Electronic Colloquium on Computational Complexity, Report No. 70 (2015)

Information Complexity Density and Simulation of Protocols

Himanshu Tyagi*[1], Shaileshh Venkatakrishnan †[2], Pramod Viswanath ‡[2], and Shun Watanabe§[3]

[1]Indian Institute of Science, Bangalore.
[2]University of Illinois, Urbana-Champaign.
[3]Tokyo University of Agriculture and Technology

Abstract

A simulation of an interactive protocol entails the use of an interactive communication to produce the output of the protocol to within a fixed statistical distance ε. Recent works in the TCS community have propagated that the *information complexity* of the protocol plays a central role in characterizing the minimum number of bits that the parties must exchange for a successful simulation, namely the *distributional communication complexity* of simulating the protocol. Several simulation protocols have been proposed with communication complexity depending on the information complexity of the simulated protocol. However, in the absence of any general lower bounds for distributional communication complexity, the conjectured central role of information complexity is far from settled. We fill this gap and show that the distributional communication complexity of ε-simulating a protocol is bounded below by the ε-tail λ_ε of the *information complexity density*, a random variable with information complexity as its expected value. For protocols with bounded number of rounds, we give a simulation protocol that yields a matching upper bound. Thus, it is not information complexity but λ_ε that governs the distributional communication complexity.

As applications of our bounds, in the amortized regime for product protocols, we identify the exact second order term, together with the precise dependence on ε. For general protocols such as a mixture of two product protocols or for the amortized case when the repetitions are not independent, we derive a general formula for the leading asymptotic term. These results sharpen and significantly extend known results in the amortized regime. In the single-shot regime, our lower bound clarifies the dependence of communication complexity on ε. We illustrate this with an example that exhibits an arbitrary separation between distributional communication complexity and information complexity for all sufficiently small ε.

*htyagi@ece.iisc.ernet.in
†bjjvnkt2@illinois.edu
‡pramodv@illinois.edu
§shunwata@cc.tuat.ac.jp

ISSN 1433-8092

Contents

1 Introduction

Two parties observing random variables X and Y seek to run an interactive protocol π with inputs X and Y. The parties have access to private as well as shared public randomness. What is the minimum number of bits that they must exchange in order to simulate π to within a fixed statistical distance ε? This question is of importance to the theoretical computer science as well as the information theory communities. On the one hand, it is related closely to the communication complexity problem [Yao79], which in turn is an important tool for deriving lower bounds for computational complexity [KW88] and for space complexity of streaming algorithms [AMS96]. On the other hand, it is a significant generalization of the classical information theoretic problem of distributed data compression [SW73], replacing data to be compressed with an interactive protocol and allowing interactive communication as opposed to the usual one-sided communication.

In recent years, it has been argued that the distributional communication complexity for simulating a protocol π is related closely to its *information complexity*[1] $\mathtt{IC}(\pi)$ defined as follows:

$$\mathtt{IC}(\pi) \overset{\text{def}}{=} I(\Pi \wedge X|Y) + I(\Pi \wedge Y|X),$$

where $I(X \wedge Y|Z)$ denotes the conditional mutual information between X and Y given Z (*cf.* [Sha48, CK11]). For a protocol π with communication complexity $\|\pi\|$ (the depth of the binary protocol tree), a simulation protocol requiring $\tilde{\mathcal{O}}(\sqrt{\mathtt{IC}(\pi)\|\pi\|})$ bits of communication was given in [BBCR10] and one requiring $2^{\mathcal{O}(\mathtt{IC}(\pi))}$ bits of communication was given in [Bra12]. Interestingly, it was shown in [BR11] that the amortized distributional communication complexity of simulating n copies of a protocol π for vanishing simulation error is bounded above by[2] $\mathtt{IC}(\pi)$. While a matching lower bound was also derived in [BR11], it is not valid in our context – [BR11] considered function computation and used a coordinate-wise error criterion. In fact, none of the works discussed above gave a precise definition of simulation[3] of a protocol, perhaps owing to their focus on the upper bounds for distributional communication complexity of protocol simulation rather than the lower bounds. A general version of the simulation problem was considered in [YGA12], but only bounded round simulation protocols were considered. Nevertheless, we can readily modify the lower bound argument in [BR11] and use the continuity of conditional mutual information to formally obtain the required lower bound and thereby a characterization of the amortized distributional communication complexity for vanishing simulation error. Specifically, denoting by $D(\pi^n)$ the distributional communication complexity of simulating n copies of a protocol π with vanishing simulation error, we have

$$\lim_{n\to\infty} \frac{1}{n} D(\pi^n) = \mathtt{IC}(\pi).$$

Perhaps motivated by this characterization, or a folklore version of it, the research in this area has focused on designing simulation protocols for π requiring communication of length depending on $\mathtt{IC}(\pi)$; the results cited above belong to this category as well. However, the central role of $\mathtt{IC}(\pi)$ in the distributional communication complexity of protocol simulation is far from settled and many important questions remain unanswered. For instance, (a) Does $\mathtt{IC}(\pi)$ suffice to capture the

[1] For brevity, we do not display the dependence of $\mathtt{IC}(\pi)$ on the (fixed) distribution P_{XY}.

[2] Braverman and Rao actually used their general simulation protocol as a tool for deriving the amortized distributional communication complexity of function computation. This result was obtained independently by Ma and Ishwar in [MI11] using standard information theoretic techniques.

[3] Prior literature makes no explicit distinction between simulation and compression of protocols. However, the difference is significant and is discussed in Remark 2 below.

dependence of distributional communication complexity on the simulation error ε? (b) Does information complexity have an operational role in simulating π^n besides being the leading asymptotic term? (c) How about the simulation of more complicated protocols such as a mixture $\pi_{\texttt{mix}}$ of two product protocols π_1^n and π_2^n – does $\texttt{IC}(\pi_{\texttt{mix}})$ still constitute the leading asymptotic term in the communication complexity of simulating $\pi_{\texttt{mix}}$?

In this paper, we answer all these questions in the negative by exhibiting another quantity that plays such a fundamental role and can differ from information complexity significantly. To this end, we introduce the notion of *information complexity density* of a protocol π with inputs X and Y generated from a fixed distribution P_{XY}.

Definition 1 (Information complexity density). The *information complexity density* of a private coin protocol π is given by the function

$$\texttt{ic}(\tau; x, y) = \log \frac{\mathrm{P}_{\Pi|XY}(\tau|x,y)}{\mathrm{P}_{\Pi|X}(\tau|x)} + \log \frac{\mathrm{P}_{\Pi|XY}(\tau|x,y)}{\mathrm{P}_{\Pi|Y}(\tau|y)},$$

for all observations x and y of the two parties and all transcripts τ, where $\mathrm{P}_{\Pi XY}$ denotes the joint distribution of the observation of the two parties and the random transcript Π generated by π.

Note that $\texttt{IC}(\pi) = \mathbb{E}\left[\texttt{ic}(\Pi; X, Y)\right]$. We show that it is the *ε-tail of the information complexity density* $\texttt{ic}(\Pi; X, Y)$, *i.e.*, the supremum[4] over values of λ such that $\Pr\left(\texttt{ic}(\Pi; X, Y) > \lambda\right) \le \varepsilon$, which governs the communication complexity of simulating a protocol with simulation error less than ε and not the information complexity of the protocol. The information complexity $\texttt{IC}(\pi)$ becomes the leading term in communication complexity for simulating π only when roughly

$$\texttt{IC}(\pi) \gg \sqrt{\mathrm{Var}(\texttt{ic}(\Pi; X, Y)) \log(1/\varepsilon)}.$$

This condition holds, for instance, in the amortized regime considered in [BR11]. However, the ε-tail of $\texttt{ic}(\Pi; X, Y)$ can differ significantly from $\texttt{IC}(\pi)$, the mean of $\texttt{ic}(\Pi; X, Y)$. In Appendix A, we provide an example protocol with inputs of size 2^n such that for $\varepsilon = 1/n^3$, the ε-tail of $\texttt{ic}(\Pi; X, Y)$ is greater than $2n$ while $\texttt{IC}(\pi)$ is very small, just $\tilde{\mathcal{O}}(n^{-2})$.

1.1 Summary of results

Our main results are bounds for distributional communication complexity $D_\varepsilon(\pi)$ for ε-simulating a protocol π. The key quantity in our bounds is the ε-tail λ_ε of $\texttt{ic}(\Pi; X, Y)$.

Lower bound. Our main contribution is a general lower bound for $D_\varepsilon(\pi)$. We show that for every private coin protocol π, $D_\varepsilon(\pi) \gtrsim \lambda_\varepsilon$. In fact, this bound does not rely on the structure of random variable Π and is valid for the more general problem of simulating a correlated random variable.

Prior to this work, there was no lower bound that captures both the dependence on simulation error ε as well as the underlying probability distribution. On the one hand, the lower bound above yields many sharp results in the amortized regime. It gives the leading asymptotic term in the communication complexity for simulating any sequence of protocols, and not just product protocols. For product protocols, it yields the precise dependence of communication complexity on ε as well as the exact second-order asymptotic term. On the other hand, it clarifies the dependence of $D_\varepsilon(\pi)$ on ε even in the single-shot regime. For instance, the result of [Bra12] is often stated as $D_\varepsilon(\pi) = \mathcal{O}(2^{\texttt{IC}(\pi)})$, a form which hides the dependence on ε. However, our lower bound shows that

[4] Formally, our lower bound uses lower ε-tail $\sup\{\lambda : \Pr\left(\texttt{ic}(\Pi; X, Y) > \lambda\right) > \varepsilon\}$ and the upper bound uses upper ε-tail $\inf\{\lambda : \Pr\left(\texttt{ic}(\Pi; X, Y) > \lambda\right) < \varepsilon\}$. For many interesting cases, the two coincide.

this can be misleading. Consider the example protocol in Appendix A. On evaluating our lower bound for this protocol, for $\varepsilon = 1/n^3$ we get $D_\varepsilon(\pi) = \Omega(n)$ which is far more than $2^{\mathtt{IC}(\pi)}$ since $\mathtt{IC}(\pi) = \tilde{\mathcal{O}}(n^{-2})$. Remarkably, [GKR14b, GKR14a] exhibited exponential separation between the distributional communication complexity of computing a function and the information complexity of that function even for a large ε. Our simple example shows a much stronger separation between $D_\varepsilon(\pi)$ and $\mathtt{IC}(\pi)$, albeit for a small ε.

Upper bound. To establish our asymptotic results, we propose a new simulation protocol, which is of independent interest. For a protocol π with bounded rounds of interaction, using our proposed protocol we can show that $D_\varepsilon(\pi) \lesssim \lambda_\varepsilon$. Much as the protocol of [BR11], our simulation protocol simulates one round at a time, and thus, the slack in our upper bound does depend on the number of rounds.

Note that while the operative term in the lower bound and the upper bound is the ε-tail of $\mathtt{ic}(\Pi; X, Y)$, the lower bound approaches it from below and the upper bound approaches it from above. It is often the case that these two limits match and the leading term in our bounds coincide. See Figure 1 for an illustration of our bounds.

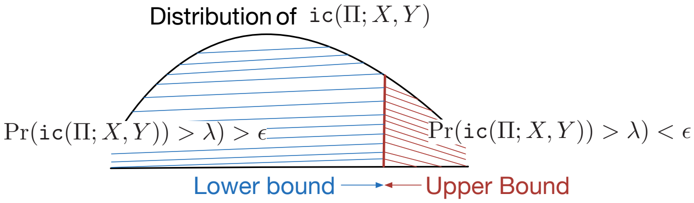
Distribution of $\mathtt{ic}(\Pi; X, Y)$
$\Pr(\mathtt{ic}(\Pi; X, Y)) > \lambda) > \epsilon$
$\Pr(\mathtt{ic}(\Pi; X, Y)) > \lambda) < \epsilon$
Lower bound
Upper Bound

Figure 1: Illustration of lower and upper bounds for $D_\varepsilon(\pi)$

Amortized regime: second-order asymptotics. Denote by π^n the n-fold product protocol obtained by applying π to each coordinate (X_i, Y_i) for inputs X^n and Y^n. Consider the communication complexity $D_\varepsilon(\pi^n)$ of ε-simulating π^n for *independent and identically distributed* (IID) (X^n, Y^n) generated from P^n_{XY}. Using the bounds above, we can obtain the following sharpening of the results of [BR11]: With $\mathtt{V}(\pi)$ denoting the variance of $\mathtt{ic}(\Pi; X, Y)$,

$$D_\varepsilon(\pi^n) = n\mathtt{IC}(\pi) + \sqrt{n\mathtt{V}(\pi)}Q^{-1}(\varepsilon) + o(\sqrt{n}),$$

where $Q(x)$ is equal to the probability that a standard normal random variable exceeds x and $Q^{-1}(\varepsilon) \approx \sqrt{\log(1/\varepsilon)}$. On the other hand, the arguments in[5] [BR11] or [YGA12] give us

$$D_\varepsilon(\pi^n) \geq n\mathtt{IC}(\pi) - n\varepsilon[\|\pi\| + \log|\mathcal{X}||\mathcal{Y}|] - \varepsilon\log(1/\varepsilon).$$

But the precise communication requirement is not less but $\sqrt{n\mathtt{V}(\pi)\log(1/\varepsilon)}$ *more than* $n\mathtt{IC}(\pi)$.

General formula for amortized communication complexity. The lower and upper bounds above can be used to derive a formula for the first-order asymptotic term, the coefficient of n, in $D_\varepsilon(\pi_n)$ for any sequence of protocols π_n with inputs $X_n \in \mathcal{X}^n$ and $Y_n \in \mathcal{Y}^n$ generated from any sequence of distributions $\mathrm{P}_{X_nY_n}$. We illustrate our result by the following example.

Example 1 (Mixed protocol). Consider two protocols $\pi_{\mathtt{h}}$ and $\pi_{\mathtt{t}}$ with inputs X and Y such that $\mathtt{IC}(\pi_{\mathtt{h}}) > \mathtt{IC}(\pi_{\mathtt{t}})$. For n IID observations (X^n, Y^n) drawn from P_{XY}, we seek to simulate the mixed protocol $\pi_{\mathtt{mix,n}}$ defined as follows: Party 1 first flips a (private) coin with probability p of heads and sends the outcome Π_0 to Party 2. Depending on the outcome of the coin, the parties execute $\pi_{\mathtt{h}}$

[5] The proof in [BR11] uses the inequality $\mathtt{IC}(\pi) \leq \|\pi\|$, a multiparty extension of which is available in [CN08, MT10].

or $\pi_{\mathtt{t}}$ n times, i.e., they use $\pi_{\mathtt{h}}^n$ if $\Pi_0 = \mathtt{h}$ and $\pi_{\mathtt{t}}^n$ if $\Pi_0 = \mathtt{t}$. What is the amortized communication complexity of simulating the mixed protocol $\pi_{\mathtt{mix,n}}$? Note that

$$\mathtt{IC}(\pi_{\mathtt{mix,n}}) = n\left[p\mathtt{IC}(\pi_{\mathtt{h}}) + (1-p)\mathtt{IC}(\pi_{\mathtt{t}})\right].$$

Is it true that in the manner of [BR11] the leading asymptotic term in $D_\varepsilon(\pi_{\mathtt{mix,n}})$ is $\mathtt{IC}(\pi_{\mathtt{mix,n}})$? In fact, it is not so. Our general formula implies that for all $p \in (0,1)$,

$$D_\varepsilon(\pi_{\mathtt{mix,n}}) = n\mathtt{IC}(\pi_{\mathtt{h}}) + o(n)$$

This is particularly interesting when p is very small and $\mathtt{IC}(\pi_{\mathtt{h}}) \gg \mathtt{IC}(\pi_{\mathtt{t}})$.

The results above illustrate the central thesis of this paper: It is the ε-tail of $\mathtt{ic}(\Pi; X, Y)$ and not just $\mathtt{IC}(\pi)$ that governs the communication complexity of ε-simulating a protocol π.

1.2 Proof techniques

Proof for the lower bound. We present a new method for deriving lower bounds on distributional communication complexity. Our proof relies on a reduction argument that utilizes an ε-simulation to generate an information theoretically secure secret key for X and Y (for a definition of the latter, see [Mau93, AC93] or Section 4). Heuristically, a protocol can be simulated using fewer bits of communication than its length because of the correlation in the observations X and Y. Due to this correlation, when simulating the protocol, the parties agree on more bits (generate more *common randomness*) than what they communicate. These extra bits can be extracted as an information theoretically secure secret key for the two parties using the *leftover hash lemma* (*cf.* [BBCM95, RW05]). A lower bound on the number of bits communicated can be derived using an upper bound for the maximum possible length of a secret key that can be generated using interactive communication; the latter was derived recently in [TW14a, TW14b].

Protocol for the upper bound. We simulate a given protocol one round at a time. Simulation of each round consists of two subroutines: Interactive Slepian-Wolf compression and message reduction by public randomness. The first subroutine is an interactive version of the classical Slepian-Wolf compression [SW73] for sending X to an observer of Y which is of optimal instantaneous rate. The second subroutine uses an idea that appeared first in [RR11] (see, also, [Mur14, YAG14]) and reduces the number of bits communicated in the first by realizing a portion of the required communication by the shared public randomness. This is possible since we are not required to recover a given random variable Π, but only simulate it to within a fixed statistical distance.

The proposed protocol is closely related to that proposed in [BR11]. However, there are some crucial differences. The protocol in [BR11], too, uses public randomness to sample each round of the protocol, before transmitting it using an interactive communication of size incremented in steps. However, our information theoretic approach provides a systematic method for choosing this step size. Furthermore, our protocol for sampling the protocol from public randomness is significantly different from that in [BR11] and relies on randomness extraction techniques. In particular, the protocol in [BR11] does not attain the asymptotically optimal bounds achieved by our protocol.

Technical approach. While we utilize new, bespoke techniques for deriving our lower and upper bounds, casting our problem in an information theoretic framework allows us to build upon the developments in this classic field. In particular, we rely on the *information spectrum approach* of Han and Verdú, introduced in the seminal paper [HV93] (see the textbook [Han03] for a detailed account). In this approach, the classical measures of information such as entropy and mutual information are viewed as expectations of the corresponding *information densities*, and the notion of “typical sets” is replaced by sets where these information densities are bounded uniformly. The set of values taken by an information density (such as $h(x) = -\log \mathrm{P}_X(x)$) is called its *spectrum*. Coding theorems of classical information theory consider IID repetitions and rely on the so-called the *asymptotic equipartition property* [CT06] which essentially corresponds to the concentration of spectrums on small intervals. For *single-shot* problems such concentrations are not available and we have to work with the whole span of the spectrum.

Our main technical contribution in this paper is the extension of the information spectrum method to handle interactive communication. Our results rely on the analysis of appropriately chosen information densities and, in particular, will rely on the spectrum of the information complexity density $\mathtt{ic}(\Pi; X, Y)$. As is usually the case, different components of our analysis require bounds on these information densities in different directions, which in turn renders our bounds loose and incurs a gap equal to the length of the corresponding information spectrum. To overcome this shortcoming, we use the *spectrum slicing* technique of Han [Han03][6] to divide the information spectrum into small portions with information densities closely bounded from both sides. While in our upper bounds spectrum slicing is used to carefully choose the parameters of the protocol, it is required in our lower bounds to identify a set of inputs where a given simulation will require a large number of bits to be communicated.

1.3 Organization

A formal statement of the problem along with the necessary preliminaries is given in the next section. Section 3 contains all our results. In Section 4, we review the information theoretic secret key agreement problem, the leftover hash lemma, and the data exchange problem, all of which will be instrumental in our proofs. The formal proof of our lower bound is contained in Section 5 and that of our upper bound in Section 6. The final section contains a derivation of our asymptotic results.

1.4 Notations

Random variables are denoted by capital letters such as X, Y, *etc.* realizations by small letters such as x, y, *etc.* and their range sets by corresponding calligraphic letters such as $\mathcal{X}$, $\mathcal{Y}$, *etc.*. Protocols are denoted by appropriate subscripts or superscripts with π, the corresponding random transcripts by the same sub- or superscripts with Π; τ is used as a placeholder for realizations of random transcripts. All the logarithms in this paper are to the base 2.

The following convention, described for the entropy density, shall be used for all information densities used in this paper. We shall abbreviate the entropy density $h_{\mathrm{P}_X}(x) = -\log \mathrm{P}_X(x)$ by $h(x)$, when there is no confusion about P_X, and the random variable $h(X)$ corresponds to drawing X from the distribution P_X.

Whenever there is no confusion, we will not display the dependence of distributional communication complexity on the underlying distribution. In most of our discussion, the latter remains fixed.

2 Problem Statement

Two parties observe correlated random variables X and Y, with Party 1 observing X and Party 2 observing Y, generated from a fixed distribution P_{XY} and taking values in finite sets $\mathcal{X}$ and $\mathcal{Y}$,

[6] The spectrum slicing technique was introduced in [Han03] to derive the error exponents of various problems for general sources and a rate-distortion function for general sources.

respectively. An *interactive protocol* π (for these two parties) consists of shared public randomness U, private randomness[7] $U_{\mathcal{X}}$ and $U_{\mathcal{Y}}$, and interactive communication $\Pi_1, ..., \Pi_r$. The parties communicate alternatively with Party 1 transmitting in the odd rounds and Party 2 in the even rounds. Specifically, Π_i is a string of bits determined by the previous transmissions $\Pi_1, ..., \Pi_{i-1}$ together with $(X, U_{\mathcal{X}}, U)$ for odd i and $(Y, U_{\mathcal{Y}}, U)$ for even i. The number of rounds of communication r is a random stopping-time such that the event $\{r = t\}$ is determined by the transcript $\Pi_1, ..., \Pi_t$; we denote the overall transcript of the protocol[8] by Π. The length of a protocol π, $\|\pi\|$, is the maximum number of bits that are communicated in any execution of the protocol.

A random variable F is said to be *recoverable* by π for Party 1 (or Party 2) if F is function of $(X, U, U_{\mathcal{X}}, \Pi)$ (or $(Y, U, U_{\mathcal{Y}}, \Pi)$).

A protocol with a constant U is called a *private coin protocol*, that with a constant $(U_{\mathcal{X}}, U_{\mathcal{Y}})$ is called a *public coin protocol*, and with $(U, U_{\mathcal{X}}, U_{\mathcal{Y}})$ constant is called a *deterministic protocol*.

When we execute the protocol π above, the overall *view* of the parties consists of random variables $(XY\Pi\Pi)$, where the two Πs correspond to the transcript of the protocol seen by the two parties. A simulation of the protocol consists of another protocol which generates almost the same view as that of the original protocol. We are interested in the simulation of private coin protocols, using arbitrary[9] protocols; public coin protocols can be simulated by simulating for each fixed value of public randomness the resulting private coin protocol.

Definition 2 (**ε-Simulation of a protocol**)**.** Let π be a private coin protocol. Given $0 \le \varepsilon < 1$, a protocol $\pi_{\mathtt{sim}}$ constitutes an ε-simulation of π if there exist $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Y}}$, respectively, recoverable by $\pi_{\mathtt{sim}}$ for Party 1 and Party 2 such that

$$d_{\mathtt{var}}\left(\mathrm{P}_{\Pi\Pi XY}, \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}\right) \le \varepsilon, \tag{1}$$

where $d_{\mathtt{var}}\left(\mathrm{P}, \mathrm{Q}\right) = \frac{1}{2}\sum_x |\mathrm{P}_x - \mathrm{Q}_x|$ denotes the variational or the statistical distance between P and Q.

Definition 3 (**Distributional communication complexity**)**.** The ε-error distributional communication complexity $D_\varepsilon\left(\pi|\mathrm{P}_{XY}\right)$ of simulating a private coin protocol π is the minimum length of an ε-simulation of π. The distribution P_{XY} remains fixed throughout our analysis; for brevity, we shall abbreviate $D_\varepsilon\left(\pi|\mathrm{P}_{XY}\right)$ by $D_\varepsilon\left(\pi\right)$.

Problem. Given a protocol π and a joint distribution P_{XY} for the observations of the two parties, we seek to characterize $D_\varepsilon\left(\pi\right)$.

Remark 1 (**Deterministic protocols**). Note that a deterministic protocol corresponds to an *interactive function*, and for such protocols,

$$d_{\mathtt{var}}\left(\mathrm{P}_{\Pi\Pi XY}, \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}\right) = 1 - \Pr\left(\Pi = \Pi_{\mathcal{X}} = \Pi_{\mathcal{Y}}\right).$$

Therefore, a protocol is an ε-simulation of a deterministic protocol if and only if it computes the corresponding interactive function with probability of error less than ε. Furthermore, randomization does not help in this case, and it suffices to use deterministic simulation protocols. Thus, our results below provide tight bounds for distributional communication complexity of interactive

[7] The random variables $U, U_{\mathcal{X}}, U_{\mathcal{Y}}$ are mutually independent and independent jointly of (X, Y).

[8] We allow Π_i to be constant and allow it to depend only on the local observation (and not on the previous communication $\Pi_1, ..., \Pi_{i-1}$. This description of an interactive protocol is the most general possible and is equivalent to the usual protocol-tree based description (*cf.* [BBCR10, BR11]).

[9] Since we are not interested in minimizing the amount of randomness used in a simulation, and private randomness can always be sampled from public randomness, we can restrict ourselves to public protocols for simulating.

functions and, in fact, of all functions which are *information theoretically securely computable* for the distribution P_{XY}, since computing these functions is tantamount to computing an interactive function [NTW15] (see, also, [Bea89, Kus92]).

Remark 2 (**Compression of protocols**). A protocol $\pi_{\mathtt{com}}$ constitutes an ε-compression of a given protocol π if it recovers $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Y}}$ for Party 1 and Party 2 such that

$$\Pr\left(\Pi = \Pi_{\mathcal{X}} = \Pi_{\mathcal{Y}}\right) \geq 1 - \varepsilon.$$

Note that randomization does not help in this case either. In fact, for deterministic protocols simulation and compression coincide. In general, however, compression is a more demanding task than simulation and our results show that in many cases, (such as the amortized regime), compression requires strictly more communication than simulation. Specifically, our results for ε-simulation in this paper can be modified to get corresponding results for ε-compression by replacing the information complexity density $\mathtt{ic}(\tau; x, y)$ by

$$h(\tau|x) + h(\tau|y) = -\log \mathrm{P}_{\Pi|X}\left(\tau|x\right) \mathrm{P}_{\Pi|Y}\left(\tau|y\right).$$

The proofs remain essentially the same and, in fact, simplify significantly.

3 Main Results

We derive a lower bound for $D_\varepsilon\left(\pi\right)$ which applies to all private coin protocols π and, in fact, applies to the more general problem of communication complexity of sampling a correlated random variable. For protocols with bounded number of rounds of interaction, *i.e.*, protocols with $r = r(X, Y, U, U_{\mathcal{X}}, U_{\mathcal{Y}}) \leq r_{\max}$ with probability 1, we present a simulation protocol which yields upper bounds for $D_\varepsilon\left(\pi\right)$ of similar form as our lower bounds. In particular, in the asymptotic regime our bounds improve over previously known bounds and are tight.

3.1 Lower bound

We prove the following lower bound.

Theorem 1. *Given $0 \leq \varepsilon < 1$ and a protocol π, for arbitrary $0 < \eta < 1/3$*

$$D_\varepsilon\left(\pi\right) \geq \sup\{\lambda : \Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) \geq \varepsilon + \varepsilon'\} - \lambda', \tag{2}$$

where $(x)_+$ denotes $\max\{0, x\}$ and the fudge parameters ε' and λ' depend on η as well as appropriately chosen information spectrums and will be described below in (4) and (5).

When the fudge parameters ε' and λ' are negligible, the right-side of the bound above is close to ε-tail of $\mathtt{ic}(\Pi; X, Y)$. Indeed, the fudge parameters turn out to be negligible in many cases of interest. For instance, for the amortized case ε' can be chosen to be arbitrarily small. The parameter λ' is related to the length of the interval in which the underlying information densities lie with probability greater than $1 - \varepsilon'$, the essential length of spectrums. For the amortized case with product protocols, by the central limit theorem the related essential spectrums are of length $\Lambda = \mathcal{O}(\sqrt{n})$ and $\lambda' = \log \Lambda$. On the other hand, λ_ε is $\mathcal{O}(n)$. Thus, the $\log n$ order fudge parameter λ' is negligible in this case. The same is true also for the example protocol in Appendix A.

Remark 3. The result above does not rely on the interactive nature of Π and is valid for simulation of any random variable. Specifically, for any joint distribution $\mathrm{P}_{\Pi XY}$, an ε-simulation satisfying (1) must communicate at least as many bits as the right-side of (2), which is roughly equal to the largest value λ_ε of λ such that $\Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) > \varepsilon$.

The fudge parameters. The fudge parameters ε' and λ' in Theorem 1 depend on the spectrums of the following information densities:

(i) *Information complexity density:* This density is described in Definition 1 and will play a pivotal role in our results.

(ii) *Entropy density of* (X, Y)*:* This density, given by $h(X, Y) = -\log \mathrm{P}_{XY}(X, Y)$, captures the randomness in the data and plays a fundamental role in the compression of the collective data of the two parties (*cf.* [Han03]).

(iii) *Conditional entropy density of* X *given* $Y\Pi$*:* The conditional entropy density $h(X|Y) = -\log \mathrm{P}_{X|Y}(X|Y)$ plays a fundamental role in the compression of X for an observer of Y [MK95, Han03]. We shall use the conditional entropy density $h(X|Y\Pi)$ in our bounds.

(iv) *Sum conditional entropy density of* $(X\Pi, Y\Pi)$*:* The sum conditional entropy density is given by $h(X \triangle Y) = -\log \mathrm{P}_{X|Y}(X|Y)\,\mathrm{P}_{Y|X}(Y|X)$ has been shown recently to play a fundamental role in the communication complexity of the data exchange problem [TVW15]. We shall use the sum conditional entropy density $h(X\Pi \triangle Y\Pi)$.

(v) Information density of X given Y is given by $i(X \wedge Y) \overset{\text{def}}{=} h(X) - h(X|Y)$.

Let $[\lambda^{(1)}_{\min}, \lambda^{(1)}_{\max}]$, $[\lambda^{(2)}_{\min}, \lambda^{(2)}_{\max}]$, and $[\lambda^{(3)}_{\min}, \lambda^{(3)}_{\max}]$ denote the "essential" spectrums of information densities $\zeta_1 = h(X, Y)$, $\zeta_2 = h(X|Y\Pi)$, and $\zeta_3 = h(X\Pi \triangle Y\Pi)$, respectively. Concretely, let the tail events $\mathcal{E}_i = \{\zeta_i \notin [\lambda^{(i)}_{\min}, \lambda^{(i)}_{\max}]\}$, $i = 1, 2, 3$, satisfy

$$\Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_3) \le \varepsilon_{\mathtt{tail}}, \tag{3}$$

where $\varepsilon_{\mathtt{tail}}$ can be chosen to be appropriately small. Further, let $\Lambda_i = \lambda^{(i)}_{\max} - \lambda^{(i)}_{\min}$, $i = 1, 2, 3$, denote the corresponding effective spectrum lengths. The parameters ε' and λ' in Theorem 1 are given by

$$\varepsilon' = \varepsilon_{\mathtt{tail}} + 2\eta \tag{4}$$

and

$$\lambda' = 2\log \Lambda_1 \Lambda_3 + \log \Lambda_2 - \log(1 - 3\eta) + 9\log 1/\eta + 3, \tag{5}$$

where $0 < \eta < 1/3$ is arbitrary. If $\Lambda_i = 0$, $i = 1, 2, 3$, we can replace it with 1 in the bound above. Thus, our spectrum slicing approach allows us to reduce the dependence of λ' on spectrum lengths Λ_i's from linear to logarithmic.

3.2 Upper bound

We prove the following upper bound.

Theorem 2. *For every* $0 \le \varepsilon < 1$ *and every protocol* π*,*

$$D_\varepsilon(\pi) \le \inf\left\{\lambda : \Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) \le \varepsilon - \varepsilon'\right\} + \lambda',$$

where the fudge parameters ε' *and* λ' *depend on the maximum number of rounds of interaction in* π *and on appropriately chosen information spectrums.*

Remark 4. In contrast to the lower bound given in the previous section, the upper bound above relies on the interactive nature of π. Furthermore, the fudge parameters ε' and λ' depend on the number of rounds, and the upper bound may not be useful when the number of rounds is not negligible compared to ε-tail of the information complexity density. However, we will see that the above upper bound is tight for the amortized regime, even up to the second-order asymptotic term.

The simulation protocol. Our simulation protocol simulates the given protocol π round-by-round, starting from Π_1 to Π_r. Simulation of each round consists of two subroutines: Interactive Slepian-Wolf compression and message reduction by public randomness.

The first subroutine uses an interactive version of the classical Slepian-Wolf compression [SW73] (see [MK95] for a single-shot version) for sending X to an observer of Y. The standard (noninteractive) Slepian-Wolf coding entails hashing X to l values and sending the hash values to the observer of Y. The number of hash values l is chosen to take into account the worst-case performance of the protocol. However, we are not interested in the worst-case performance of each round, but of the overall multiround protocol. As such, we seek to compress X using the least possible instantaneous rate. To that end, we increase the number of hash values gradually, Δ at a time, until the receiver decodes X and sends back an ACK. We apply this subroutine to each round i, say i odd, with Π_i in the role of X and $(Y, \Pi_1 ...., \Pi_{i-1})$ in the role of Y. Similar interactive Slepian-Wolf compression schemes have been considered earlier in different contexts (*cf.* [FS02, Orl90, YH10, HTW14, TVW15]).

The second subroutine reduces the number of bits communicated in the first by realizing a portion of the required communication by the shared public randomness U. Specifically, instead of transmitting hash values of Π_i, we transmit hash values of a random variable $\hat{\Pi}_i$ generated in such a manner that some of its corresponding hash bits can be extracted from U and the overall joint distributions do not change by much. Since U is independent of (X, Y), the number k of hash bits that can be realized using public randomness is the maximum number of random hash bits of Π_i that can be made almost independent of (X, Y), a good bound for which is given by the leftover hash lemma. The overall simulation protocol for Π_i now communicates $l - k$ instead of l bits. A similar technique for message reduction appears in a different context in [RR11, Mur14, YAG14].

The overall performance of the protocol above is still suboptimal because the saving of k bits is limited by the worst-case performance. To remedy this shortcoming, we once again take recourse to spectrum slicing to ensure that our saving k is close to the best possible for each realization (Π, X, Y).

Note that our protocol above is closely related to that proposed in [BR11]. However, the information theoretic form here makes it amenable to techniques such as spectrum slicing, which leads to tighter bounds than those established in [BR11].

The fudge parameters. The fudge parameters ε' and λ' in Theorem 2 depend on the spectrum of various conditional information densities. Our simulation protocol simulates π one round at a time. Simulation of each round consists of two subroutines: Interactive Slepian-Wolf compression and message reduction by public randomness. To optimize the performance of each subroutine, we slice the spectrum of the respective conditional information density involved. Specifically, for odd round t, we slice the spectrum of $h(\Pi_t|Y\Pi^{t-1}) = -\log \mathrm{P}_{\Pi_t|Y\Pi^{t-1}} \left(\Pi_t|Y, \Pi^{t-1}\right)$ for interactive Slepian-Wolf compression and $h(\Pi_t|X\Pi^{t-1}) = -\log \mathrm{P}_{\Pi_t|X\Pi^{t-1}} \left(\Pi_t|X, \Pi^{t-1}\right)$ for the substitution of message by public randomness; for even rounds, the role of X and Y is interchanged. Each round involves some residuals related to the two conditional information densities. Then, the fudge parameters ε' and λ' are accumulations of the residuals of each round. See Remark 5 in Section 6.5 for explicit expressions for ε' and λ'.

3.3 Amortized regime: second-order asymptotics

It was shown in [BR11] that information complexity of a protocol equals the amortized communication rate for simulating the protocol, *i.e.*,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}D_\varepsilon(\pi^n|\mathrm{P}^n_{XY}) = \mathtt{IC}(\pi),$$

where P^n_{XY} denotes the n-fold product of the distribution P_{XY}, namely the distribution of random variables $(X_i, Y_i)_{i=1}^n$ drawn IID from P_{XY}, and π^n corresponds to running the same protocol π on every coordinate (X_i, Y_i). Thus, $\mathtt{IC}(\pi)$ is the first-order term (coefficient of n) in the communication complexity of simulating the n-fold product of the protocol. However, the analysis in [BR11] sheds no light on finer asymptotics such as the second-order term or the dependence of $D_\varepsilon(\pi^n|\mathrm{P}^n_{XY})$ on[10] ε. On the one hand, it even remains unclear from [BR11] if a positive ε reduces the amortized communication rate or not. On the other hand, the amortized communication rate yields only a loose bound for $D_\varepsilon(\pi^n|\mathrm{P}^n_{XY})$ for a finite, fixed n. A better estimate of $D_\varepsilon(\pi^n|\mathrm{P}^n_{XY})$ at a finite n and for a fixed ε can be obtained by identifying the second-order asymptotic term. Such second-order asymptotics were first considered in [Str62] and have received a lot of attention in information theory in recent years following [Hay09, PPV10].

Our lower bound in Theorem 1 and upper bound in Theorem 2 show that the leading term in $D_\varepsilon(\pi^n|\mathrm{P}^n_{XY})$ is roughly the ε-tail λ_ε of the random variable $\mathtt{ic}(\Pi^n; X^n, Y^n) = \sum_{i=1}^n \mathtt{ic}(\Pi_i; X_i, Y_i)$, a sum of n IID random variables. By the central limit theorem the first-order asymptotic term in λ_ε equals $n\mathbb{E}\left[\mathtt{ic}\Pi; X, Y\right] = \mathtt{IC}(\pi)$, recovering the result of [BR11]. Furthermore, the second-order asymptotic term depends on the variance $\mathtt{V}(\pi)$ of $\mathtt{ic}(\Pi; X, Y)$, *i.e.*, on

$$\mathtt{V}(\pi) \overset{\text{def}}{=} \mathrm{Var}\left[\mathtt{ic}(\Pi; X, Y)\right].$$

We have the following result.

Theorem 3. *For every $0 < \varepsilon < 1$ and every protocol π with $\mathtt{V}(\pi) > 0$,*

$$D_\varepsilon(\pi^n|\mathrm{P}^n_{XY}) = n\mathtt{IC}(\pi) + \sqrt{n\mathtt{V}(\pi)}Q^{-1}(\varepsilon) + o(\sqrt{n}),$$

where $Q(x)$ is equal to the probability that a standard normal random variable exceeds x.

As a corollary, we obtain the so-called *strong converse.*

Corollary 4. *For every $0 < \varepsilon < 1$, the amortized communication rate*

$$\lim_{n\to\infty}\frac{1}{n}D_\varepsilon(\pi^n|\mathrm{P}^n_{XY}) = \mathtt{IC}(\pi).$$

Corollary 4 implies that the amortized communication complexity of simulating protocol π cannot be smaller than its information complexity even if we allow a positive error. Thus, if the length of the simulation protocol $\pi_{\mathtt{sim}}$ is "much smaller" than $n\mathtt{IC}(\pi)$, the corresponding simulation error $\varepsilon = \varepsilon_n$ must approach 1. But how fast does this ε_n converge to 1? Our next result shows that this convergence is exponentially rapid in n.

Theorem 5. *Given a protocol π and an arbitrary $\delta > 0$, for any simulation protocol $\pi_{\mathtt{sim}}$ with*

$$\|\pi_{\mathtt{sim}}\| \le n[\mathtt{IC}(\pi) - \delta],$$

there exists a constant $E = E(\delta) > 0$ such that for every n sufficiently large, it holds that

$$d_{\mathtt{var}}\left(\mathrm{P}_{\Pi^n\Pi^n X^n Y^n}, \mathrm{P}_{\Pi^n_{\mathcal{X}}\Pi^n_{\mathcal{Y}} X^n Y^n}\right) \ge 1 - 2^{-En}.$$

[10] The lower bound in [BR11] gives only the *weak converse* which holds only when $\varepsilon = \varepsilon_n \to 0$ as $n \to \infty$.

A similar converse was first shown for the channel coding problem in information theory by Arimoto [Ari73] (see [DK79, PV10] for further refinements of this result), and has been studied for other classical information theory problems as well. To the best of our knowledge, Corollary 5 is the first instance of an Arimoto converse for a problem involving interactive communication.

In the TCS literature, such converse results have been termed *direct product theorems* and have been considered in the context of the (distributional) communication complexity problem (for computing a given function) [BRWY13, BW14, JPY12]. Our lower bound in Theorem 1, too, yields a direct product theorem for the communication complexity problem. We state this simple result in the passing, skipping the details since they closely mimic Theorem 5. Specifically, given a function f on $\mathcal{X} \times \mathcal{Y}$, by slight abuse of notations and terminologies, let $D_\varepsilon(f) = D_\varepsilon(f|\mathrm{P}_{XY})$ be the communication complexity of computing f. As noted in Remark 3, Theorem 1 is valid for an arbitrary random variables Π, and not just an interactive protocol. Then, by following the proof of Theorem 5 with $F = f(X,Y)$ replacing Π in the application of Theorem 1, we get the following direct product theorem.

Theorem 6. *Given a function f and an arbitrary $\delta > 0$, for any function computation protocol π computing estimates $F_{\mathcal{X},n}$ and $F_{\mathcal{Y},n}$ of f^n at the Party 1 and Party 2, respectively, and with length*

$$\|\pi\| \leq n[H(F|X) + H(F|Y) - \delta], \tag{6}$$

there exists a constant $E = E(\delta) > 0$ such that for every n sufficiently large, it holds that

$$\Pr\left(F_{\mathcal{X},n} = F_{\mathcal{Y},n} = F^n\right) \leq 2^{-En},$$

where $F^n := (F_1, ..., F_n)$ and $F_i := f(X_i, Y_i)$, $1 \leq i \leq n$.

Recall that [BR11, MI11] showed that the first order asymptotic term in the amortized communication complexity for function computation was shown to equal the information complexity $\mathtt{IC}(f)$ of the function, namely the infimum over $\mathtt{IC}(\pi)$ for all interactive protocols π that recover f with 0 error. Ideally, we would like to show an Arimoto converse for this problem, *i.e.*, replace the threshold on the right-side of (6) with $n[\mathtt{IC}(f) - \delta]$. The direct product result above is weaker than such an Arimoto converse, and proving the Arimoto converse for the function computation problem is work in progress. Nevertheless, the simple result above is not comparable with the known direct product theorems in [BRWY13, BW14] and can be stronger in some regimes[11].

3.4 General formula for amortized communication complexity

Consider arbitrary distributions $\mathrm{P}_{X_nY_n}$ on $\mathcal{X}^n \times \mathcal{Y}^n$ and arbitrary protocols π_n with inputs X_n and Y_n taking values in $\mathcal{X}^n$ and $\mathcal{Y}^n$, for each $n \in \mathbb{N}$. For vanishing simulation error ε_n, how does $D_{\varepsilon_n}(\pi_n|\mathrm{P}_{X_nY_n})$ evolve as a function of n?

The previous section, and much of the theoretical computer science literature, has focused on the case when $\mathrm{P}_{X_nY_n} = \mathrm{P}_{XY}^n$ and the same protocol π is executed on each coordinate. In this section, we identify the first-order asymptotic term in $D_{\varepsilon_n}(\pi_n|\mathrm{P}_{X_nY_n})$ for a general sequence of distributions[12] $\{\mathrm{P}_{X_nY_n}\}_{n=1}^\infty$ and a general sequence of protocols $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^\infty$. Formally, the amortized (distributional) communication complexity of $\boldsymbol{\pi}$ for $\{\mathrm{P}_{X_nY_n}\}_{n=1}^\infty$ is given by[13]

$$D(\boldsymbol{\pi}) \overset{\text{def}}{=} \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} D_\varepsilon(\pi_n|\mathrm{P}_{X_nY_n}).$$

[11] The result in [BRWY13, BW14] shows a direct product theorem when we communicate less than $n\mathtt{IC}(f)/\mathtt{poly}(\log n)$.

[12] We do not require $\mathrm{P}_{X_nY_n}$ to be even consistent.

[13] Although $D(\boldsymbol{\pi})$ also depends on $\{\mathrm{P}_{X_nY_n}\}_{n=1}^\infty$, we omit the dependency in our notation.

Our goal is to characterize $D(\boldsymbol{\pi})$ for any given sequences P_n and $\boldsymbol{\pi}$. We seek a general formula for $D(\boldsymbol{\pi})$ under minimal assumptions. Since we do not make any assumptions on the underlying distribution, we cannot use any measure concentration results. Instead, we take recourse to probability limits of information spectrums introduced by Han and Verdú in [HV93] for handling this situation (*cf.* [Han03]). Specifically, for a sequence of protocols $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^{\infty}$ and a sequence of observations $(\mathbf{X}, \mathbf{Y}) = \{(X_n, Y_n)\}_{n=1}^{\infty}$, the *sup information complexity* is defined as

$$\overline{\mathtt{IC}}(\boldsymbol{\pi}) \overset{\text{def}}{=} \inf \left\{ \alpha \mid \lim_{n\to\infty} \Pr\left(\frac{1}{n}\mathtt{ic}(\Pi_n; X_n, Y_n) > \alpha\right) = 0 \right\},$$

where, with a slight abuse of notation, Π_n is the transcript of protocol π_n for observations (X_n, Y_n). The result below shows that it is $n\overline{\mathtt{IC}}(\boldsymbol{\pi})$, and not $\mathtt{IC}(\pi_n)$, that determines the communication complexity in general.

Theorem 7. *For every sequence of protocols* $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^{\infty}$,

$$D(\boldsymbol{\pi}) = \overline{\mathtt{IC}}(\boldsymbol{\pi}).$$

The proof uses Theorem 1 and Theorem 2 with carefully chosen spectrum-slice sizes.

For the case when $\pi_n = \pi^n$ and $\mathrm{P}_{X_nY_n} = \mathrm{P}_{XY}^n$, it follows from the law of large numbers that $\overline{\mathtt{IC}}(\boldsymbol{\pi}) = \mathtt{IC}(\pi)$ and we recover the result of [BR11]. However, the utility of the general formula goes far beyond this simple amortized regime. Example 1 provides one such instance. In this case, $\overline{\mathtt{IC}}(\boldsymbol{\pi})$ can be easily shown to equal $\mathtt{IC}(\pi_{\mathtt{h}})$ for any bias of the coin Π_0.

4 Background: Secret Key Agreement and Data Exchange

Our proofs draw from various techniques in cryptography and information theory. In particular, we use our recent results on information theoretic secret key agreement and data exchange, which are reviewed in this section together with the requisite background.

4.1 Secret key agreement by public discussion

The problem of two party secret key agreement by public discussion was alluded to in [BBR88], but a proper formulation and an asymptotically optimal construction appeared first in [Mau93, AC93]. Consider two parties with the first and the second party, respectively, observing the random variable X and Y. Using an interactive protocol π and their local observations, the parties agree on a secret key. A random variable K constitutes a secret key if the two parties form estimates that agree with K with probability close to 1 and K is concealed, in effect, from an eavesdropper with access to the transcript Π and a side-information Z. Formally, let $K_{\mathcal{X}}$ and $K_{\mathcal{Y}}$, respectively, be recoverable by π for the first and the second party. Such random variables $K_{\mathcal{X}}$ and $K_{\mathcal{Y}}$ with common range $\mathcal{K}$ constitute an *ε-secret key* if the following condition is satisfied:

$$d_{\mathtt{var}}\left(\mathrm{P}_{K_{\mathcal{X}}K_{\mathcal{Y}}\Pi Z}, \mathrm{P}_{\mathtt{unif}}^{(2)} \times \mathrm{P}_{\Pi Z}\right) \le \varepsilon,$$

where

$$\mathrm{P}_{\mathtt{unif}}^{(2)}(k_{\mathcal{X}}, k_{\mathcal{Y}}) = \frac{\mathbb{1}(k_{\mathcal{X}} = k_{\mathcal{Y}})}{|\mathcal{K}|}.$$

The condition above ensures both reliable *recovery*, requiring $\Pr(K_{\mathcal{X}} \neq K_{\mathcal{Y}})$ to be small, and information theoretic *secrecy*, requiring the distribution of $K_{\mathcal{X}}$ (or $K_{\mathcal{Y}}$) to be almost independent of the eavesdropper's side information (Π, Z) and to be almost uniform. See [TW14a] for a discussion.

Definition 4. Given $0 \leq \varepsilon < 1$, the supremum over lengths $\log |\mathcal{K}|$ of an ε-secret key is denoted by $S_\varepsilon(X, Y|Z)$, and for the case when Z is constant by $S_\varepsilon(X, Y)$.

By its definition, $S_\varepsilon(X, Y|Z)$ has the following monotonicity property.

Lemma 8 (**Monotonicity**)**.** *For any deterministic protocol π,*

$$S_\varepsilon(X, Y|Z) \geq S_\varepsilon(X\Pi, Y\Pi|Z\Pi).$$

Furthermore, if $V_\mathcal{X}$ and $V_\mathcal{Y}$ can be recovered by π for the first and the second party, respectively, then

$$S_\varepsilon(X, Y|Z) \geq S_\varepsilon(XV_\mathcal{X}, V_\mathcal{Y}|Z\Pi).$$

The claim holds since the two parties can generate a secret key by first running π and then generating a secret key for the case when the first party observes (X, Π), the second party observes (Y, Π) and the eavesdropper observes (Z, Π). Similarly, the second inequality holds since the parties can ignore a portion of their observations and generate a secret key from $(X, V_\mathcal{X})$ and $(Y, V_\mathcal{Y})$.

4.1.1 Leftover hash lemma: A tool for generating secret keys

A key tool for generating secret keys is the *leftover hash lemma* [BBR88, ILL89, BBCM95, RW05, Ren05] which, given a random variable X and an l-bit eavesdropper's observation Z, allows us to extract roughly $H_{\min}(\mathrm{P}_X) - l$ bits of uniform bits, independent of Z. We shall use a slightly more general form. Given random variables X and Z, let

$$H_{\min}\left(\mathrm{P}_{XZ} \mid \mathrm{Q}_Z\right) \overset{\text{def}}{=} \sup_{x,z} -\log \frac{\mathrm{P}_{XZ}\left(x, z\right)}{\mathrm{Q}_Z\left(z\right)}.$$

We define[14] the *conditional min-entropy* of X given Z by

$$H_{\min}\left(\mathrm{P}_{XZ} \mid Z\right) \overset{\text{def}}{=} \sup_{\mathrm{Q}_Z : \mathtt{supp}(\mathrm{P}_Z) \subset \mathtt{supp}(\mathrm{Q}_Z)} H_{\min}\left(\mathrm{P}_{XZ} \mid \mathrm{Q}_Z\right).$$

Further, let $\mathcal{F}$ be a *2-universal family* of mappings $f : \mathcal{X} \to \mathcal{K}$, *i.e.*, for each $x' \neq x$, the family $\mathcal{F}$ satisfies

$$\frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \mathbb{1}(f(x) = f(x')) \leq \frac{1}{|\mathcal{K}|}.$$

Lemma 9 (**Leftover Hash**)**.** *Consider random variables X, Z and V taking values in countable sets $\mathcal{X}$, $\mathcal{Z}$, and a finite set $\mathcal{V}$, respectively. Let S be a random seed such that f_S is uniformly distributed over a 2-universal family $\mathcal{F}$. Then, for $K_S = f_S(X)$*

$$\mathbb{E}_S \left\{d_{\mathtt{var}}\left(\mathrm{P}_{K_S V Z}, \mathrm{P}_{\mathtt{unif}} \mathrm{P}_{VZ}\right)\right\} \leq \frac{1}{2}\sqrt{|\mathcal{K}||\mathcal{V}|2^{-H_{\min}(\mathrm{P}_{XZ}|Z)}},$$

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

The version above is a straightforward modification of the leftover hash lemma in, for instance, [Ren05] and can be derived in a similar manner.

As an application of the leftover hash lemma above, we get the following useful result.

Lemma 10. *Consider random variables X, Y, Z and V taking values in countable sets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and a finite set $\mathcal{V}$, respectively. Then,*

$$S_{2\varepsilon}(X, Y|ZV) \geq S_\varepsilon(X, Y|Z) - \log|\mathcal{V}| - 2\log(1/2\varepsilon).$$

The proof is relegated to Appendix B.

[14] There is no agreement over the definition conditional min-entropy; the form adopted here is convenient for our use.

4.1.2 Conditional independence testing upper bound for secret key lengths

Next, we recall the *conditional independence testing* upper bound for $S_\varepsilon(X,Y)$, which was established in [TW14a, TW14b]. In fact, the general upper bound in [TW14a, TW14b] is a single-shot upper bound on the secret key length for a multiparty secret key agreement problem with side information at the eavesdropper. Below, we recall a specialization of the general result for the two party case with no side information at the eavesdropper. In order to state the result, we need the following concept from binary hypothesis testing.

Consider a binary hypothesis testing problem with null hypothesis P and alternative hypothesis Q, where P and Q are distributions on the same alphabet $\mathcal{V}$. Upon observing a value $v \in \mathcal{V}$, the observer needs to decide if the value was generated by the distribution P or the distribution Q. To this end, the observer applies a stochastic test T, which is a conditional distribution on $\{0,1\}$ given an observation $v \in \mathcal{V}$. When $v \in \mathcal{V}$ is observed, the test T chooses the null hypothesis with probability $\mathrm{T}(0|v)$ and the alternative hypothesis with probability $T(1|v) = 1 - T(0|v)$. For $0 \le \varepsilon < 1$, denote by $\beta_\varepsilon(\mathrm{P},\mathrm{Q})$ the infimum of the probability of error of type II given that the probability of error of type I is less than ε, *i.e.*,

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

where

$$\begin{aligned}\mathrm{P}[\mathrm{T}] &= \sum_v \mathrm{P}(v)\mathrm{T}(0|v),\\ \mathrm{Q}[\mathrm{T}] &= \sum_v \mathrm{Q}(v)\mathrm{T}(0|v).\end{aligned}$$

The following upper bound for $S_\varepsilon(X,Y)$ was established in [TW14a, TW14b].

Theorem 11 (**Conditional independence testing bound**). *Given $0 \le \varepsilon < 1$, $0 < \eta < 1-\varepsilon$, the following bound holds:*

$$S_\varepsilon\left(X,Y\right) \le -\log \beta_{\varepsilon+\eta}\big(\mathrm{P}_{XY}, \mathrm{Q}_X\mathrm{Q}_Y\big) + 2\log(1/\eta),$$

for all distributions Q_X and Q_Y on $\mathcal{X}$ and $\mathcal{Y}$, respectively.

We close by noting a further upper bound for $\beta_\varepsilon(\mathrm{P},\mathrm{Q})$, which is easy to derive.

Lemma 12. *For every $0 \le \varepsilon < 1$ and λ,*

$$-\log \beta_\varepsilon(\mathrm{P},\mathrm{Q}) \le \lambda - \log\left(\mathrm{P}\left(\log\frac{\mathrm{P}\left(X\right)}{\mathrm{Q}\left(X\right)} < \lambda\right) - \varepsilon\right)_+,$$

where $(x)_+ = \max\{0,x\}$. As a corollary, we obtain the following upper bound for $S_\varepsilon(X,Y)$:

$$S_\varepsilon\left(X,Y\right) \le \lambda - \log\left(\Pr\left(\log\frac{\mathrm{P}_{XY}\left(X,Y\right)}{\mathrm{Q}_X\left(X\right)\mathrm{Q}_Y\left(Y\right)} < \lambda\right) - \varepsilon - \eta\right)_+ + 2\log(1/\eta),$$

for all distributions Q_X and Q_Y.

4.2 The data exchange problem

The next primitive that will be used in the reduction argument in our lower bound proof is a protocol for data exchange. The parties observing X and Y seek to know each other's data. What is the minimum length of interactive communication required? This basic problem, first studied in [OG84], is in effect a two-party extension of the classical Slepian-Wolf compression [SW73] (see [CN04] for a multiparty version). In a recent work [TVW15], we derived tight lower and upper bounds for the length of a protocol that, for a given distribution P_{XY}, will facilitate data exchange with probability of error less than ε. We only review the proposed protocol and its performance here; first, we formally define the data exchange problem.

Definition 5. For $0 \leq \varepsilon < 1$, a protocol π attains ε-*data exchange* if there exist $\hat{Y}$ and $\hat{X}$ which are recoverable by π for the first and the second party, respectively, and satisfy

$$\mathrm{P}(\hat{X} = X,\ \hat{Y} = Y) \geq 1 - \varepsilon.$$

Note that data exchange corresponds to simulating a (deterministic) interactive protocol π where $\Pi_1(X) = X$ and $\Pi_2 = Y$; attaining ε-data exchange is tantamount to ε-simulation of π. In fact, the specific protocol for data exchange proposed in [TVW15] can be recovered as a special case of our simulation protocol in Section 6. The next result paraphrases [TVW15, Theorem 2] and can also be recovered as a special case of Lemma 21.

We paraphrase the result form [TVW15] in a form that is more suited for our application here. The data exchange protocol proposed in [TVW15] relies on slicing the spectrum of $h(X|Y)$ (or $h(Y|X)$). Let $\mathcal{E}_{\mathtt{tail}}$ denote the tail event $h(X|Y) \notin [\lambda'_{\mathrm{min}}, \lambda'_{\mathrm{max}}]$. The protocol entails slicing the essential spectrum $[\lambda'_{\mathrm{min}}, \lambda'_{\mathrm{max}}]$ into N parts of length Δ each, *i.e.*,

$$N = \frac{\lambda'_{\mathrm{max}} - \lambda'_{\mathrm{min}}}{\Delta}.$$

Theorem 13 ([TVW15, Theorem 2], Lemma 21)**.** *Given $\Delta > 0, \xi > 0$, and N as above, there exists a deterministic protocol for ε-data exchange satisfying the following properties:*

(i) Denoting by $\mathcal{E}_{\mathtt{error}}$ the error event, it holds that

$$\mathrm{P}_{XY}\left(\mathcal{E}_{\mathtt{error}} \cap \{h(X \triangle Y) \leq \lambda\}\right) \leq \mathrm{P}_{XY}\left(\mathcal{E}_{\mathtt{tail}}\right) + N2^{-\xi},$$

which further yields that the probability of error ε is bounded above as

$$\varepsilon \leq \mathrm{P}_{XY}\left(h(X \triangle Y) > \lambda\right) + \mathrm{P}_{XY}\left(\mathcal{E}_{\mathtt{tail}}\right) + N2^{-\xi};$$

(ii) the protocol communicates no more than $\lambda + \Delta + N + \xi$ bits;

(iii) for every (X, Y) such that $\lambda'_{\mathrm{min}} < h(X|Y) < \lambda'_{\mathrm{max}}$, the transcript of the protocol can take no more than $2^{h(X\triangle Y)+\Delta+\xi}$ values.

Note that property (iii) above, though not explicitly stated in [TVW15, Theorem 2] or in the general Lemma 21 below, follows simply from the proofs of these results. It makes the subtle observation that while, for each (X, Y) such that $\lambda'_{\mathrm{min}} < h(X|Y) < \lambda'_{\mathrm{max}}$, $h\left(X \triangle Y\right) + \Delta + N + \xi$ bits are communicated to interactively generate the transcript, the number of (variable length) transcripts is no more than[15] $h\left(X \triangle Y\right) + \Delta + N + \xi$. Property (ii) above was crucial to establish

[15] The N-bit ACK-NACK feedback used in the protocol can be determined from the length of the transcript.

the communication complexity results of [TVW15]; property (iii) was not relevant in the context of that work. On the other hand, here we shall use the protocol of Theorem 13 in our reduction to secret key agreement in the next section and will treat the communication used in data exchange as eavesdropper's side information. As such, it suffices to bound the number of values taken by the transcript; the number of bits actually communicated in the interactive protocol is a loose upper bound on the former quantity.

It is perhaps interesting that our simulation protocol given in Section 6 is used both in our upper bound to compress a given protocol and in our lower bound to complete the reduction argument.

5 Proof of Lower Bound

As described in the introduction, our proof of Theorem 1 relies on generating a secret key for X and Y from a given ε-simulation $\pi_{\mathtt{sim}}$ of π. However, there are two caveats in the heuristic approach described in the introduction:

First, to extract secret keys from the generated common randomness we rely on the leftover hash lemma. In particular, the bits are extracted by applying a 2-universal hash family to the common randomness generated. However, the range-size of the hash family must be selected based on the min-entropy of the generated common randomness, which is not easy to estimate. To remedy this, we communicate more using a data-exchange protocol proposed in [TVW15] to make the collective observations (X, Y) available to both the parties; a good bound for the communication complexity of this protocol is available. The generated common randomness now includes (X, Y) for which the min-entropy can be easily bounded and the size of the aforementioned extracted secret key can be tracked. A similar *common randomness completion and decomposition* technique was introduced in [Tya13] to characterize a class of securely computable functions.

Second, our methodology described above requires bounds on various information densities in different directions. A direct application of this method will result in a gap equal to the effective length of various spectrums involved. To remedy this, we apply the methodology described above not to the original distribution P_{XY} but a conditional distribution $\mathrm{P}_{XY|\mathcal{E}}$ where the event $\mathcal{E}$ is an appropriately chosen event contained in single slices of various spectrums involved. Such a conditioning is allowed since we are interested in the worst-case communication complexity of the simulation protocol.

We now describe the proof of Theorem 1 in detail. To make the exposition clear, we have divided the proof into five steps.

Given a (private coin) protocol π, let $\pi_{\mathtt{sim}}$ be its ε-simulation and $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Y}}$ be the corresponding estimates of the transcript Π for Party 1 and Party 2, respectively.

5.1 From simulation to probability of error

We first use a coupling argument to replace the ε-simulation condition with an ε probability of error condition. Recall the maximal coupling lemma.

Lemma 14 (Maximal Coupling Lemma [Str65]). *For any two distributions* P *and* Q *on the same set, there exists a joint distribution* P_{XY} *with* $X \sim \mathrm{P}$ *and* $Y \sim \mathrm{Q}$ *such that*

$$\Pr\left(X \neq Y\right) = d_{\mathtt{var}}\left(\mathrm{P}, \mathrm{Q}\right).$$

Using the maximal coupling lemma, for each fixed x, y there exists a joint distribution $\mathrm{P}_{\Pi\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}|X=x,Y=y}$ such that

$$\Pr\left(\Pi = \Pi_{\mathcal{X}} = \Pi_{\mathcal{Y}} | X = x, Y = y\right) = 1 - d_{\mathtt{var}}\left(\mathrm{P}_{\Pi\Pi|X=x,Y=y}, \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}|X=x,Y=y}\right);$$

Consequently,

$$\begin{aligned}\Pr\left(\Pi = \Pi_{\mathcal{X}} = \Pi_{\mathcal{Y}}\right) &= 1 - \sum_{x,y} \mathrm{P}_{XY}\left(x,y\right) d_{\mathtt{var}}\left(\mathrm{P}_{\Pi\Pi|X=x,Y=y}, \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}|X=x,Y=y}\right) \\ &= 1 - d_{\mathtt{var}}\left(\mathrm{P}_{\Pi\Pi XY}, \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}\right) \\ &\geq 1 - \varepsilon. \end{aligned} \tag{7}$$

As pointed in footnote 9, we restrict ourselves to public coin protocols $\pi_{\mathtt{sim}}$ using shared public randomness U. For concreteness (and convenience of proof), we define the joint distribution for $(\Pi\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XYU)$ as

$$\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}\Pi XYU} = \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}\Pi XY}\mathrm{P}_{U|\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}. \tag{8}$$

Note that the marginal $\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XYU}$ remains as in the original protocol. In particular, (X,Y) is jointly independent of U.

5.2 From partial knowledge to omniscience

As explained in the heuristic proof above, instead of extracting a secret key from the common randomness generated by the protocol $\pi_{\mathtt{sim}}$, we first use the data exchange protocol of Theorem 13 to make all the data available to both the parties, which was termed *attaining omniscience*[16] in [CN04]. In particular, the parties run the protocol $\pi_{\mathtt{sim}}$ followed by a data exchange protocol for $(X\Pi, Y\Pi)$ to recover (X,Y) at both the parties. Once both the parties have access to (X,Y), they can extract a secret key from (X,Y) which will be used in the reduction in our final step.

Formally, with the notations introduced in Section 4.2, let $\pi_{\mathtt{DE}}$ be the data exchange protocol of Theorem 13 with X and Y replaced by $(X\Pi)$ and $(Y\Pi)$, respectively, with N_2 and Δ_2 denoting N and Δ, respectively, and with $\lambda = \lambda_{\max}^{(3)}$, $\lambda'_{\min} = \lambda_{\min}^{(2)}$, $\lambda'_{\max} = \lambda_{\max}^{(2)}$. Then, denoting by $\mathcal{E}_{\mathtt{error}}$ the error event for the protocol $\pi_{\mathtt{DE}}$ Theorem 13(i) yields

$$\Pr\left(\mathcal{E}_{\mathtt{error}} \cap \mathcal{E}_3^c\right) \leq \Pr\left(\mathcal{E}_2\right) + N_2 2^{-\xi}, \tag{9}$$

where $\mathcal{E}_2$ and $\mathcal{E}_3$ are as in (3). Furthermore, for every realization $(X,Y) \notin \mathcal{E}_3$ the number possible transcripts $\Pi_{\mathtt{DE}}$ is no more than

$$2^{h(X\Pi \triangle Y\Pi)+\Delta_2+\xi}. \tag{10}$$

We seek to use $\pi_{\mathtt{DE}}$ for recovering Y and X, respectively, at Party 1 and Party 2 by running $\pi_{\mathtt{DE}}$ successively after $\pi_{\mathtt{sim}}$. However, $\pi_{\mathtt{sim}}$ yields $X\Pi_{\mathcal{X}}$ and $Y\Pi_{\mathcal{Y}}$ at Party 1 and Party 2, respectively, while the data exchange protocol $\pi_{\mathtt{DE}}$ facilitates data exchange when the two parties observe $X\Pi$ and $Y\Pi$. We can easily fix this gap using (7).

Specifically, denote by $\hat{X}$ and $\hat{Y}$ the estimates of X and Y formed at Party 2 and Party 1 in $\pi_{\mathtt{DE}}$. Note that $\pi_{\mathtt{DE}}$ is a deterministic protocol and $\hat{X}$ and $\hat{Y}$ are functions of (X, Y, Π, Π). Denote by $\mathcal{A}$ the set

$$\mathcal{A} = \{(\tau_{\mathcal{X}}, \tau_{\mathcal{Y}}, \tau, x, y) : \tau_{\mathcal{X}} = \tau_{\mathcal{Y}} = \tau\}$$

and by $\mathcal{B}$ the set

$$\mathcal{B} = \{(\tau_{\mathcal{X}}, \tau_{\mathcal{Y}}, \tau, x, y) : \hat{X}(x, y, \tau, \tau) = x, \hat{Y}(x, y, \tau, \tau) = y\},$$

[16] Csiszár and Narayan considered a multiterminal version of the data exchange problem in [CN04] and connected the minimum (amortized) rate of communication needed to the maximum (amortized) secret key rate.

which is the same as $\mathcal{E}^c_{\texttt{error}}$ for $\mathcal{E}_{\texttt{error}}$ in (9). Then, by (7) and (9)

$$\begin{aligned}
&\Pr\left(\{\hat{X}(X, Y, \Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}}) = X, \hat{Y}(X, Y, \Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}}) = Y\} \cap \mathcal{E}_3^c\right) \\
&\geq \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}\Pi XY}\left(\mathcal{A} \cap \mathcal{B} \cap \mathcal{E}_3^c\right) \\
&\geq \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}\Pi XY}\left(\mathcal{A}\right) + \Pr\left(\mathcal{E}_3^c\right) - \mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}\Pi XY}\left(\mathcal{B}^c \cap \mathcal{E}_3^c\right) - 1 \\
&\geq 1 - \varepsilon - \Pr\left(\mathcal{E}_2\right) - \Pr\left(\mathcal{E}_3\right) - N_2 2^{-\xi}.
\end{aligned} \tag{11}$$

5.3 From simulation to secret keys: A rough sketch of the reduction

The first step in our proof is to replace the simulation condition (1) with the probability of error condition (7) for the joint distribution $\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}\Pi XYU}$ in (8).

Next, we "complete the common randomness," *i.e.*, we communicate more to facilitate the recovery of Y and X at Party 1 and Party 2, respectively. To that end, upon executing $\pi_{\texttt{sim}}$, the parties run the data exchange protocol $\pi_{\texttt{DE}}$ of Theorem 13 for $(X\Pi)$ and $(Y\Pi)$, with $(X, \Pi_{\mathcal{X}})$ and $(Y, \Pi_{\mathcal{Y}})$ in place of $(X\Pi)$ and $(Y\Pi)$, respectively. Condition (7) guarantees that the combined protocol $(\pi_{\texttt{sim}}, \pi_{\texttt{DE}})$ recovers Y and X at Party 1 and Party 2 with probability of error less than ε.

We now sketch our reduction argument. Consider the secret key agreement for X and Y when the eavesdropper observes U. By the independence of (X, Y) and U, $S_\eta(XU, YU|U) = S_\eta(X, Y)$, and further, the result of [TW14a] shows that $S_\eta(X, Y)$ is bounded above, roughly, by the *mutual information density* $i(X \wedge Y) = \log \mathrm{P}_{XY}(X, Y) / \mathrm{P}_X(X) \mathrm{P}_Y(Y)$, *i.e.*,

$$S_\eta(XU, YU|U) = S_\eta(X, Y) \lesssim i(X \wedge Y). \tag{12}$$

On the other hand, we can generate a secret key using the following protocol:

1. Run the combined protocol $(\pi_{\texttt{sim}}, \pi_{\texttt{DE}})$ to attain data exchange for X and Y, resulting in a common randomness of size roughly $h(X, Y|U) = h(X, Y)$.

2. The data exchange protocol $\pi_{\texttt{DE}}$ for $(X\Pi)$ and $(Y\Pi)$ communicates roughly $h\left(X\Pi \triangle Y\Pi\right)$ bits for every fixed realization (X, Y, Π). Thus, the combined protocol $(\pi_{\texttt{sim}}, \pi_{\texttt{DE}})$, which allows both the parties to recover (X, Y), communicates no more than $\|\pi_{\texttt{sim}}\| + h\left(X\Pi \triangle Y\Pi\right)$ bits for every fixed realization (X, Y, Π). Using the leftover hash lemma, we can extract a secret key of rate roughly $h(X, Y) - \|\pi_{\texttt{sim}}\| - h\left(X\Pi \triangle Y\Pi\right)$.

The following approximate inequalities summarize our reduction:

$$\begin{aligned}
S_\eta(XU, YU|U) &\geq S_\eta(X\hat{Y}, \hat{X}Y|\Pi_{\texttt{sim}}\Pi_{\texttt{DE}}U) \\
&\gtrsim S_\eta(X\hat{Y}, \hat{X}Y|U) - \|\pi_{\texttt{sim}}\| - h\left(X\Pi \triangle Y\Pi\right) \\
&\approx h(X, Y) - \|\pi_{\texttt{sim}}\| - h\left(X\Pi \triangle Y\Pi\right),
\end{aligned} \tag{13}$$

where the first inequality is by Lemma 8 and the the second by Lemma 9.

We note that the generation of secret keys from data exchange was first proposed in [CN04] in an amortized, IID setup and was shown to yield a secret key of asymptotically optimal rate.

From (12) and (13) it follows that

$$\|\pi_{\texttt{sim}}\| \gtrsim h(X, Y) - h\left(X\Pi \triangle Y\Pi\right) - i(X \wedge Y) = \texttt{ic}(\Pi; X, Y),$$

which is the required lower bound.

Clearly, the steps above are not precise. We have used instantaneous communication and common randomness lengths in our bounds whereas a formal treatment will require us to use worst-case performance bounds for these quantities. Unfortunately, such worst-case bounds do not yield our desired lower bound for $D_\varepsilon(\pi)$. To fill this gap, we apply the arguments above not for the original distribution $\mathrm{P}_{\Pi_\mathcal{X}\Pi_\mathcal{Y}\Pi XYU}$ but for the conditional distribution $\mathrm{P}_{\Pi_\mathcal{X}\Pi_\mathcal{Y}\Pi XYU|\mathcal{E}}$ where the event $\mathcal{E}$ is carefully constructed in such a manner that the aforementioned worst-case bounds are close to instantaneous bounds for all realizations. Specifically, $\mathcal{E}$ is selected by appropriately slicing the spectrums of the various information densities that appear in the worst-case bounds.

5.4 From original to conditional probabilities

To identify an appropriate critical event for conditioning, we take recourse to spectrum slicing. Specifically, we identify an appropriate subset of intersection of slices of spectrums (ii) and (iv) described in Section 3.1. For the combined protocol $(\pi_{\mathtt{sim}}, \pi_{\mathtt{DE}})$ and the estimates $(\hat{X}, \hat{Y})$ as above, and $\lambda_{\min}^{(i)}, \lambda_{\max}^{(i)}$, $i = 1, 2, 3$, as in Section 3.1, let

$$\begin{aligned}
\mathcal{E}_{\mathtt{sim}} &= \{\Pi = \Pi_\mathcal{X} = \Pi_\mathcal{Y}\},\\
\mathcal{E}_{\mathtt{DE}} &= \{\hat{X}(X, Y, \Pi_\mathcal{X}, \Pi_\mathcal{Y}) = X,\ \hat{Y}(X, Y, \Pi_\mathcal{X}, \Pi_\mathcal{Y}) = Y\},\\
\mathcal{E}_\lambda &= \{\mathtt{ic}(\Pi; X, Y) \geq \lambda\}\\
\mathcal{E}_i^{(1)} &= \{\lambda_{\min}^{(1)} + (i-1)\Delta_1 \leq h(X, Y) \leq \lambda_{\min}^{(1)} + i\Delta_1\}, \quad 1 \leq i \leq N_1,\\
\mathcal{E}_j^{(3)} &= \{\lambda_{\min}^{(3)} + (j-1)\Delta_3 \leq h\left(X\Pi \triangle Y\Pi\right) \leq \lambda_{\min}^{(3)} + j\Delta_3\}, \quad 1 \leq j \leq N_3,
\end{aligned}$$

where

$$N_1 = \frac{\lambda_{\max}^{(1)} - \lambda_{\min}^{(1)}}{\Delta_1} \text{ and } N_3 = \frac{\lambda_{\max}^{(3)} - \lambda_{\min}^{(3)}}{\Delta_3}.$$

Note that $\cup_i \mathcal{E}_i^{(1)} = \mathcal{E}_1^c$ and $\cup_j \mathcal{E}_j^{(3)} = \mathcal{E}_3^c$, where the events $\mathcal{E}_1$ and $\mathcal{E}_3$ are as in (3). Finally, define the event $\mathcal{E}_{ij}$ as follows:

$$\mathcal{E}_{ij} = \mathcal{E}_{\mathtt{sim}} \cap \mathcal{E}_{\mathtt{DE}} \cap \mathcal{E}_\lambda \cap \mathcal{E}_i^{(1)} \cap \mathcal{E}_j^{(3)}, \quad 1 \leq i \leq N_1, 1 \leq j \leq N_3.$$

The next lemma says that (at least) one of the events $\mathcal{E}_{ij}$ has significant probability, and this particular event will be used as the critical event in our proofs.

Lemma 15. *There exists i, j such that*

$$\Pr\left(\mathcal{E}_{ij}\right) \geq \frac{\Pr\left(\mathcal{E}_\lambda\right) - \varepsilon - \varepsilon_{\mathtt{tail}} - N_2 2^{-\xi}}{N_1 N_3} \overset{\text{def}}{=} \alpha. \tag{14}$$

Proof. Note that the event $\mathcal{E}_{\mathtt{sim}} \cap \mathcal{E}_{\mathtt{DE}} \cap \mathcal{E}_3^c$ is the same as the event $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E}_3^c$ of (11). Therefore,

$$\begin{aligned}
\Pr\left(\mathcal{E}_{\mathtt{sim}} \cap \mathcal{E}_{\mathtt{DE}} \cap \mathcal{E}_\lambda \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\right) &\geq \Pr\left(\mathcal{E}_\lambda\right) + \Pr\left(\mathcal{E}_{\mathtt{sim}} \cap \mathcal{E}_{\mathtt{DE}} \cap \mathcal{E}_3^c\right) + \Pr\left(\mathcal{E}_1^c\right) - 2\\
&\geq \Pr\left(\mathcal{E}_\lambda\right) - \varepsilon - \Pr\left(\mathcal{E}_2\right) - \Pr\left(\mathcal{E}_3\right) - N_2 2^{-\xi} - \Pr\left(\mathcal{E}_1\right)\\
&\geq \Pr\left(\mathcal{E}_\lambda\right) - \varepsilon - \varepsilon_{\mathtt{tail}} - N_2 2^{-\xi},
\end{aligned}$$

where the second inequality uses (11) and and the third uses (3). The proof is completed upon noting that $\{\mathcal{E}_{ij}\}_{i,j}$ constitutes a partition of $\mathcal{E}_{\mathtt{sim}} \cap \mathcal{E}_{\mathtt{DE}} \cap \mathcal{E}_\lambda \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c$ with $N_1 N_3$ parts. ■

5.5 From simulation to secret keys: The formal reduction proof

We are now in a position to complete the proof of our lower bound. For brevity, let $\mathcal{E}$ denote the event $\mathcal{E}_{ij}$ of Lemma 15 satisfying $\Pr(\mathcal{E}) \geq \alpha$.

Our proof essentially formalizes the steps outlined in Section 5.3, but for the conditional distribution given $\mathcal{E}$. With an abuse of notation, let $S_\eta(X, Y|Z, \mathcal{E})$ denote the maximum length of an η-secret key for two parties observing X and Y, and the eavesdropper's side information Z, when the distribution of (X, Y, Z) is given by $\mathrm{P}_{XYZ|\mathcal{E}}$. Then, using Lemma 12 with $\mathrm{Q}_X = \mathrm{P}_X$ and $\mathrm{Q}_Y = \mathrm{P}_Y$, we get the following bound in place of (12):

$$\begin{aligned} S_{2\eta}(X, Y|\mathcal{E}) &\leq \gamma - \log\left(\Pr\left(\left\{(x,y) : \log\frac{\mathrm{P}_{XY|\mathcal{E}}(x,y)}{\mathrm{P}_X(x)\,\mathrm{P}_Y(y)} < \gamma\right\} \,\middle|\, \mathcal{E}\right) - 3\eta\right)_+ + 2\log(1/\eta) \\ &\leq \gamma - \log\left(\Pr\left(\left\{(x,y) : \log\frac{\mathrm{P}_{XY}(x,y)}{\mathrm{P}_X(x)\,\mathrm{P}_Y(y)} < \gamma + \log\alpha\right\} \,\middle|\, \mathcal{E}\right) - 3\eta\right)_+ + 2\log(1/\eta), \end{aligned} \tag{15}$$

where $0 < \eta < 1/3$ is arbitrary and in the previous inequality we have used

$$\mathrm{P}_{XY|\mathcal{E}}(x, y|\mathcal{E}) \leq \frac{\mathrm{P}_{XY}(x,y)}{\Pr(\mathcal{E})} \leq \frac{\mathrm{P}_{XY}(x,y)}{\alpha}.$$

To replace (13), note tha by Lemma 8

$$\begin{aligned} S_{2\eta}(X, Y|\mathcal{E}) &\geq S_{2\eta}(X\Pi_{\mathtt{sim}}\Pi_{\mathtt{DE}}, Y\Pi_{\mathtt{sim}}\Pi_{\mathtt{DE}}|U, \Pi_{\mathtt{sim}}, \Pi_{\mathtt{DE}}, \mathcal{E}) \\ &\geq S_{2\eta}(X\hat{Y}, \hat{X}Y|U, \Pi_{\mathtt{sim}}, \Pi_{\mathtt{DE}}, \mathcal{E}). \end{aligned} \tag{16}$$

Next, note that by (10) the transcript $\Pi_{\mathtt{sim}}\Pi_{\mathtt{DE}}$ takes no more than $2^{\|\pi_{\mathtt{sim}}\| + h(X\Pi \triangle Y\Pi) + \Delta_2 + \xi}$ values for every realization $(X, Y) \notin \mathcal{E}_3$. However, when the event $\mathcal{E} = \mathcal{E}_{ij}$ holds, $h(X\Pi \triangle Y\Pi) \leq \lambda_{\min}^{(3)} + j\Delta_3$. It follows by Lemma 22 that

$$\begin{aligned} &S_{2\eta}(X\hat{Y}, \hat{X}Y|U\Pi_{\mathtt{sim}}\Pi_{\mathtt{DE}}, \mathcal{E}) \\ &\geq S_\eta(X\hat{Y}, \hat{X}Y|U, \mathcal{E}) - \|\pi_{\mathtt{sim}}\| - \lambda_{\min}^{(3)} - j\Delta_3 - \Delta_2 - \xi - 2\log(1/2\eta). \end{aligned} \tag{17}$$

Also, since $\{X = \hat{X}, Y = \hat{Y}\}$ holds when we condition on $\mathcal{E}$,

$$\begin{aligned} S_\eta(X\hat{Y}, \hat{X}Y|U, \mathcal{E}) &= S_\eta(XY, XY|U, \mathcal{E}) \\ &\geq H_{\min}(\mathrm{P}_{XYU|\mathcal{E}} \mid U) - 2\log(1/2\eta), \end{aligned} \tag{18}$$

where the previous inequality is by the leftover hash lemma. Furthermore, by using

$$\mathrm{P}_{XYU|\mathcal{E}}(x, y, u) \leq \frac{\mathrm{P}_{XYU}(x,y,u)}{\Pr(\mathcal{E})} \leq \frac{\mathrm{P}_{XYU}(x,y,u)}{\alpha}$$

we can bound $H_{\min}(\mathrm{P}_{XYU|\mathcal{E}} \mid U)$ as follows:

$$\begin{aligned} H_{\min}(\mathrm{P}_{XYU|\mathcal{E}} \mid U) &\geq \min_{x,y,u} -\log\frac{\mathrm{P}_{XYU|\mathcal{E}}(x,y,u)}{\mathrm{P}_U(u)} \\ &\geq \min_{x,y,u} -\log\frac{\mathrm{P}_{XYU}(x,y,u)\,\mathbb{1}(\mathrm{P}_{XYU|\mathcal{E}}(x,y,u) > 0)}{\alpha\mathrm{P}_U(u)} \end{aligned}$$

$$\begin{aligned} &= \min_{x,y\in\mathcal{E}_i^{(1)}} h_{\mathrm{P}_{XY}}(x,y) + \log\alpha \\ &\geq \lambda_{\text{min}}^{(1)} + (i-1)\Delta_1 + \log\alpha. \end{aligned} \tag{19}$$

Thus, on combining (16)-(19), we get

$$S_{2\eta}(X,Y|\mathcal{E}) \geq [\lambda_{\text{min}}^{(1)} + (i-1)\Delta_1 - \lambda_{\text{min}}^{(3)} - j\Delta_3 + \log\alpha] - \Delta_2 - \xi - 4\log(1/2\eta) - \|\pi_{\mathtt{sim}}\|. \tag{20}$$

To get a matching form of the upper bound (15) for $S_{2\eta}(X,Y|\mathcal{E})$, note that since[17]

$$-\mathtt{ic}_{\mathrm{P}_{\Pi XY}}(\tau;x,y) = i_{\mathrm{P}_{XY}}(x\wedge y) - h_{\mathrm{P}_{XY}}(x,y) + h_{\mathrm{P}_{\Pi XY}}((x,\tau)\triangle(y,\tau)),$$

and since under $\mathcal{E}$

$$\begin{aligned} h_{\mathrm{P}_{XY}}(x,y) &\leq \lambda_{\text{min}}^{(1)} + i\Delta_1, \\ h_{\mathrm{P}_{XY\Pi}}((x,\tau)\triangle(y,\tau)) &\geq \lambda_{\text{min}}^{(3)} + (j-1)\Delta_3, \end{aligned}$$

it holds that

$$\begin{aligned} &\Pr\left(\{(x,y) : i_{\mathrm{P}_{XY}}(x\wedge y) < \gamma + \log\alpha\} \,\middle|\, \mathcal{E}\right) \\ &\geq \Pr\left(\left\{(x,y,\tau) : -\mathtt{ic}_{\mathrm{P}_{XY\Pi}}(x,y,\tau) < \gamma - \lambda_{\text{min}}^{(1)} - i\Delta_1 + \lambda_{\text{min}}^{(3)} + (j-1)\Delta_3 + \log\alpha\right\} \,\middle|\, \mathcal{E}\right). \end{aligned}$$

On choosing

$$\gamma = -\lambda + \lambda_{\text{min}}^{(1)} + i\Delta_1 - \lambda_{\text{min}}^{(3)} - (j-1)\Delta_3 - \log\alpha,$$

it follows from (15) that

$$\begin{aligned} &S_{2\eta}(X,Y|\mathcal{E}) \\ &\leq -\lambda + [\lambda_{\text{min}}^{(1)} + i\Delta_1 - \lambda_{\text{min}}^{(3)} - (j-1)\Delta_3 - \log\alpha] - \log\left(\Pr\left(\mathcal{E}_\lambda \mid \mathcal{E}\right) - 3\eta\right)_+ + 2\log(1/\eta) \\ &\leq -\lambda + [\lambda_{\text{min}}^{(1)} + i\Delta_1 - \lambda_{\text{min}}^{(3)} - (j-1)\Delta_3 - \log\alpha] - \log(1-3\eta) + 2\log(1/\eta), \end{aligned} \tag{21}$$

where the equality holds since $\Pr\left(\mathcal{E}_\lambda \mid \mathcal{E}\right) = 1$.

Thus, by (20) and (21), we get

$$\begin{aligned} \|\pi_{\mathtt{sim}}\| &\geq \lambda + 2\log\alpha - \Delta_1 - \Delta_2 - \Delta_3 - \xi - 6\log(1/\eta) + \log(1-3\eta) + 4 \\ &= \lambda + 2\log(\Pr\left(\mathcal{E}_\lambda\right) - \varepsilon - \varepsilon_{\mathtt{tail}} - \eta) - 2\log N_1N_3 - (\Delta_1 + \Delta_2 + \Delta_3) - \log N_2 \\ &\qquad - 7\log(1/\eta) + \log(1-3\eta) + 4. \end{aligned}$$

where the equality holds for $\xi = -\log\eta + \log N_2$. Note that the maximum value of the right-side above, when maximized over N_i and Δ_i under the constraint $N_i\Delta_i = \Lambda_i$, $i = 1,2,3$, occurs for $\Delta_1 = \Delta_3 = 2$ and $\Delta_2 = 1$. Substituting this choice of parameters, we get

$$\begin{aligned} \|\pi_{\mathtt{sim}}\| &\geq \lambda + 2\log(\Pr\left(\mathcal{E}_\lambda\right) - \varepsilon - \varepsilon_{\mathtt{tail}} - \eta) - 2\log\Lambda_1\Lambda_3 - \log\Lambda_2 - 7\log(1/\eta) + \log(1-3\eta) + 3. \\ &\geq \lambda - 2\log\Lambda_1\Lambda_3 - \log\Lambda_2 - 9\log(1/\eta) + \log(1-3\eta) + 3. \end{aligned}$$

where the final inequality holds for every λ such that $\Pr\left(\mathcal{E}_\lambda\right) \geq \varepsilon + \varepsilon_{\mathtt{tail}} + 2\eta$; Theorem 1 follows upon maximizing the right side-over all such λ. ■

[17] For clarity, we display the dependence of each information density on the underlying distribution in the remainder of this section.

6 Simulation Protocol and the Upper Bound

In this section, we formally present an ε-simulation of a given interactive protocol π with bounded rounds. For clarity, we build the simulation protocol in steps.

6.1 Sending X using one-sided communication

We start with the well-known Slepian-Wolf compression problem [SW73] where Party 1 wants to transmit X itself to Party 2 using as few bits as possible. This corresponds to simulating the deterministic protocol $\Pi = \Pi_1 = X$. See Remark 1 in Section 2 for a discussion on simulation of deterministic protocols.

For encoder, we use a hash function that is randomly chosen from a 2-universal hash family $\mathcal{F}_l(\mathcal{X})$; for decoder, we use a kind of joint typical decoder [CT06]. Let the *typical set* $\mathcal{T}_{\mathrm{P}_{X|Y}}$ be given by

$$\mathcal{T}_{\mathrm{P}_{X|Y}} = \left\{(x,y) : h_{\mathrm{P}_{X|Y}}(x|y) \leq l - \gamma\right\} \tag{22}$$

for a slack parameter $\gamma > 0$. Our first protocol is given below:

Protocol 1: Slepian-Wolf compression

```
**Input**: Observations $X$ and $Y$, uniform public randomness $U_{\mathsf{hash}}$, and a parameter $l$
**Output**: Estimate $\hat{X}$ of $X$ at party 2
Both parties use $U_{\mathsf{hash}}$ to select $f$ from $\mathcal{F}_l(\mathcal{X})$
Party 1 sends $\Pi_{\mathsf{sim},1} = f(X)$
**if** *Party 2 finds a unique* $x \in \mathcal{T}_{\mathrm{P}_{X|Y}}$ *with hash value* $f(x) = \Pi_{\mathsf{sim},1}$ **then**
    set $\hat{X} = x$
**else**
    protocol declares an error
```

The following result is from [MK95], [Han03, Lemma 7.2.1] (see, also, [Kuz12]).

Lemma 16 (Performance of Protocol 1). *For every $\gamma > 0$, the protocol above satisfies*

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

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 $h_{\mathrm{P}_{X|Y}}(X|Y)$.

In fact, the use of the typical set in (22) is not crucial in Protocol 1 and its performance analysis: For a given measure Q_{XY}, we can define another typical set $\mathcal{T}_{\mathrm{Q}_{X|Y}}$ by replacing $h_{\mathrm{P}_{X|Y}}(x|y)$ with $h_{\mathrm{Q}_{X|Y}}(x|y)$ in (22) even though the underlying distribution of (X,Y) is P_{XY}. Then, the error probability is bounded as

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

which implies that X can be sent by using roughly as many bits as the ε-tail of $h_{\mathrm{Q}_{X|Y}}(X|Y)$ under P_{XY}. This modification simplifies our performance analysis of the more involved protocols in the following sections.

6.2 Sending X using interactive communication

Protocol 1 aims at minimizing the worst-case communication length over all realization of (X, Y). However, our goal here is to simulate a multiround interactive protocol and as such, we need not account for the worst-case communication length in each round. Instead, we shall optimize the worst-case communication length for the combined interactive protocol. The protocol below is a modification of Protocol 1 and uses roughly $h(X|Y)$ bits for transmitting X instead of its ε-tail.

The new protocol proceeds as the previous one but relies on *spectrum-slicing* to adapt the length of communication to the specific realization of (X, Y): It increases the size of the hash output gradually, starting with $\lambda_1 = \lambda_{\min}$ and increasing the size Δ-bits at a time until either Party 2 decodes X or $\lambda_{\max}$ bits have been sent. After each transmission, Party 2 sends either an ACK-NACK feedback signal. The protocol stops when an ACK symbol is received.

Fix an auxiliary distribution Q_{XY}. For $\lambda^{\min}_{\mathrm{Q}_{X|Y}}, \lambda^{\max}_{\mathrm{Q}_{X|Y}}, \Delta_{\mathrm{Q}_{X|Y}} > 0$ with $\lambda^{\max}_{\mathrm{Q}_{X|Y}} > \lambda^{\min}_{\mathrm{Q}_{X|Y}}$, let

$$N_{\mathrm{Q}_{X|Y}} = \frac{\lambda^{\max}_{\mathrm{Q}_{X|Y}} - \lambda^{\min}_{\mathrm{Q}_{X|Y}}}{\Delta_{\mathrm{Q}_{X|Y}}},$$

and

$$\lambda^{(i)}_{\mathrm{Q}_{X|Y}} = \lambda^{\min}_{\mathrm{Q}_{X|Y}} + (i-1)\Delta_{\mathrm{Q}_{X|Y}}, \quad 1 \le i \le N_{\mathrm{Q}_{X|Y}}.$$

Further, let

$$\mathcal{T}^{(0)}_{\mathrm{Q}_{X|Y}} := \left\{(x, y) \mid h_{\mathrm{Q}_{X|Y}}(x|y) \ge \lambda^{\max}_{\mathrm{Q}_{X|Y}} \text{ or } h_{\mathrm{Q}_{X|Y}}(x|y) < \lambda^{\min}_{\mathrm{Q}_{X|Y}}\right\}, \tag{23}$$

and for $1 \le i \le N_{\mathrm{Q}_{X|Y}}$, let $\mathcal{T}^{(i)}_{\mathrm{Q}_{X|Y}}$ denote the ith slice of the spectrum given by

$$\mathcal{T}^{(i)}_{\mathrm{Q}_{X|Y}} = \left\{(x, y) \mid \lambda^{(i)}_{\mathrm{Q}_{X|Y}} \le h_{\mathrm{Q}_{X|Y}}(x|y) < \lambda^{(i)}_{\mathrm{Q}_{X|Y}} + \Delta_{\mathrm{Q}_{X|Y}}\right\}.$$

Note that $\mathcal{T}^{(0)}_{\mathrm{Q}_{X|Y}}$ corresponds to $\mathcal{T}^{c}_{\mathrm{Q}_{X|Y}}$ in the previous section and will be counted as an error event.

Our protocol is described in Protocol 2. For every $(x, y) \in \mathcal{T}^{(i)}_{\mathrm{Q}_{X|Y}}$, $1 \le i \le N_{\mathrm{Q}_{X|Y}}$, the following lemma provides a bound on the error.

Lemma 17 (**Performance of Protocol 2**). *For $(x, y) \in \mathcal{T}^{(i)}_{\mathrm{Q}_{X|Y}}$, $1 \le i \le N_{\mathrm{Q}_{X|Y}}$, 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 2 sends at most $(l + (i-1)\Delta_{\mathrm{Q}_{X|Y}} + i)$ bits and has probability of error bounded above as follows:*

$$\Pr\left(\hat{X} \ne x \mid X = x, Y = y\right) \le i 2^{\lambda^{\min}_{\mathrm{Q}_{X|Y}} + \Delta_{\mathrm{Q}_{X|Y}} - l}.$$

Proof. Since $(x, y) \in \mathcal{T}^{(i)}_{\mathrm{Q}_{X|Y}}$, an error occurs if there exists a $\hat{x} \ne x$ such that $(\hat{x}, y) \in \mathcal{T}^{(j)}_{\mathrm{Q}_{X|Y}}$ and $\Pi_{\mathsf{sim},2k-1} = f_{2k-1}(\hat{x})$ for $1 \le k \le j$ for some $j \le i$. Therefore, the probability of error is bounded above as

$$\Pr\left(\hat{X} \ne x \mid X = x, Y = y\right) \le \sum_{j=1}^{i} \sum_{\hat{x} \ne x} \Pr\left(f_{2k-1}(x) = f_{2k-1}(\hat{x}), \forall\, 1 \le k \le j\right) \mathbb{1}\left((\hat{x}, y) \in \mathcal{T}^{(j)}_{\mathrm{Q}_{X|Y}}\right)$$

Protocol 2: Interactive Slepian-Wolf compression

```
Input: Observations $X$ and $Y$ with distribution $\mathrm{P}_{XY}$, uniform public randomness $U_{\mathsf{hash}}$,
       auxiliary distribution $\mathrm{Q}_{XY}$, and parameters $\gamma$, $\lambda^{\min}_{\mathrm{Q}_{X|Y}}$, $\Delta_{\mathrm{Q}_{X|Y}}$, $N_{\mathrm{Q}_{X|Y}}$, and $l$
Output: Estimate $\hat{X}$ of $X$ at party 2
Both parties use $U_{\mathsf{hash}}$ to select $f_1$ from $\mathcal{F}_l(\mathcal{X})$
Party 1 sends $\Pi_{\mathsf{sim},1} = f_1(X)$
if Party 2 finds a unique $x \in \mathcal{T}^{(1)}_{\mathrm{Q}_{X|Y}}$ with hash value $f_1(x) = \Pi_{\mathsf{sim},1}$ then
    set $\hat{X} = x$
    send back $\Pi_{\mathsf{sim},2} = \mathrm{ACK}$
else
    send back $\Pi_{\mathsf{sim},2} = \mathrm{NACK}$
while $2 \leq i \leq N_{\mathrm{Q}_{X|Y}}$ and party 2 did not send an ACK do
    Both parties use $U_{\mathsf{hash}}$ to select $f_i$ from $\mathcal{F}_{\Delta_{\mathrm{Q}_{X|Y}}}(\mathcal{X})$, independent of $f_1, \ldots, f_{i-1}$
    Party 1 sends $\Pi_{\mathsf{sim},2i-1} = f_i(X)$
    if Party 2 finds a unique $x \in \mathcal{T}^{(i)}_{\mathrm{Q}_{X|Y}}$ with hash value $f_j(x) = \Pi_{\mathsf{sim},2j-1}$, $\forall\, 1 \leq j \leq i$ then
        set $\hat{X} = x$
        send back $\Pi_{\mathsf{sim},2i} = \mathrm{ACK}$
    else
        if More than one such $x$ found then
            protocol declares an error
        else
            send back $\Pi_{\mathsf{sim},2i} = \mathrm{NACK}$
    Reset $i \to i + 1$
if No $\hat{X}$ found at party 2 then
    Protocol declares an error
```

$$
\begin{aligned}
&\leq \sum_{j=1}^{i} \sum_{\hat{x} \neq x} \frac{1}{2^{l+(j-1)\Delta_{\mathrm{Q}_{X|Y}}}} \mathbb{1}\left((\hat{x}, y) \in \mathcal{T}^{(j)}_{\mathrm{Q}_{X|Y}}\right) \\
&= \sum_{j=1}^{i} \sum_{\hat{x} \neq x} \frac{1}{2^{l+(j-1)\Delta_{\mathrm{Q}_{X|Y}}}} \left|\left\{\hat{x} \mid (\hat{x}, y) \in \mathcal{T}^{(j)}_{\mathrm{Q}_{X|Y}}\right\}\right| \\
&\leq i 2^{\lambda^{\min}_{\mathrm{Q}_{X|Y}} + \Delta_{\mathrm{Q}_{X|Y}} - l},
\end{aligned}
$$

where the first inequality follows from the union bound, the second inequality follows from the property of 2-universal hash family, and the third inequality follows from the fact that

$$
|\{\hat{x} \mid (\hat{x}, y) \in \mathcal{T}^{(j)}_{\mathrm{Q}_{X|Y}}\}| \leq 2^{\lambda^{(j)}_{\mathrm{Q}_{X|Y}} + \Delta_{\mathrm{Q}_{X|Y}}}.
$$

Note that the protocol sends l bits in the first transmission, and $\Delta_{\mathrm{Q}_{X|Y}}$ bits and 1-bit feedback in every subsequence transmission. Therefore, no more than $(l + (i-1)\Delta_{\mathrm{Q}_{X|Y}} + i)$ bits are sent. ■

Corollary 18. *Protocol 2 with $l = \lambda^{\min}_{\mathrm{Q}_{X|Y}} + \Delta_{\mathrm{Q}_{X|Y}} + \gamma$ sends at most $(h_{\mathrm{Q}_{X|Y}}(X|Y) + \Delta_{\mathrm{Q}_{X|Y}} + \gamma +$*

$N_{\mathrm{Q}_{X|Y}}$) bits when the observations are[18] $(X, Y) \notin \mathcal{T}^{(0)}_{\mathrm{Q}_{X|Y}}$*, and has probability of error less than*

$$\Pr\left(\hat{X} \neq X\right) \leq \Pr\left((X, Y) \in \mathcal{T}^{(0)}_{\mathrm{Q}_{X|Y}}\right) + N_{\mathrm{Q}_{X|Y}} 2^{-\gamma}.$$

6.3 Simulation of Π_1 using interactive communication

We now proceed to simulating the first round of our given interactive protocol π. Note that using Protocol 2, we can send Π_1 using roughly $h(\Pi_1|Y)$ bits. This protocol uses a public randomness U_{hash} only to choose hash functions, which is convenient for our probability of error analysis, and can be easily derandomized. We now present a scheme which uses another independent portion of public randomness U_{sim} to reduce the rate of the communication further. However, the scheme will only allow the parties to simulate Π_1 (rather than recover it with small probability of error) and cannot be derandomized.

Specifically, our next protocol uses X and $U = (U_{\mathsf{hash}}, U_{\mathsf{sim}})$ to simulate Π_1 in such a manner that U_{sim} can be treated, in effect, as a portion of the communication used in Protocol 2. Note that since U_{sim} is independent of (X, Y), the portion of communication which is equivalent to U_{sim} must as well be almost independent of (X, Y). Such a portion can be guaranteed by noting that the communication used in Protocol 2 is simply a random hash of Π_1 drawn from a 2-universal family, and therefore, its appropriately small portion can have the desired independence property by the leftover hash lemma. In fact, since the Markov condition $\Pi_1 -\!\!\circ\!\!- X -\!\!\circ\!\!- Y$ holds, it suffices guarantee the independent of X instead of (X, Y).

Protocol 3: Simulation of Π_1

Input: Observations X and Y with distribution P_{XY}, uniform public randomness $U = (U_{\mathsf{hash}}, U_{\mathsf{sim}})$, auxiliary distribution $\mathrm{Q}_{\Pi_1 Y}$, and parameters γ, $\lambda^{\min}_{\mathrm{Q}_{\Pi_1|Y}}$, $\Delta_{\mathrm{Q}_{\Pi_1|Y}}$, $N_{\mathrm{Q}_{\Pi_1|Y}}$ and k

Output: Estimates $\Pi_{1\mathcal{X}}$ and $\Pi_{1\mathcal{Y}}$ of Π_1

1. Two parties share k random bits U_{sim} and an h chosen from $\mathcal{H}_k(\mathrm{supp}(\Pi_1))$ using U_{hash}

2. Party 1 generates a sample $\Pi_{1\mathcal{X}}$ using $\mathrm{P}_{\Pi_1|X f(\Pi_1)}\left(\cdot|X, U_{\mathsf{sim}}\right)$

3. Parties use Protocol 2 with auxiliary distribution $\mathrm{Q}_{\Pi_1 Y}$, and parameters γ, $\lambda^{\min}_{\mathrm{Q}_{\Pi_1|Y}}$, $\Delta_{\mathrm{Q}_{\Pi_1|Y}}$, $N_{\mathrm{Q}_{\Pi_1|Y}}$, and $l = \lambda^{\min}_{\mathrm{Q}_{\Pi_1|Y}} + \Delta_{\mathrm{Q}_{\Pi_1|Y}} + \gamma$ to send $\Pi_{1\mathcal{X}}$ to Party 2 by treating U_{sim} as the first k bits of communication obtained via the hash function f

Our simulation protocol is described in Protocol 3. Let the quantities such as $\lambda^{\min}_{\mathrm{Q}_{\Pi_1|Y}}, \Delta_{\mathrm{Q}_{\Pi_1|Y}}$, and $N_{\mathrm{Q}_{\Pi_1|Y}}$ be defined analogously to the corresponding quantities in Section 6.2 with Π_1 replacing X. The following lemma provides a bound on the simulation error for Protocol 3.

Lemma 19 (**Performance of Protocol 3**)**.** *Protocol 3 sends at most*

$$\left(h_{\mathrm{Q}_{\Pi_1|Y}}(\Pi_{1\mathcal{X}}|Y) + \Delta_{\mathrm{Q}_{\Pi_1|Y}} + N_{\mathrm{Q}_{\Pi_1|Y}} + \gamma - k\right)_+$$

bits when $(\Pi_{1\mathcal{X}}, Y) \notin \mathcal{T}^{(0)}_{\mathrm{Q}_{\Pi_1|Y}}$*, and has simulation error*

$$d_{\mathsf{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1 XY}\right) \leq \Pr\left((\Pi_1, Y) \in \mathcal{T}^{(0)}_{\mathrm{Q}_{\Pi_1|Y}}\right) + N_{\mathrm{Q}_{\Pi_1|Y}} 2^{-\gamma} + \frac{1}{2}\sqrt{2^{k - H_{\min}(\mathrm{P}_{\Pi_1 X}|\mathrm{Q}_X)}}$$

for any auxiliary distribution Q_X on $\mathcal{X}$.

[18] When $h_{\mathrm{Q}_{X|Y}}(X|Y) < \lambda^{\min}_{\mathrm{Q}_{X|Y}}$, Protocol 2 may transmit more than $(h_{\mathrm{Q}_{X|Y}}(X|Y) + \Delta_{\mathrm{Q}_{X|Y}} + \gamma + N_{\mathrm{Q}_{X|Y}})$ bits.

Proof. Consider the following simple protocol for simulating Π_1 at Party 2:

1. Party 1 generates a sample Π_1 using $\mathrm{P}_{\Pi_1|X}(\cdot|X)$.
2. Both parties use Protocol 2 with auxiliary distribution $\mathrm{Q}_{\Pi_1 Y}$, and parameters γ, $\lambda^{\min}_{\mathrm{Q}_{\Pi_1|Y}}$, $\Delta_{\mathrm{Q}_{\Pi_1|Y}}$, $N_{\mathrm{Q}_{\Pi_1|Y}}$, and $l = \lambda^{\min}_{\mathrm{Q}_{\Pi_1|Y}} + \Delta_{\mathrm{Q}_{\Pi_1|Y}} + \gamma$ to generate an estimate $\hat{\Pi}_1$ of Π_1 at Party 2.

In this protocol, $l_{\mathsf{wst}} = \lambda^{\min}_{\mathrm{Q}_{\Pi_1|Y}} + N_{\mathrm{Q}_{\Pi_1|Y}}\Delta_{\mathrm{Q}_{\Pi_1|Y}} + \gamma$ bits of hash values will be sent for the worst (Π_1, Y). We divide these l_{wst} hash values into two parts, the fist k bits and the last $l_{\mathsf{wst}} - k$ bits; let f and f', respectively, denote the hash function producing the first and the second parts. Protocol 3 replaces, in effect, f with shared randomness U_{sim} for an appropriately chosen value of k.

Note that the joint distribution of the random variables involved in the simple protocol above satisfies[19]

$$\begin{aligned}
&\mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\hat{\Pi}_1XY}(v, v', \tau, \hat{\tau}, x, y) \\
&= \mathrm{P}_{f(\Pi_1)X}(v, x)\mathrm{P}_{\Pi_1|Xf(\Pi_1)}(\tau|x, v)\mathrm{P}_{f'(\Pi_1)|\Pi_1}(v'|\tau)\mathrm{P}_{Y|X}(y|x)\mathrm{P}_{\hat{\Pi}_1|f(\Pi_1)f'(\Pi_1)\Pi_1XY}(\hat{\tau}|v, v', \tau, x, y).
\end{aligned} \tag{24}$$

Note that the simple protocol above is deterministic and therefore by Remark 1

$$\begin{aligned}
d_{\mathsf{var}}\left(\mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\hat{\Pi}_1XY}, \mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\Pi_1XY}\right) &= \Pr\left(\Pi_1 \neq \hat{\Pi}_1\right) \\
&\leq \Pr\left((\Pi_1, Y) \in \mathcal{T}^{(0)}_{\mathrm{Q}_{\Pi_1|Y}}\right) + N_{\mathrm{Q}_{\Pi_1|Y}}2^{-\gamma},
\end{aligned} \tag{25}$$

where the inequality is by Corollary 18.

On the other hand, the joint distribution of random variables involved in Protocol 3 can be factorized as

$$\begin{aligned}
&\mathrm{P}_{U_{\mathsf{sim}}f'(\Pi_{1\mathcal{X}})\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}(u, u', \tau, \hat{\tau}, x, y) \\
&= \mathrm{P}_{U_{\mathsf{sim}}}(u)\mathrm{P}_X(x)\mathrm{P}_{\Pi_1|Xf(\Pi_1)}(\tau|x, u)\mathrm{P}_{f'(\Pi_1)|\Pi_1}(u'|\tau)\mathrm{P}_{Y|X}(y|x)\mathrm{P}_{\hat{\Pi}_1|f(\Pi_1)f'(\Pi_1)\Pi_1XY}(\hat{\tau}|u, u', \tau, x, y).
\end{aligned} \tag{26}$$

Therefore, the simulation error for Protocol 3 is bounded as

$$\begin{aligned}
&d_{\mathsf{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1XY}\right) \\
&\leq d_{\mathsf{var}}\left(\mathrm{P}_{U_{\mathsf{sim}}f'(\Pi_1)\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}, \mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\Pi_1XY}\right) \\
&\leq d_{\mathsf{var}}\left(\mathrm{P}_{U_{\mathsf{sim}}f'(\Pi_1)\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}, \mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\hat{\Pi}_1XY}\right) + d_{\mathsf{var}}\left(\mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\hat{\Pi}_1XY}, \mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\Pi_1XY}\right) \\
&= d_{\mathsf{var}}\left(\mathrm{P}_{U_{\mathsf{sim}}}\mathrm{P}_X, \mathrm{P}_{f(\Pi_1)X}\right) + d_{\mathsf{var}}\left(\mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\hat{\Pi}_1XY}, \mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\Pi_1XY}\right) \\
&\leq d_{\mathsf{var}}\left(\mathrm{P}_{U_{\mathsf{sim}}}\mathrm{P}_X, \mathrm{P}_{f(\Pi_1)X}\right) + \Pr\left((\Pi_1, Y) \in \mathcal{T}^{(0)}_{\mathrm{Q}_{\Pi_1|Y}}\right) + N_{\mathrm{Q}_{\Pi_1|Y}}2^{-\gamma},
\end{aligned}$$

where the first inequality is by the monotonicity of $\|\cdot\|$, the second inequality is by the triangular inequality, the equality is by the fact that replacing $\mathrm{P}_{U_{\mathsf{sim}}}\mathrm{P}_X$ with $\mathrm{P}_{f(\Pi_1)X}$ is the only difference between the factorizations in (26) and (24), and the final inequality is by (25). The desired bound on simulation error for Protocol 3 follows by using Lemma 9 to get

$$d_{\mathsf{var}}\left(\mathrm{P}_{U_{\mathsf{sim}}}\mathrm{P}_X, \mathrm{P}_{f(\Pi_1)X}\right) \leq \frac{1}{2}\sqrt{2^{k - H_{\min}(\mathrm{P}_{\Pi_1X}|\mathrm{Q}_X)}}.$$

Since Protocol 3 uses shared randomness U_{sim} instead of sending $f(\Pi_1)$, it communicates k fewer bits in comparison with the simple protocol above, which completes the proof. ■

[19] When the protocol terminate before $N_{\mathrm{Q}_{\Pi_1|Y}}$th round, a part of $(f(\Pi_1), f'(\Pi_1))$ may not be sent.

6.4 Improved simulation of Π_1

In Protocol 3 we were able to reduce the communication by roughly $H_{\min}(\mathrm{P}_{\Pi_1 X}|\mathrm{Q}_X)$ bits by simulating a Π_1 such that if we use Protocol 2 for sending Π_1 to Party 2, a portion of the required communication can be treated as shared public randomness. However, this is the least reduction in communication we can obtain in the worst-case. In this section, we slice the spectrum of $h_{\mathrm{P}_{\Pi_1|X}}(\Pi_1|X)$ to obtain an instantaneous reduction of roughly $h_{\mathrm{P}_{\Pi_1|X}}(\Pi_1|X)$ bits.

Denote by J a random variable which takes the value $j \in \{0, 1, \dots, N_{\mathrm{P}_{\Pi_1|X}}\}$ if $(\Pi_1, X) \in \mathcal{T}^{(j)}_{\mathrm{P}_{\Pi_1|X}}$. In our modified protocol, Party 1 first samples J and sends it to Party 2. Then, they proceed with Protocol 3 for $\mathrm{P}_{\Pi_1 XY|J=j}$ by selecting k to be less than $H_{\min}(\mathrm{P}_{\Pi_1 X|J=j}|\mathrm{Q}_X)$ for an appropriately chosen Q_X. Let $\mathcal{J}_{\mathsf{g}}$ be the set of "good" indices $j > 0$ with

$$\mathrm{P}_J(j) \geq \frac{1}{N^2_{\mathrm{P}_{\Pi_1|X}}};$$

it holds that

$$\mathrm{P}_J\left(\mathcal{J}^c_{\mathsf{g}}\right) < \Pr\left((\Pi_1, X) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_1|X}}\right) + \frac{1}{N_{\mathrm{P}_{\Pi_1|X}}}.$$

Note that for $j \in \mathcal{J}_{\mathsf{g}}$, with $\mathrm{Q}_X = \mathrm{P}_X$, we have

$$\begin{aligned} H_{\min}(\mathrm{P}_{\Pi_1 X|J=j}|\mathrm{P}_X) &= \min_{\tau,x} -\log \frac{\mathrm{P}_{\Pi_1 X|J}(\tau, x|j)}{\mathrm{P}_X(x)} \\ &= \min_{\tau,x} -\log \frac{\mathrm{P}_{\Pi_1|X}(\tau|x)}{\mathrm{P}_J(j)} \\ &\geq \lambda^{\min}_{\mathrm{P}_{\Pi_1|X}} + (j-1)\Delta_{\mathrm{P}_{\Pi_1|X}} - 2\log N_{\mathrm{P}_{\Pi_1|X}}. \end{aligned}$$

Protocol 4: Improved simulation of Π_1

```
Input: Observations $X$ and $Y$ with distribution $\mathrm{P}_{XY}$, uniform public randomness
       $U = (U_{\mathsf{hash}}, U_{\mathsf{sim}})$, and parameters $\lambda^{\min}_{\mathrm{P}_{\Pi_1|Y}}$, $\Delta_{\mathrm{P}_{\Pi_1|Y}}$, $N_{\mathrm{P}_{\Pi_1|Y}}$, $\lambda^{\min}_{\mathrm{P}_{\Pi_1|X}}$, $\Delta_{\mathrm{P}_{\Pi_1|X}}$, $N_{\mathrm{P}_{\Pi_1|X}}$,
       and $\gamma$
Output: Estimates $\Pi_{1\mathcal{X}}$ and $\Pi_{1\mathcal{Y}}$ of $\Pi_1$
Party 1 generate $J \sim \mathrm{P}_{J|X}(\cdot|X)$, and send it to Party 2.
if $J = j \in \mathcal{J}_g$ then
    Parties use Protocol 3 with auxiliary distribution $\mathrm{P}_{\Pi_1 Y}$, parameters $\gamma$, $\lambda^{\min}_{\mathrm{P}_{\Pi_1|Y}}$, $\Delta_{\mathrm{P}_{\Pi_1|Y}}$,
    $N_{\mathrm{P}_{\Pi_1|Y}}$, and $k = \lambda^{\min}_{\mathrm{P}_{\Pi_1|X}} + (j-1)\Delta_{\mathrm{P}_{\Pi_1|X}} - 2\log N_{\mathrm{P}_{\Pi_1|X}} - 2\gamma + 2$ to simulate $\Pi_{1\mathcal{X}}$ and
    $\Pi_{1\mathcal{Y}}$ for the distribution $\mathrm{P}_{\Pi_1 XY|J=j}$
else
    protocol declares an error
```

Our modified simulation protocol is described in Protocol 4. The following lemma provides a bound on the simulation error.

Lemma 20 (**Performance of Protocol 4**)**.** *Protocol 4 sends at most*

$$\left(h_{\mathrm{P}_{\Pi_1|Y}}(\Pi_{1\mathcal{X}}|Y) - h_{\mathrm{P}_{\Pi_1|X}}(\Pi_{1\mathcal{X}}|X) + N_{\mathrm{P}_{\Pi_1|Y}} + 3\log N_{\mathrm{P}_{\Pi_1|X}} + \Delta_{\mathrm{P}_{\Pi_1|Y}} + \Delta_{\mathrm{P}_{\Pi_1|X}} + 3\gamma\right)_+$$

bits when $(\Pi_{1\mathcal{X}}, Y) \notin \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_1|Y}}$*, and has simulation error*

$$\begin{aligned}&d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1XY}\right)\\ &\leq \Pr\left((\Pi_1, Y) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_1|Y}}\right) + \Pr\left((\Pi_1, X) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_1|X}}\right) + \left(N_{\mathrm{P}_{\Pi_1|Y}} + 1\right)2^{-\gamma} + \frac{1}{N_{\mathrm{P}_{\Pi_1|X}}}.\end{aligned}$$

Proof. First, we have

$$\begin{aligned}&d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1XY}\right)\\ &\leq d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XYJ}, \mathrm{P}_{\Pi_1\Pi_1XYJ}\right)\\ &= \sum_j \mathrm{P}_J(j) d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY|J=j}, \mathrm{P}_{\Pi_1\Pi_1XY|J=j}\right)\\ &\leq \sum_{j\in\mathcal{J}_{\mathtt{g}}} \mathrm{P}_J(j) d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY|J=j}, \mathrm{P}_{\Pi_1\Pi_1XY|J=j}\right) + \mathrm{P}_J\left(\mathcal{J}_{\mathtt{g}}^c\right)\\ &\leq \sum_{j\in\mathcal{J}_{\mathtt{g}}} \mathrm{P}_J(j) d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY|J=j}, \mathrm{P}_{\Pi_1\Pi_1XY|J=j}\right) + \Pr\left((\Pi_1, X) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_1|X}}\right) + \frac{1}{N_{\mathrm{P}_{\Pi_1|X}}}.\end{aligned}$$

Then, we apply Lemma 19 with $\mathrm{Q}_X = \mathrm{P}_X$ for each $j \in \mathcal{J}_{\mathtt{g}}$, and get

$$\begin{aligned}&d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY|J=j}, \mathrm{P}_{\Pi_1\Pi_1XY|J=j}\right)\\ &\leq \Pr\left((\Pi_1, Y) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_1|Y}} \mid J = j\right) + N_{\mathrm{P}_{\Pi_1|Y}} 2^{-\gamma} + \frac{1}{2}\sqrt{2^{k - H_{\min}(\mathrm{P}_{\Pi_1X|J=j}|\mathrm{P}_X)}}\\ &\leq \Pr\left((\Pi_1, Y) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_1|Y}} \mid J = j\right) + \left(N_{\mathrm{P}_{\Pi_1|Y}} + 1\right)2^{-\gamma}. \qquad (27)\end{aligned}$$

Thus, we have the desired bound on simulation error.

Next, we prove the claimed bound on the number of bits sent by the protocol. By Lemma 19, the fact that J can be sent by using at most $\log N_{\mathrm{P}_{\Pi_1|X}} + 1$ bits and the choice of k in Protocol 4, for $J = j$ the protocol above communicates at most

$$\begin{aligned}&h_{\mathrm{Q}_{\Pi_1|Y}}(\Pi_{1\mathcal{X}}|Y) + \Delta_{\mathrm{Q}_{\Pi_1|Y}} + N_{\mathrm{Q}_{\Pi_1|Y}} + \gamma + \log N_{\mathrm{P}_{\Pi_1|X}} + 2 - k\\ &\leq h_{\mathrm{Q}_{\Pi_1|Y}}(\Pi_{1\mathcal{X}}|Y) - \lambda^{\min}_{\mathrm{P}_{\Pi_1|X}} - (j-1)\Delta_{\mathrm{P}_{\Pi_1|X}} + \Delta_{\mathrm{Q}_{\Pi_1|Y}} + N_{\mathrm{Q}_{\Pi_1|Y}} + 3\log N_{\mathrm{P}_{\Pi_1|X}} + 3\gamma.\\ &\leq h_{\mathrm{Q}_{\Pi_1|Y}}(\Pi_{1\mathcal{X}}|Y) - h_{\mathrm{P}_{\Pi_1|X}}(\Pi_{1\mathcal{X}}|X) + \Delta_{\mathrm{P}_{\Pi_1|X}} + \Delta_{\mathrm{Q}_{\Pi_1|Y}} + N_{\mathrm{Q}_{\Pi_1|Y}} + 3\log N_{\mathrm{P}_{\Pi_1|X}} + 3\gamma,\end{aligned}$$

where the previous inequality holds since for $\Pi_{1\mathcal{X}}$ generated by $\mathrm{P}_{\Pi_1|Xf(\Pi_1)J}(\cdot|X, U_{\mathsf{sim}}, j)$

$$\lambda^{\min}_{\mathrm{P}_{\Pi_1|X}} + j\Delta_{\mathrm{P}_{\Pi_1|X}} \geq h_{\mathrm{P}_{\Pi_1|X}}(\Pi_{1\mathcal{X}}|X),$$

for each $j \in \mathcal{J}_{\mathtt{g}}$. ■

6.5 Simulation of Π

We are now in a position to describe our complete simulation protocol. Consider an interactive protocol π with maximum number of rounds $r_max = d < \infty$. We simply apply Protocol 4 for each round Π_t of Π. Our overall simulation protocol is described in Protocol 5. In each round we use Protocol 4 assuming that the simulation up to the previous round has succeeded, where, for the rounds with even numbers, we use Protocol 4 by interchanging the role of Party 1 and Party 2.

The following lemma provides a bound on the simulation error.

Protocol 5: Simulation of Π

```
Input: Observations $X$ and $Y$ with distribution $\mathrm{P}_{XY}$, uniform public randomness
       $U = (U_{t,\mathsf{hash}}, U_{t,\mathsf{sim}} : t = 1, \ldots, d)$, and parameters $\lambda^{\min}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$, $\Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$,
       $N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$, $\lambda^{\min}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$, $\Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$, $N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$ for $t = 1, \ldots, d$ and $\gamma$.
Output: Estimates $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Y}}$ of $\Pi$
while Total communication is less than $l_{\max}$ bits, and simulation not ended do
    Party 1 and Party 2, respectively, use estimates $\Pi^{t-1}_{\mathcal{X}}$ and $\Pi^{t-1}_{\mathcal{Y}}$ for $\Pi^{t-1}$ ;
    Parties use Protocol 4 for simulating $\mathrm{P}_{\Pi_t(X\Pi^{t-1})(Y\Pi^{t-1})}$ with parameters $\lambda^{\min}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$,
    $\Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$, $N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$, $\lambda^{\min}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$, $\Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$, $N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$ and $\gamma$ ;
    Update $t \to t + 1$
if Total communication exceeds $l_{\max}$ bits then
    Declare an error
```

Lemma 21 (**Performance of Protocol 5**)**.** *Protocol 5 sends at most $l_{\max}$ bits, and has simulation error*

$$\begin{aligned}
&d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right)\\
&\leq \Pr\left(\mathtt{ic}(\Pi; X, Y) + \sum_{t=1}^{d} \delta_t > l_{\max}\right)\\
&\quad + \sum_{t=1}^{d}\Bigg[4\Pr\left((\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right) + 4\Pr\left((\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)\\
&\quad + 3\left(N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 2\right)2^{-\gamma} + \frac{3}{N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}} + \frac{3}{N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}}\Bigg],
\end{aligned}$$

where

$$\delta_t = \begin{cases} N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + 3\log N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 3\gamma & \textit{odd } t \\ N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 3\log N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + 3\gamma & \textit{even } t \end{cases}.$$

Remark 5. The fudge parameters ε' and λ' can be explicitly given by

$$\begin{aligned}
\varepsilon' &= \sum_{t=1}^{d}\Bigg[4\Pr\left((\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right) + 4\Pr\left((\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)\\
&\quad + 3\left(N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 2\right)2^{-\gamma} + \frac{3}{N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}} + \frac{3}{N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}}\Bigg],\\
\lambda' &= \sum_{t=1}^{d}\delta_t.
\end{aligned}$$

Proof. Consider a virtual protocol which does not terminate even if the total number of bits exceed $l_{\max}$. Denote the output of this protocol by $\bar{\Pi}_X = (\bar{\Pi}_{1\mathcal{X}}, \ldots, \bar{\Pi}_{d\mathcal{X}})$ and $\bar{\Pi}_Y = (\bar{\Pi}_{1\mathcal{Y}}, \ldots, \bar{\Pi}_{d\mathcal{Y}})$. We have

$$d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right)$$

$$\begin{aligned}
&\leq d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}, \mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}\right) + d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right)\\
&\leq \Pr\left((\Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}}) \neq (\bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}})\right) + d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right). \qquad (28)
\end{aligned}$$

First, we bound the second term of (28). By using triangular inequality repeatedly and by using Lemma 20, we have

$$\begin{aligned}
&d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right)\\
&\leq d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{1\mathcal{X}}\bar{\Pi}_{1\mathcal{Y}}\cdots\bar{\Pi}_{(d-1)\mathcal{X}}\bar{\Pi}_{(d-1)\mathcal{Y}}\bar{\Pi}_{d\mathcal{X}}\bar{\Pi}_{d\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1\cdots\Pi_{(d-1)}\Pi_{(d-1)}\bar{\Pi}_{d\mathcal{X}}\bar{\Pi}_{d\mathcal{Y}}XY}\right)\\
&\quad + d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_1\Pi_1\cdots\Pi_{(d-1)}\Pi_{(d-1)}\bar{\Pi}_{d\mathcal{X}}\bar{\Pi}_{d\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1\cdots\Pi_{(d-1)}\Pi_{(d-1)}\Pi_d\Pi_d XY}\right)\\
&= d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{1\mathcal{X}}\bar{\Pi}_{1\mathcal{Y}}\cdots\bar{\Pi}_{(d-1)\mathcal{X}}\bar{\Pi}_{(d-1)\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1\cdots\Pi_{(d-1)}\Pi_{(d-1)}XY}\right)\\
&\quad + d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{d\mathcal{X}}\bar{\Pi}_{d\mathcal{Y}}(X\Pi^{d-1})(Y\Pi^{d-1})}, \mathrm{P}_{\Pi_d\Pi_d(X\Pi^{d-1})(Y\Pi^{d-1})}\right)\\
&=\\
&\vdots\\
&= \sum_{t=1}^{d} d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{t\mathcal{X}}\bar{\Pi}_{t\mathcal{Y}}(X\Pi^{t-1})(Y\Pi^{t-1})}, \mathrm{P}_{\Pi_t\Pi_t(X\Pi^{t-1})(Y\Pi^{t-1})}\right)\\
&\leq \sum_{t:\mathrm{odd}} \Bigg[\Pr\left((\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right) + \Pr\left((\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)\\
&\quad + \left(N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + 1\right)2^{-\gamma} + \frac{1}{N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}}\Bigg]\\
&\quad + \sum_{t:\mathrm{even}} \Bigg[\Pr\left((\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right) + \Pr\left((\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)\\
&\quad + \left(N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 1\right)2^{-\gamma} + \frac{1}{N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}}\Bigg]\\
&\leq \sum_{t=1}^{d} \Bigg[\Pr\left((\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right) + \Pr\left((\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)\\
&\quad + \left(N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 2\right)2^{-\gamma} + \frac{1}{N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}} + \frac{1}{N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}}\Bigg]. \qquad (29)
\end{aligned}$$

By denoting

$$\begin{aligned}
l(X, Y, \bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}}) := &\sum_{t:\mathrm{odd}} h_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{X}}|Y, \bar{\Pi}^{t-1}_{\mathcal{Y}}) - h_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{X}}|X, \bar{\Pi}^{t-1}_{\mathcal{X}})\\
&+ \sum_{t:\mathrm{even}} h_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{Y}}|X, \bar{\Pi}^{t-1}_{\mathcal{X}}) - h_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{Y}}|Y, \bar{\Pi}^{t-1}_{\mathcal{Y}}),
\end{aligned}$$

Since $(\Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}})$ coincides with $(\bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}})$ when the accumulated message length of the protocol generating $(\bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}})$ does not exceed $l_{\max}$, and since the message length of each round is bounded by each term of $l(X, Y, \bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}})$ plus δ_t by Lemma 20 unless $(\bar{\Pi}_{t\mathcal{X}}, (Y, \bar{\Pi}^{t-1}_{\mathcal{Y}})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$ or $(\bar{\Pi}_{t\mathcal{Y}}, (X, \bar{\Pi}^{t-1}_{\mathcal{X}})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$, we have

$$\Pr\left((\Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}}) \neq (\bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}})\right)$$

$$
\begin{aligned}
&\leq \Pr\left(l(X, Y, \bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}}) + \sum_{t=1}^{d} \delta_t > l_{\max} \right) \\
&\quad + \Pr\left(\bigcup_{t:\text{odd}} (\bar{\Pi}_{t\mathcal{X}}, (Y, \bar{\Pi}_{\mathcal{Y}}^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} \text{ or } \bigcup_{t:\text{even}} (\bar{\Pi}_{t\mathcal{Y}}, (X, \bar{\Pi}_{\mathcal{X}}^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} \right)
\end{aligned} \tag{30}
$$

Since

$$
\Pr\left((X, Y, \bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}}) \in \mathcal{E} \right) \leq \Pr\left((X, Y, \Pi, \Pi) \in \mathcal{E} \right) + d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY} \right)
$$

for any event $\mathcal{E}$, it follows from (30) that

$$
\begin{aligned}
&\Pr\left((\Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}}) \neq (\bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}}) \right) \\
&\leq \Pr\left(l(X, Y, \Pi, \Pi) + \sum_{t=1}^{d} \delta_t > l_{\max} \right) \\
&\quad + \Pr\left(\bigcup_{t:\text{odd}} (\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} \text{ or } \bigcup_{t:\text{even}} (\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} \right) \\
&\quad + 2 d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY} \right) \\
&\leq \Pr\left(l(X, Y, \Pi, \Pi) + \sum_{t=1}^{d} \delta_t > l_{\max} \right) \\
&\quad + \sum_{t=1}^{d} \left[\Pr\left((\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} \right) + \Pr\left((\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} \right) \right] \\
&\quad + 2 d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY} \right).
\end{aligned} \tag{31}
$$

Thus, by combining this bound with (28) and (29), and by noting

$$
l(X, Y, \Pi, \Pi) = \mathtt{ic}(\Pi; X, Y),
$$

we have the desired bound on simulation error. ■

7 Asymptotic Optimality

We now present the proofs of Theorem 3 and Theorem 7. Both the proofs rely on carefully choosing the slice-sizes in the lower and upper bounds.

7.1 Proof of Theorem 3

We start with the upper bound. Note that, for IID random variables (Π^n, X^n, Y^n), the spectrums of $h(\Pi_t^n | Z^n, (\Pi^{t-1})^n)$ for[20] $Z = X$ or Y have width $O(\sqrt{n})$. Therefore, the parameters Δs and Ns that appear in the fudge parameters can be chosen as $O(n^{1/4})$. Specifically, by standard measure

[20] We use this notation throughout this section to avoid repetition.

concentration bounds (for bounded random variables), for every $\nu > 0$, there exists a constant[21] $c > 0$ such that with

$$\lambda^{\min}_{\mathrm{P}_{\Pi_t^n|Z^n(\Pi^{t-1})^n}} = nH(\Pi_t|Z,\Pi^{t-1}) - c\sqrt{n},$$
$$\lambda^{\max}_{\mathrm{P}_{\Pi_t^n|Z^n(\Pi^{t-1})^n}} = nH(\Pi_t|Z,\Pi^{t-1}) + c\sqrt{n},$$

the following bound holds:

$$\Pr\left((\Pi_t^n, (Z^n, (\Pi^{t-1})^n)) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t^n|Z^n(\Pi^{t-1})^n}} \right) \le \nu. \tag{32}$$

Let T denote the third central moment of the random variable $\mathtt{ic}(\Pi; X, Y)$. For

$$\lambda_n = n\mathtt{IC}(\pi) + \sqrt{n\mathtt{V}(\pi)}Q^{-1}\left(\varepsilon - 9d\nu - \frac{T^3}{2\mathtt{V}(\pi)^{3/2}\sqrt{n}}\right),$$

choosing $\Delta_{\mathrm{P}_{\Pi_t^n|Z^n(\Pi^{t-1})^n}} = N_{\mathrm{P}_{\Pi_t^n|Z^n(\Pi^{t-1})^n}} = \gamma = \sqrt{2c}n^{1/4}$, and $l_{\max} = \lambda_n + \sum_{t=1}^d \delta_t$ in Theorem 2 (for the definition of the fudge parameters, see Remark 5), we get a protocol of length $l_{\max}$ and satisfying

$$d_{\mathtt{var}}\left(\mathrm{P}_{\Pi^n_{\mathcal{X}}\Pi^n_{\mathcal{Y}}X^nY^n}, \mathrm{P}_{\Pi^n\Pi^nX^nY^n}\right) \le \Pr\left(\sum_{i=1}^n \mathtt{ic}(\Pi_i; X_i, Y_i) > \lambda_n\right) + 9d\nu$$

for sufficiently large n. Here, note that $\delta_t = O(n^{1/4})$. Thus, the Berry-Esséen theorem (*cf.* [Fel71]) and the observation above gives a protocol of length $l_{\max}$ attaining ε-simulation. Therefore, using the Taylor approximation of $Q(\cdot)$ yields the achievability of the claimed protocol length.

For the lower bound, we fix sufficiently small constant $\delta > 0$, and we set $\lambda^{(1)}_{\min} = n(H(X,Y) - \delta)$, $\lambda^{(1)}_{\max} = n(H(X,Y) + \delta)$, $\lambda^{(2)}_{\min} = n(H(X|Y,\Pi) - \delta)$, $\lambda^{(2)}_{\max} = n(H(X|Y,\Pi) + \delta)$, $\lambda^{(3)}_{\min} = n(H(X\Pi \triangle Y\Pi) - \delta)$, $\lambda^{(3)}_{\max} = n(H(X\Pi \triangle Y\Pi) + \delta)$, respectively. Then, by standard measure concentration bounds imply that the tail probability $\varepsilon_{\mathtt{tail}}$ in (3) is bounded above by $\frac{c}{n}$ for some constant $c > 0$. We also set $\eta = \frac{1}{n}$. For these choices of parameters, we note that the fudge parameter is $\lambda' = O(\log n)$. Thus, by setting

$$\begin{aligned}\lambda = \lambda_n &= n\mathtt{IC}(\pi) + \sqrt{n\mathtt{V}(\pi)}Q^{-1}\left(\varepsilon + \frac{c+2}{n} + \frac{T^3}{2\mathtt{V}(\pi)^{3/2}\sqrt{n}}\right) \\ &= n\mathtt{IC}(\pi) + \sqrt{n\mathtt{V}(\pi)}Q^{-1}(\varepsilon) + O(\log n),\end{aligned}$$

where the final equality is by the Tailor approximation, an application of the Berry-Esséen theorem to the bound in (2) gives the desired lower bound on the protocol length. ■

7.2 Proof of Theorem 5

Theorem 1 implies that if a protocol $\pi_{\mathtt{sim}}$ is such that

$$\log \|\pi_{\mathtt{sim}}\| < \lambda - \lambda', \tag{33}$$

[21] Although the constant depends on random variables appearing in each round, since the number of rounds is bounded, we take the maximum constant so that (32) holds for every t.

then its simulation error must be larger than

$$\Pr\left(\mathtt{ic}\left(\Pi^n; X^n, Y^n\right) > \lambda\right) - \varepsilon'. \tag{34}$$

To compute fudge parameters, we set $\lambda_{\text{min}}^{(1)} = n(H(X,Y) - \delta)$, $\lambda_{\text{max}}^{(1)} = n(H(X,Y) + \delta)$, $\lambda_{\text{min}}^{(2)} = n(H(X|Y,\Pi) - \delta)$, $\lambda_{\text{max}}^{(2)} = n(H(X|Y,\Pi) + \delta)$, $\lambda_{\text{min}}^{(3)} = n(H(X\Pi\triangle Y\Pi) - \delta)$, $\lambda_{\text{max}}^{(3)} = n(H(X\Pi\triangle Y\Pi) + \delta)$, respectively. By the Chernoff bound, there exists $E_1 > 0$ such that

$$\varepsilon_{\mathtt{tail}} \leq 2^{-E_1 n}.$$

Furthermore, $\Lambda_i = O(n)$ for $i = 1, 2, 3$. We set $\eta = 2^{-\frac{\delta}{27}n}$. It follows that

$$\varepsilon' \leq 2^{-E_1 n} + 2^{-\frac{\delta}{27}n} \tag{35}$$

and

$$\lambda' \leq \frac{\delta}{3}n + O(\log n). \tag{36}$$

Finally, upon setting

$$\lambda = n\mathtt{IC}(\pi) - \frac{\delta}{3} \tag{37}$$

and applying the Chernoff bound once more, we obtain a constant $E_2 > 0$ such that

$$\Pr\left(\mathtt{ic}\left(\Pi^n; X^n, Y^n\right) > \lambda\right) \geq 1 - 2^{-E_2 n}. \tag{38}$$

The result follows upon combining (33)-(38). ■

7.3 Proof of Theorem 7

For a sequence of protocols $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^{\infty}$ and a sequence of observations $(\mathbf{X}, \mathbf{Y}) = \{(X_n, Y_n)\}_{n=1}^{\infty}$, let

$$\underline{H}(\boldsymbol{\Pi}_t|\mathbf{Z}, \boldsymbol{\Pi}^{t-1}) = \sup\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(\Pi_{n,t}|Z_n\Pi_n^{t-1}) < \alpha\right) = 0\right\}, \tag{39}$$

$$\overline{H}(\boldsymbol{\Pi}_t|\mathbf{Z}, \boldsymbol{\Pi}^{t-1}) = \inf\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(\Pi_{n,t}|Z_n\Pi_n^{t-1}) > \alpha\right) = 0\right\}, \tag{40}$$

where $\mathbf{Z} = \mathbf{X}$ or $\mathbf{Y}$, $\boldsymbol{\Pi}_t = \{\Pi_{n,t}\}_{n=1}^{\infty}$ and $\boldsymbol{\Pi}_n^{t-1} = \{\Pi_n^{t-1}\}_{n=1}^{\infty}$ are sequences of transcripts of tth round and up to tth rounds, respectively. For achievability part, we fix arbitrary small $\delta > 0$, and set

$$\lambda^{\text{min}}_{\mathrm{P}_{\Pi_{n,t}|Z_n\Pi_n^{t-1}}} = n\left(\underline{H}(\boldsymbol{\Pi}_t|\mathbf{Z}, \boldsymbol{\Pi}^{t-1}) - \delta\right),$$

$$\lambda^{\text{max}}_{\mathrm{P}_{\Pi_{n,t}|Z_n\Pi_n^{t-1}}} = n\left(\overline{H}(\boldsymbol{\Pi}_t|\mathbf{Z}, \boldsymbol{\Pi}^{t-1}) + \delta\right),$$

$\Delta_{\mathrm{P}_{\Pi_{n,t}|Z_n\Pi_n^{t-1}}} = N_{\mathrm{P}_{\Pi_{n,t}|Z_n\Pi_n^{t-1}}} = \gamma = \sqrt{2\delta n}$. We set

$$l_{\text{max}} = n\left(\overline{\mathtt{IC}}(\boldsymbol{\pi}) + \delta\right) + \sum_{t=1}^{d}$$

$$= n\left(\overline{\mathtt{IC}}(\boldsymbol{\pi}) + \delta\right) + O(\sqrt{n}).$$

Then, by Theorem 2, by the definition of $\overline{\mathtt{IC}}(\boldsymbol{\pi})$ and by (39) and (40), there exists a simulation protocol of length $l_{\max}$ with vanishing simulation error. Since $\delta > 0$ is arbitrary, we have the desired achievability bound.

For converse part, we fix arbitrary $\delta > 0$, and set $\lambda_{\min}^{(1)} = n(\underline{H}(\mathbf{X}, \mathbf{Y}) - \delta)$, $\lambda_{\max}^{(1)} = n(\overline{H}(\mathbf{X}, \mathbf{Y}) + \delta)$, $\lambda_{\min}^{(2)} = n(\underline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) - \delta)$, $\lambda_{\max}^{(2)} = n(\overline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) + \delta)$, $\lambda_{\min}^{(3)} = n(\underline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) - \delta)$, $\lambda_{\max}^{(3)} = n(\overline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) + \delta)$, respectively, where

$$\begin{aligned}
\underline{H}(\mathbf{X}, \mathbf{Y}) &= \sup\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(X_nY_n) < \alpha\right) = 0\right\}, \\
\overline{H}(\mathbf{X}, \mathbf{Y}) &= \inf\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(X_nY_n) > \alpha\right) = 0\right\}, \\
\underline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) &= \sup\left\{\alpha : \Pr\left(h(X_n|Y_n\Pi_n) < \alpha\right) = 0\right\}, \\
\overline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) &= \inf\left\{\alpha : \Pr\left(h(X_n|Y_n\Pi_n) > \alpha\right) = 0\right\}, \\
\underline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) &= \sup\left\{\alpha : \Pr\left(-h(X_n\Pi_n\triangle Y_n\Pi_n) < \alpha\right) = 0\right\}, \\
\overline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) &= \inf\left\{\alpha : \Pr\left(-h(X_n\Pi_n\triangle Y_n\Pi_n) > \alpha\right) = 0\right\}.
\end{aligned}$$

Then, by the definitions, we find that the tail probability $\varepsilon_{\mathtt{tail}}$ in (3) converges to 0. We also set $\eta = (1/n)$. For these choices of parameters, we note that the fudge parameter is $\lambda' = O(\log n)$. Thus, by using the bound in (2) for

$$\lambda = \lambda_n = n\left(\overline{\mathtt{IC}}(\boldsymbol{\pi}) + \delta\right), \tag{41}$$

and by taking $\delta \to 0$, we have the desired converse bound. ■

A Example Protocol

To illustrate the utility of our lower bound, we consider a protocol π which takes very few values most of the time, but with very small probability it can send many different transcripts. The proposed protocol can be ε-simulated using very few bits of communication on average. But in the worst-case it requires as many bits of communication for ε-simulation as needed for data exchange, for all $\varepsilon > 0$ small enough.

Specifically, let $\mathcal{X} = \mathcal{Y} = \{1, \dots, 2^n\}$ and let π be a deterministic protocol such that the transcript $\tau(x, y)$ for (x, y) is given by

$$\tau(x,y) = \begin{cases} a & \text{if } x > \delta 2^n, y > \delta 2^n \\ b & \text{if } x > \delta 2^n, y \le \delta 2^n \\ c & \text{if } x \le \delta 2^n, y > \delta 2^n \\ (x,y) & \text{if } x \le \delta 2^n, y \le \delta 2^n \end{cases}$$

for some small $\delta > 0$, which will be specified later. Clearly, this protocol is interactive.

Let (X, Y) be the uniform random variables on $\mathcal{X} \times \mathcal{Y}$. Then,

$$\Pr\left(\Pi \notin \{a, b, c\}\right) = \delta^2.$$

Since

$$\mathrm{P}_{\Pi|X}(\tau(x,y)|x) = \begin{cases} 1-\delta & \text{if } x > \delta 2^n, y > \delta 2^n \\ \delta & \text{if } x > \delta 2^n, y \le \delta 2^n \\ 1-\delta & \text{if } x \le \delta 2^n, y > \delta 2^n \\ \frac{1}{2^n} & \text{if } x \le \delta 2^n, y \le \delta 2^n \end{cases}$$

and similarly for $\mathrm{P}_{\Pi|Y}(\tau(x,y)|y)$, we have

$$\mathtt{ic}(\tau(x,y); x, y) = \begin{cases} 2\log(1/(1-\delta)) & \text{if } x > \delta 2^n, y > \delta 2^n \\ \log(1/\delta) + \log(1/(1-\delta)) & \text{if } x > \delta 2^n, y \le \delta 2^n \\ \log(1/\delta) + \log(1/(1-\delta)) & \text{if } x \le \delta 2^n, y > \delta 2^n \\ 2n & \text{if } x \le \delta 2^n, y \le \delta 2^n \end{cases}.$$

Consider $\delta = \frac{1}{n}$, and $\varepsilon = \frac{1}{n^3}$. Note that for any $\lambda < 2n$,

$$\Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) \ge \Pr\left(\Pi\{a, b, c\}\right) = \delta^2 = \frac{1}{n^2} > \varepsilon,$$

and

$$\Pr\left(\mathtt{ic}(\Pi; X, Y) > 2n\right) = 0.$$

Thus, the ε-tail of information complexity density $\lambda_\varepsilon = \sup\{\lambda : \Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) > \varepsilon\}$ is given by

$$\lambda_\varepsilon = 2n. \tag{42}$$

On the other hand, we have

$$\begin{aligned} \mathtt{IC}(\pi) &= H(\Pi|X) + H(\Pi|Y) \\ &\le 2\delta[h(\delta) + \log n - \log(1/\delta)] + 2(1-\delta)h_b(\delta) \\ &\le \tilde{\mathcal{O}}(\delta^2) \end{aligned}$$

where $h_b(\cdot)$ is the binary entropy function.

Also, to evaluate the lower bound of Theorem 1, we bound the fudge parameters in that bound. To that end, we fix $\varepsilon_{\texttt{tail}} = 0$ and bound the spectrum lengths $\Lambda_1, \Lambda_2, \Lambda_3$. Since (X, Y) is uniform, $h(X, Y) = 2n$ and so, $\Lambda_1 = 0$. Also, note that with probability 1 the conditional entropy density $h(X|\Pi, Y)$ is either 0 or $\log(\delta 2^n)$, which implies $\Lambda_2 = \mathcal{O}(n)$. A similar argument shows that $\Lambda_3 = \mathcal{O}(n)$. Therefore, the fudge parameter

$$\lambda' = \mathcal{O}(\log \Lambda_1 \Lambda_2 \Lambda_3) = \mathcal{O}(\log n),$$

which in view of (42) and Theorem 1 gives $D_\varepsilon(\pi) = \Omega(2n)$. ■

B Proof of Lemma 22

Lemma 22. *Consider random variables X, Y, Z and V taking values in countable sets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and a finite set $\mathcal{V}$, respectively. Then, for every $0 < \varepsilon < 1/2$,*

$$S_{2\varepsilon}(X, Y|ZV) \geq S_\varepsilon(X, Y|Z) - \log|\mathcal{V}| - 2\log(1/2\varepsilon).$$

Proof. Consider random variables $K'_{\mathcal{X}}$ and $K'_{\mathcal{Y}}$ with a common range $\mathcal{K}'$ such that $(K'_{\mathcal{X}}, K'_{\mathcal{Y}})$ constitutes an ε-secret key for X and Y given eavesdropper's observation Z, recoverable using an interactive protocol π'. Let $\mathrm{Q}_{K'_{\mathcal{X}} K'_{\mathcal{Y}} \Pi' Z V}$ denote the distribution $\mathrm{P}'^{(2)}_{\texttt{unif}} \mathrm{P}_{\Pi' Z V}$, where $\mathrm{P}'^{(2)}_{\texttt{unif}}$ denotes the distribution

$$\mathrm{P}'^{(2)}_{\texttt{unif}}(k_{\mathcal{X}}, k_{\mathcal{Y}}) = \frac{\mathbb{1}(k_{\mathcal{X}} = k_{\mathcal{Y}})}{|\mathcal{K}'|}, \quad \forall\, k_{\mathcal{Y}}, k_{\mathcal{Y}} \in \mathcal{K}'.$$

Then, by definition of an ε-secret key, it holds that

$$d_{\texttt{var}}\left(\mathrm{P}_{K'_{\mathcal{X}} K'_{\mathcal{Y}} \Pi' Z}, \mathrm{Q}_{K'_{\mathcal{X}} K'_{\mathcal{Y}} \Pi' Z}\right) \leq \varepsilon. \tag{43}$$

Note that $H_{\min}(\mathrm{Q}_{K'_{\mathcal{X}} \Pi' Z} \mid \Pi' Z) \geq \log|\mathcal{K}'|$. Therefore, by Lemma 9 there exists a function $K_{\mathcal{X}} = K(K'_{\mathcal{X}})$ taking values in a set $\mathcal{K}$ with $\log|\mathcal{K}| \geq \log|\mathcal{K}'| - \log|\mathcal{V}| - 2\log(1/2\varepsilon)$ such that

$$d_{\texttt{var}}\left(\mathrm{Q}_{K_{\mathcal{X}} \Pi' Z V}, \mathrm{P}_{\texttt{unif}} \mathrm{Q}_{\Pi' Z V}\right) \leq \varepsilon, \tag{44}$$

where $\mathrm{P}_{\texttt{unif}}$ denotes the uniform distribution on the set $\mathcal{K}$. Upon letting $K_{\mathcal{Y}} = K(K'_{\mathcal{Y}})$ and defining $\mathrm{P}^{(2)}_{\texttt{unif}}$ analogously to $\mathrm{P}'^{(2)}_{\texttt{unif}}$ with $\mathcal{K}$ in place of $\mathcal{K}'$, we have

$$\begin{aligned} d_{\texttt{var}}\left(\mathrm{P}_{K_{\mathcal{X}} K_{\mathcal{Y}} \Pi' Z V}, \mathrm{P}^{(2)}_{\texttt{unif}} \mathrm{P}_{\Pi' Z V}\right) &\leq d_{\texttt{var}}\left(\mathrm{Q}_{K_{\mathcal{X}} K_{\mathcal{Y}} \Pi' Z V}, \mathrm{P}^{(2)}_{\texttt{unif}} \mathrm{P}_{\Pi' Z V}\right) + \varepsilon \\ &= d_{\texttt{var}}\left(\mathrm{Q}_{K \Pi' Z V}, \mathrm{P}_{\texttt{unif}} \mathrm{P}_{\Pi' Z V}\right) + \varepsilon \\ &\leq 2\varepsilon, \end{aligned}$$

where the first inequality is by (43) and the second by (44), and the equality is by the definition of Q. Therefore, $(K_{\mathcal{X}}, K_{\mathcal{Y}})$ constitutes a 2ε-secret key of length $\log|\mathcal{K}'| - \log|\mathcal{V}| - 2\log(1/2\varepsilon)$ for X and Y given eavesdropper's observation (Z, V). The claimed bound follows since K' was an arbitrary secret key for X and Y given eavesdropper's observation Z. ■

References

[AC93] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography–part i: Secret sharing. *IEEE Trans. Inf. Theory*, 39(4):1121–1132, July 1993. 1.2, 4.1

[AMS96] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proc. ACM Symposium on Theory of Computing (STOC)*, pages 20–29, 1996. 1

[Ari73] S. Arimoto. On the converse to the coding theorem for discrete memoryless channels. *IEEE Trans. Inf. Theory*, 19(3):357–359, May 1973. 3.3

[BBCM95] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer. Generalized privacy amplification. *IEEE Trans. Inf. Theory*, 41(6):1915–1923, November 1995. 1.2, 4.1.1

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proc. ACM Symposium on Theory of Computing (STOC)*, pages 67–76, 2010. 1, 8

[BBR88] C. H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM J. Comput.*, 17(2):210–229, 1988. 4.1, 4.1.1

[Bea89] D. Beaver. Perfect privacy for two party protocols. *Technical Report TR-11-89, Harvard University*, 1989. 1

[BR11] M. Braverman and A. Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011. 1, 1, 1.1, 1.1, 5, 1, 1.2, 8, 3.2, 3.3, 10, 3.3, 3.4

[Bra12] Mark Braverman. Interactive information complexity. In *Proc. ACM Symposium on Theory of Computing Conference (STOC)*, pages 505–524, 2012. 1, 1.1

[BRWY13] M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff. Direct products in communication complexity. In *FOCS*, pages 746–755, 2013. 3.3, 3.3, 11

[BW14] M. Braverman and O. Weinstein. An interactive information odometer with applications. *ECCC*, page Report No. 47, 2014. 3.3, 3.3, 11

[CK11] I. Csiszár and J. Körner. *Information theory: Coding theorems for discrete memoryless channels. 2nd edition.* Cambridge University Press, 2011. 1

[CN04] I. Csiszár and P. Narayan. Secrecy capacities for multiple terminals. *IEEE Trans. Inf. Theory*, 50(12):3047–3061, December 2004. 4.2, 5.2, 16, 5.3

[CN08] I. Csiszár and P. Narayan. Secrecy capacities for multiterminal channel models. *IEEE Trans. Inf. Theory*, 54(6):2437–2452, June 2008. 5

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience, 2006. 1.2, 6.1

[DK79] G. Dueck and J. Korner. Reliability function of a discrete memoryless channel at rates above capacity (corresp.). *Information Theory, IEEE Transactions on*, 25(1):82–85, Jan 1979. 3.3

[Fel71] W. Feller. *An Introduction to Probability Theory and its Applications, Volume II. 2nd edition.* John Wiley & Sons Inc., UK, 1971. 7.1

[FS02] M. Feder and N. Shulman. Source broadcasting with unknown amount of receiver side information. In *ITW*, pages 127–130, Oct 2002. 3.2

[GKR14a] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 176–185, 2014. 1.1

[GKR14b] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:113, 2014. 1.1

[Han03] T. S. Han. *Information-Spectrum Methods in Information Theory [English Translation]*. Series: Stochastic Modelling and Applied Probability, Vol. 50, Springer, 2003. 1.2, 6, 3.1, 3.4, 6.1

[Hay09] M. Hayashi. Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. Inf. Theory*, 55(11):4947–4966, Novemeber 2009. 3.3

[HTW14] M. Hayashi, H. Tyagi, and S. Watanabe. Secret key agreement: General capacity and second-order asymptotics. *arXiv:1411.0735*, 2014. 3.2

[HV93] T. S. Han and S. Verdú. Approximation theory of output statistics. *IEEE Trans. Inf. Theory*, 39(3):752–772, May 1993. 1.2, 3.4

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proc. ACM Symposium on Theory of Computing (STOC)*, pages 12–24, 1989. 4.1.1

[JPY12] R. Jain, A. Pereszlenyi, and P. Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *FOCS*, pages 167–176, 2012. 3.3

[Kus92] E. Kushilevitz. Privacy and communication complexity. *SIAM Journal on Math*, 5(2):273–284, 1992. 1

[Kuz12] S. Kuzuoka. On the redundancy of variable-rate slepian-wolf coding. *Proc. International Symposium on Information Theory and its Applications (ISITA)*, pages 155–159, 2012. 6.1

[KW88] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. In *Proc. Symposium on Theory of Computing (STOC)*, pages 539–550, 1988. 1

[Mau93] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory*, 39(3):733–742, May 1993. 1.2, 4.1

[MI11] N. Ma and P. Ishwar. Some results on distributed source coding for interactive function computation. *IEEE Trans. Inf. Theory*, 57(9):6180–6195, September 2011. 2, 3.3

[MK95] S. Miyake and F. Kanaya. Coding theorems on correlated general sources. *IIEICE Trans. Fundamental*, E78-A(9):1063–1070, September 1995. 3.1, 3.2, 6.1

[MT10] M. Madiman and P. Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Trans. Inf. Theory*, 56(6):2699–2713, June 2010. 5

[Mur14] J. Muramatsu. Channel coding and lossy source coding using a generator of constrained random numbers. *IEEE Trans. Inf. Theory*, 60(5):2667–2686, May 2014. 1.2, 3.2

[NTW15] P. Narayan, H. Tyagi, and S. Watanabe. *To appear, Proc. IEEE International Symposium on Information Theory*, 2015. 1

[OG84] A. Orlitsky and A. El Gamal. Communication with secrecy constraints. In *Proc. ACM Symposium on Theory of Computing (STOC)*, pages 217–224, 1984. 4.2

[Orl90] Alon Orlitsky. Worst-case interactive communication i: Two messages are almost optimal. *IEEE Trans. Inf. Theory*, 36(5):1111–1126, 1990. 3.2

[PPV10] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory*, 56(5):2307–2359, May 2010. 3.3

[PV10] Y. Polyanskiy and S. Verdú. Arimoto channel coding converse and Rényi divergence. *Proc. Conference on Communication, Control, and Computing (Allerton)*, pages 1327–1333, 2010. 3.3

[Ren05] R. Renner. Security of quantum key distribution. *Ph. D. Dissertation, ETH Zurich*, 2005. 4.1.1, 4.1.1

[RR11] J. M. Renes and R. Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Trans. Inf. Theory*, 57(11):7377–7385, November 2011. 1.2, 3.2

[RW05] R. Renner and S. Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *Proc. ASIACRYPT*, pages 199–216, 2005. 1.2, 4.1.1

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948. 1

[Str62] V. Strassen. Asymptotische abscha??tzungen in Shannon's informationstheorie. *Third Prague Conf. Inf. Theory*, pages 689–723, 1962. English translation: `http://www.math.cornell.edu/~pmlut/strassen.pdf`. 3.3

[Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36(2):423–439, 1965. 14

[SW73] D. Slepian and J. Wolf. Noiseless coding of correlated information source. *IEEE Trans. Inf. Theory*, 19(4):471–480, July 1973. 1, 1.2, 3.2, 4.2, 6.1

[TVW15] H. Tyagi, P. Viswanath, and S. Watanabe. Interactive communication for data exchange. *To appear, Proc. IEEE International Symposium on Information Theory*, 2015. 3.1, 3.2, 4.2, 4.2, 13, 4.2, 5

[TW14a] H. Tyagi and S. Watanabe. A bound for multiparty secret key agreement and implications for a problem of secure computing. In *EUROCRYPT*, pages 369–386, 2014. 1.2, 4.1, 4.1.2, 4.1.2, 5.3

[TW14b] H. Tyagi and S. Watanabe. Converses for secret key agreement and secure computing. *CoRR*, abs/1404.5715, 2014. 1.2, 4.1.2, 4.1.2

[Tya13] H. Tyagi. Common randomness principles of secrecy. *Ph. D. Dissertation, Univeristy of Maryland, College Park*, 2013. 5

[YAG14] M. H. Yassaee, M. R. Aref, and A. Gohari. Achievability proof via output statistics of random binning. *IEEE Trans. Inf. Theory*, 60(11):6760–6786, November 2014. 1.2, 3.2

[Yao79] A. C. Yao. Some complexity questions related to distributive computing. *Proc. Annual Symposium on Theory of Computing*, pages 209–213, 1979. 1

[YGA12] M. H. Yassaee, A. Gohari, and M. R. Aref. Channel simulation via interactive communications. In *Proc. IEEE Symposium on Information Theory (ISIT)*, pages 1049–1053, 2012. 1, 1.1

[YH10] En-Hui Yang and De-Ke He. Interactive encoding and decoding for one way learning: Near lossless recovery with side information at the decoder. *Information Theory, IEEE Transactions on*, 56(4):1808–1824, April 2010. 3.2

ECCC ISSN 1433-8092

http://eccc.hpi-web.de