k)$, i.e., assume $H_{\sigma_1} \geq H_{\sigma_2} \geq \cdots \geq H_{\sigma_k}$.
        Call $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R})$.
        Let $(\mathbf{R}^{\mathtt{out}}, \mathcal{O})$ be its output.
    (iii) Let
        $\sigma^{\mathtt{out}} = \{\sigma_i : \sigma_i \in \sigma \text{ s.t. } \sigma_i \not\subset A \ \forall A \in \mathcal{O}\}$
                $\bigcup\{A : A \in \mathcal{O}\}$.
        Update $\mathbf{R} = \mathbf{R}^{\mathtt{out}}$, $\sigma = \sigma^{\mathtt{out}}$, and $k = |\sigma^{\mathtt{out}}|$.
```

References

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography–part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[2] ——, "On oblivious transfer capacity," *Information Theory, Combinatorics, and Search Theory*, pp. 145–166, 2013.

[3] M. Blum, "Coin flipping by telephone a protocol for solving impossible problems," *SIGACT News*, vol. 15, no. 1, pp. 23–27, Jan. 1983.

[4] M. Braverman and A. Rao, "Information equals amortized communication," in *FOCS*, 2011, pp. 748–757.

[5] C. Chan, "On tightness of mutual dependence upperbound for secret-key capacity of multiple terminals," *arXiv:0805.3200*, 2008.

[6] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, December 2004.

[7] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM Journal on Computing*, vol. 38, no. 1, pp. 97–139, 2008.

[8] S. Even, O. Goldreich, and A. Lempel, "A randomized protocol for signing contracts," *Communications of ACM*, vol. 28, no. 6, pp. 637–647, Jun. 1985.

[9] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminal: Part I," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3973 – 3996, August 2010.

[10] T. S. Han, *Information-Spectrum Methods in Information Theory [English Translation]*. Series: Stochastic Modelling and Applied Probability, Vol. 50, Springer, 2003.

[11] M. Hayashi, H. Tyagi, and S. Watanabe, "Secret key agreement: General capacity and second-order asymptotics," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3796–3810, July 2016.

[12] M. Hayashi, "Security analysis of ε-almost dual Universal$_2$ hash functions: Smoothing of min entropy versus smoothing of Rényi entropy of order 2," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Novemeber 2009.

[13] R. Impagliazzo, L. A. Levin, and M. Luby, "Pseudo-random generation from one-way functions," in *Proc. ACM Symposium on Theory of Computing (STOC)*, 1989, pp. 12–24.

[14] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[15] A. C. A. Nascimento and A. Winter, "On the oblivious-transfer capacity of noisy resources," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2572–2581, 2008.

[16] M. O. Rabin, "How to exchange secrets with oblivious transfer," Cryptology ePrint Archive, Report 2005/187, 2005.

[17] R. Renner, "Security of quantum key distribution," *Ph. D. Dissertation, ETH Zurich*, 2005.

[18] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," in *Proc. ASIACRYPT*, 2005, pp. 199–216.

[19] H. Tyagi and S. Watanabe, "Converses for secret key agreement and secure computing," *IEEE Trans. Inf. Theory*, vol. 61, pp. 4809–4827, 2015.

[20] A. Winter, A. C. A. Nascimento, and H. Imai, "Commitment capacity of discrete memoryless channels," in *Proc. Cryptography and Coding*, 2003, pp. 35–51.

[21] S. Wolf and J. Wullschleger, “New monotones and lower bounds in unconditional two-party computation,” *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2792–2797, June 2008.

[22] A. C. Yao, “Protocols for secure computations,” in *Proc. Annual Symposium on Foundations of Computer Science (FOCS)*, 1982, pp. 160–164.

[23] E.-H. Yang and D.-K. He, “Interactive encoding and decoding for one way learning: Near lossless recovery with side information at the decoder,” *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1808–1824, April 2010.