Information Complexity Density and Simulation of Protocols

Himanshu Tyagi* Shaileshh Venkatakrishnan† Pramod Viswanath† Shun Watanabe‡

Abstract

Two parties observing correlated random variables seek to run an interactive communication protocol. How many bits must they exchange to simulate the protocol, namely to produce a view with a joint distribution within a fixed statistical distance of the joint distribution of the input and the transcript of the original protocol? We present an information spectrum approach for this problem whereby the information complexity of the protocol is replaced by its information complexity density. Our single-shot bounds relate the communication complexity of simulating a protocol to tail bounds for information complexity density. As a consequence, we obtain a strong converse and characterize the second-order asymptotic term in communication complexity for independent and identically distributed observation sequences. Furthermore, we obtain a general formula for the rate of communication complexity which applies to any sequence of observations and protocols. Connections with results from theoretical computer science and implications for the function computation problem are discussed.

I. 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 [54], which in turn is an important tool for deriving lower bounds for computational complexity [28] and

*Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India. Email: htyagi@ece.iisc.ernet.in
*Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, IL 61801, USA. Email: {bjjvnkt2, pramodv}@illinois.edu
‡Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan. Email: shunwata@cc.tuat.ac.jp

A preliminary version of this paper was presented at the 7th Innovations in Theoretical Computer Science (ITCS) conference, Cambridge, Massachusetts, USA, 2016.

for space complexity of streaming algorithms [2]. On the other hand, it is a significant generalization of the classic information theoretic problem of distributed data compression [46], 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[1] π is related closely to its *information complexity*[2] $\mathtt{IC}(\pi)$ defined as follows:

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

where $I(X \wedge Y|Z)$ denotes the conditional mutual information between X and Y given Z (cf. [45], [14]). For a protocol π with communication complexity $|\pi|$, a simulation protocol requiring $\tilde{\mathcal{O}}(\sqrt{\mathtt{IC}(\pi)|\pi|})$ bits of communication was given in [4] and one requiring $2^{\mathcal{O}(\mathtt{IC}(\pi))}$ bits of communication was given in [11] (see, also, [5]). A general version of the simulation problem was considered in [56], but only bounded round simulation protocols were considered. Interestingly, it was shown in [9] that the amortized[3] distributional communication complexity of simulating n copies of a protocol π for vanishing simulation error is bounded above by[4] $\mathtt{IC}(\pi)$. While a matching lower bound was also derived in [9], it is not valid in our context – [9] considered function computation and used a coordinate-wise error criterion. Nevertheless, we can readily modify the lower bound argument in [9] 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

[1]The difference between simulation and compression of protocols is significant and is discussed in Remark 2 below.

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

[3]Throughout the paper, "amortized" indicates that the observations are independently identically distributed and the protocol to be simulated is n copies of the same protocol.

[4]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 [33] using standard information theoretic techniques.

questions remain unanswered. For instance, (a) how does the distributional communication complexity of simulating a protocol depend on the simulation error ε? (b) Is there a general expression for distributional communication complexity which yields information complexity as the leading asymptotic term in the amortized setting? (c) The results available in the amortized setting address simulation of product protocols, namely the protocols with n-length inputs which execute the same protocol on each coordinate of the input. But how about the simulation of more complicated protocols such as a mixture $\pi_{\mathtt{mix}}$ of two product protocols π_1^n and π_2^n – does $\mathtt{IC}(\pi_{\mathtt{mix}})$ still constitute the leading asymptotic term in the communication complexity of simulating $\pi_{\mathtt{mix}}$?

The quantity $\mathtt{IC}(\pi)$ plays the same role in the simulation of protocols as $H(X)$ in the compression of X^n [45] and $H(X|Y)$ in the transmission of X^n by the first to the second party with access to Y^n [46]. The questions raised above have been addressed for these classic problems (cf. [22]). In this paper, we answer these questions for simulation of interactive protocols. We introduce another information theoretic quantity that plays a fundamental role in characterizing communication complexity of simulating a protocol and can differ from information complexity significantly. Specifically, 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

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

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 $\mathtt{IC}(\pi) = \mathbb{E}\left[\mathtt{ic}(\Pi; X, Y)\right]$. We show that it is the *ε-tail of the information complexity density* $\mathtt{ic}(\Pi; X, Y)$, i.e., the supremum[5] over values of λ such that $\Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) > \varepsilon$, 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 $\mathtt{IC}(\pi)$ becomes the leading term in communication complexity for simulating π only when, roughly,

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

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

This condition holds, for instance, in the amortized regime considered in [9]. However, the ε-tail of $\mathtt{ic}(\Pi; X, Y)$ can differ significantly from $\mathtt{IC}(\pi)$, the mean of $\mathtt{ic}(\Pi; X, Y)$. 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 $\mathtt{ic}(\Pi; X, Y)$ is greater than $2n$ while $\mathtt{IC}(\pi)$ is very small, just $\tilde{\mathcal{O}}(n^{-1})$.

A. Summary of results

We derive 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 $\mathtt{ic}(\Pi; X, Y)$.

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 captured 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 sheds light on the dependence of $D_\varepsilon(\pi)$ on ε even in the single-shot regime. For instance, our lower bound can be used to exhibit an arbitrary separation between $D_\varepsilon(\pi)$ and $\mathtt{IC}(\pi)$ when ε is not fixed. Specifically, 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 much greater than $2^{\mathcal{O}(\mathtt{IC}(\pi))}$ since $\mathtt{IC}(\pi) = \tilde{\mathcal{O}}(n^{-1})$. Remarkably, [21], [20] exhibited exponential separation between the distributional communication complexity of computing a function and the information complexity of that function even for a fixed ε, thereby establishing the optimality of the upper bound $D_\varepsilon(\pi) \leq 2^{\mathcal{O}(\mathtt{IC}(\pi))}$ given in [11]. Our simple example shows a much stronger separation between $D_\varepsilon(\pi)$ and $\mathtt{IC}(\pi)$, albeit for a vanishing ε.

Upper bound. To establish our asymptotic results, we propose a new protocol for simulating protocols with bounded number of rounds, which is of independent interest. For a protocol π with length much greater than the number of rounds of interaction, using our proposed protocol we show that $D_\varepsilon(\pi) \lesssim \lambda_\varepsilon$. Much as the protocol of [9], our simulation protocol simulates one round at a time, and thus, the slack in our upper bound depends on the number of rounds.

As pointed-out in footnote 5, our lower bound approaches the ε-tail of $\mathtt{ic}(\Pi; X, Y)$ 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.

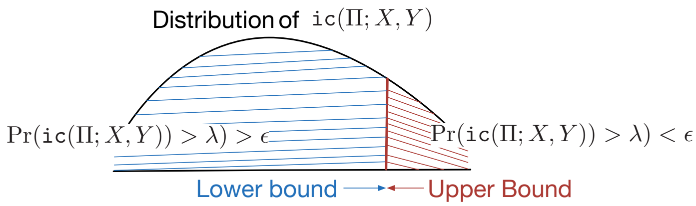

Fig. 1: Illustration of lower and upper bounds for $D_\varepsilon(\pi)$

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 [9]: 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 [9] or [56] 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 $\pi_{\mathtt{h}}$ or $\pi_{\mathtt{t}}$ n times, i.e., they use $\pi^n_{\mathtt{h}}$ if $\Pi_0 = \mathtt{h}$ and $\pi^n_{\mathtt{t}}$ 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 [9] 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 $\mathtt{IC}(\pi_{\mathtt{h}}) \gg \mathtt{IC}(\pi_{\mathtt{t}})$.

B. 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 [34], [1] or Section IV). Heuristically, a protocol can be simulated using fewer bits of communication than its length because of the correlation in 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. [7], [44]). 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 [51], [52].

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 classic Slepian-Wolf compression [46] for sending X to an observer of Y which is of optimal instantaneous rate. The second subroutine uses an idea that appeared first in [42] (see, also, [36], [55]) 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 in [9]. However, there are some crucial differences. The protocol in [9], 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 [9] and relies on randomness extraction techniques. In particular, the protocol in [9] 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 [23] (see the textbook [22] for a detailed account). In this approach, the classic measures of information such as Shannon entropy and mutual information are viewed as expectations of certain random variables referred to as *information densities*; the support of the distribution of these random variables is referred to as the corresponding *information spectrums*. For instance, Shannon entropy $H(X)$ is the expected value of entropy density $h(x) = -\log \mathrm{P}_X(x)$ which takes values in the entropy spectrum. The notion of "typical sets" is replaced by sets with bounded information densities. The coding theorems of classic information theory consider IID repetitions and rely on the so-called the *asymptotic equipartition property* (AEP) [13] which corresponds to the concentration of spectrums on small intervals. We refer to an interval of smallest length such that the information density lies in it with probability greater than $1-\varepsilon$ as an *ε-essential spectrum* or simply as an *essential spectrum* when ε is clear from the context. For *single-shot* problems, AEP does not hold and we have to work with the entire essential 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, rely on the spectrum of the information complexity density $\mathtt{ic}(\Pi; X, Y)$. 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 [22] to divide the essential spectrum into small intervals with information densities closely bounded from both sides[6]. 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.

In addition to the information complexity density described in Definition 1, we need to work with the following information densities and their spectrums:

(i) *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. [22]).

(ii) *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 [35], [22]. We shall use

[6]The spectrum slicing technique was introduced in [22] to derive the error exponents of various problems for general sources and a rate-distortion function for general sources.

the conditional entropy density $h(X|Y\Pi)$ in our bounds.

(iii) *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 [50]. We shall use the sum conditional entropy density $h(X\Pi \triangle Y\Pi)$.

(iv) Mutual information density of X and Y is given by $i(X \wedge Y) \stackrel{\text{def}}{=} h(X) - h(X|Y)$.

C. Organization

A formal statement of the problem along with the necessary preliminaries is given in the next section. Section III contains our main results. In Section IV, 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 most general and technical form of our lower bound and its proof is contained in Section V and that of our upper bound in Section VI; the proofs of the single-shot results and the asymptotic results reported in Section III are given in Section VII. We close with concluding remarks in Section VIII.

D. 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 correspondingly by $\mathcal{X}$, $\mathcal{Y}$, 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; the latter remains fixed in most of our discussion.

II. 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 $\mathcal{Y}$, 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 $\Pi_1, \Pi_2, ..., \Pi_r$. The parties communicate alternately with Party 1 transmitting in the odd rounds and Party 2 in the even rounds. Specifically, in each round i one of the party, say Party 1, communicates and transmits a string of bits $\Pi_i \in \{0,1\}^*$ determined by the previous transmissions $\Pi_1, ..., \Pi_{i-1}$ and the observations $(X, U_{\mathcal{X}}, U)$ of the communicating party. To each possible value of the bit string Π_i, a state from the state space $\{\texttt{C}, \phi\}$ is associated. If the next state is C, the other party starts communicating. If it is ϕ, the protocol stops and each party generates an output based on its local observation and transcript $\Pi^i = (\Pi_1, ..., \Pi_i)$ of the protocol. Note that the set $\mathcal{C}_i$ of possible values of Π_i and the associated next states C or ϕ for each value, are determined by a common function of $(X, U_{\mathcal{X}}, U, \Pi^{i-1})$ and $(Y, U_{\mathcal{Y}}, U, \Pi^{i-1})$ (cf. [19]), i.e., by a function of a random variable V such that

$$H(V|X, U_{\mathcal{X}}, U, \Pi^{i-1}) = H(V|Y, U_{\mathcal{Y}}, U, \Pi^{i-1}) = 0.$$

We denote the overall transcript of the protocol by Π. The *length of a protocol* π, $|\pi|$, is the maximum number of bits that are communicated in any execution of the protocol.

In the special case where $\mathcal{C}_i$ is a prefix-free set determined by Π^{i-1}, the protocol is called a *tree-protocol* (cf. [54], [31]). In this case, the set of transcripts of the protocol can be represented by a tree, termed the protocol tree, with each leaf corresponding to a particular realization of the transcript. Specifically, the protocol is defined by a binary tree where each internal node v is owned by either party, and node v is labeled either by a function $a_v : \mathcal{X} \times \mathcal{U}_{\mathcal{X}} \times \mathcal{U} \to \{0,1\}$ or $b_v : \mathcal{Y} \times \mathcal{U}_{\mathcal{Y}} \times \mathcal{U} \to \{0,1\}$. Then each leaf, or the path from the root to the leaf, corresponds to the overall transcript. Note that for a tree protocol the set of possible transcripts is prefix-free; in general, one can have protocols where this property does not hold. Our proposed protocol is indeed a tree protocol. On the other hand, our lower bound applies to the more general class of interactive protocols described above.

A random variable F is *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*, 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

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

of the original protocol. We are interested in the simulation of private coin protocols, using arbitrary[8] protocols; public-coin protocols can be simulated as private-coin protocols for each fixed value of public randomness.

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

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

where $d_{\texttt{var}}\left(\mathrm{P}, \mathrm{Q}\right) = \frac{1}{2}\sum_x |\mathrm{P}_x - \mathrm{Q}_x|$ 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*. A specific instance of this situation appears in [50] where $\Pi(X, Y) = (X, Y)$ is considered. For such protocols,

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

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 functions and even 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 [37] (see, also, [6], [30]).

Remark 2 (**Compression of protocols**). A protocol $\pi_{\texttt{com}}$ constitutes an ε-compression of a given protocol

[8]Since we are not interested in minimizing the amount of shared randomness used in a simulation, we allow arbitrary public coin protocols to be used as simulation protocols.

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 a randomized compression protocol $\pi_{\texttt{com}}$ can be derandomized to obtain a deterministic protocol with the same communication complexity. In fact, for deterministic protocols simulation and compression coincide. In general, however, compression is a more demanding task than simulation. For instance, consider the following simple example. Let inputs X and Y be constant, and let $U_{\mathcal{X}} = (U_{\mathcal{X},1}, \ldots, U_{\mathcal{X},2^r-1})$ and $U_{\mathcal{Y}} = (U_{\mathcal{Y},1}, \ldots, U_{\mathcal{Y},2^r-1})$ be two independently identically distributed sequences of independent and unbiased coin flips. Let π be a tree protocol of depth r such that the next node to communicate is given by $U_{\mathcal{X},v}$ if v is at an odd depth or $U_{\mathcal{Y},v}$ if v is at an even depth. To compress this protocol, the parties must reproduce exactly the same path as $\Pi = \Pi(U_{\mathcal{X}}, U_{\mathcal{Y}})$ from the root to a leaf, which requires roughly r bits of communication. On the other hand, to simulate the same protocol the parties need not communicate at all since the path can be sampled from the public coin $U = (U_1, ..., U_r)$ consisting of r independent and unbiased coin flips.

Indeed, our results show that in many cases, such as the amortized regime, compression requires strictly more communication than simulation. Specifically, all the results for ε-simulation in this paper can be modified to get corresponding results for ε-compression by replacing the information complexity density $\texttt{ic}(\tau; x, y)$ 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 expected value of the latter quantity exceeds that of the former. Therefore, compression requires more communication than simulation, in general. The corresponding results for compression remain essentially the same as those for simulation and have been omitted.

III. 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 a similar form as our lower bound. Instead of stating the most general technical results here, we present specific instantiations of interest: The single-shot regime, the amortized regime, and the results for the simulation of general protocols. The general lower bound is given in Section V and the general upper bound in Section VI.[9]

A. Single-shot bounds

Our first result claims that for every protocol π, $D_\varepsilon(\pi)$ is bounded below by the ε-tail of $\mathtt{ic}(\Pi; X, Y)$ up to an $\mathcal{O}(\log\log|\mathcal{X}||\mathcal{Y}|)$ term.

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

$$D_\varepsilon(\pi) \ge \sup\{\lambda : \Pr(\mathtt{ic}(\Pi; X, Y) > \lambda) \ge \varepsilon + 3\eta\} - 5\log\log|\mathcal{X}||\mathcal{Y}| - \delta(\eta), \tag{2}$$

where

$$\delta(\eta) = 9\log\frac{1}{\eta} + 5\log\log\frac{6}{\eta} + \log\frac{1}{1-3\eta} + 12.$$

A key feature of the bound above is that it brings out a precise dependence on the simulation error ε in terms of the ε-tail of $\mathtt{ic}(\Pi; X, Y)$; the expected value of $\mathtt{ic}(\Pi; X, Y)$, namely the information complexity $\mathtt{IC}(\pi)$, is a rough approximation of this ε-tail.

Next, we show a matching upper bound for $D_\varepsilon(\pi)$, albeit only for a restricted class of protocols where the length of the protocol is much less than the maximum number of rounds of interaction. Note that this restriction is significant since a protocol where parties communicate bits alternately has as many rounds of interaction as the length of the protocol. Nevertheless, for the aforementioned restricted class of protocols we show that $D_\varepsilon(\pi)$ is bounded above by the ε-tail of $\mathtt{ic}(\Pi; X, Y)$ up to an $\tilde{\mathcal{O}}(r_{\max}\sqrt{|\pi|})$ term.

Theorem 2. *Consider a protocol π with the maximum number of rounds $r_{\max} < \infty$ and $0 < \eta < 1$. Letting*

$$\varepsilon' = \eta + \frac{9\, r_{\max}}{\sqrt{|\pi|}} \quad \textit{and} \quad \lambda' = 12\, r_{\max}\sqrt{|\pi|} + 3\log\frac{11\, r_{\max}}{\eta},$$

for every $\varepsilon' < \varepsilon < 1$

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

[9]While the single-shot lower bound in Theorem 1 is tight enough to imply the converse part of Theorem 3, the single-shot upper bound in Theorem 2 does not imply the achievability part of Theorem 3. The relaxed version presented in this section exhibits an explicit form of residual terms.

B. *Amortized regime: second-order asymptotics*

It was shown in [9] 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 [9] 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 [9] 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 [47] and have received a lot of attention in information theory in recent years following [24], [40].

Our general lower bound and upper bound 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 $n\mathbb{E}\left[\mathtt{ic}(\Pi; X, Y)\right] = n\mathtt{IC}(\pi)$, recovering the result of [9]. 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 *strong converse*.

[10]The lower bound in [9] gives only the *weak converse* which holds only when $\varepsilon = \varepsilon_n \to 0$ as $n \to \infty$.

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*

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

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

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

A similar converse was first shown for the channel coding problem by Arimoto [3] (see [16], [41] for further refinements of this result), and has been studied for other classic 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 theoretical computer science 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) [10], [12], [27]. Our lower bound in Theorem 13, 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 a slight abuse of notations and terminologies, let $D_\varepsilon(f) = D_\varepsilon(f|\mathrm{P}_{XY})$ be the communication complexity of computing f. As we note in Remark 3, our general lower bound in Theorem 13 remains 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 13, 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

$$|\pi| \leq n[H(F|X) + H(F|Y) - \delta], \tag{3}$$

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 [9], [33] showed that the first order asymptotic term in the amortized communication complexity for function computation equals 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 (3) 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 [10], [12] and can be stronger in some regimes[11].

C. General formula for amortized communication complexity

Consider arbitrary distributions $\mathrm{P}_{X_n Y_n}$ 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 $D_{\varepsilon_n}(\pi_n|\mathrm{P}_{X_n Y_n})$ evolve as a function of n?

The previous section, and much of the theoretical computer science literature, has focused on the case when $\mathrm{P}_{X_n Y_n} = \mathrm{P}_{XY}^n$ and the same protocol π is executed on each coordinate. In this case, the leading asymptotic term is characterized by the information complexity of π. However, as we have seen in Example 1, for a mixed protocol the leading asymptotic term is characterized by the behavior of the "worst component" of the mixture. In this section, we formalize this observation by identifying the leading asymptotic term in $D_{\varepsilon_n}(\pi_n|\mathrm{P}_{X_n Y_n})$ for a general sequence of distributions[12] $\{\mathrm{P}_{X_n Y_n}\}_{n=1}^{\infty}$ and a general sequence of protocols $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^{\infty}$. Formally, the amortized (distributional) communication

[11]The result in [10], [12] shows a direct product theorem when we communicate less than $n\mathtt{IC}(f)/\mathtt{poly}(\log n)$.

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

complexity of $\boldsymbol{\pi}$ for $\{\mathrm{P}_{X_nY_n}\}_{n=1}^{\infty}$ is given by[13]

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

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 [23] for handling this situation (cf. [22]). 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

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

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 $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^{\infty}$,*

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

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 [9]. However, the utility of the general formula goes 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.

IV. 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.

A. Secret key agreement by public discussion

The problem of two party secret key agreement by public discussion was alluded to in [8], but a proper formulation and an asymptotically optimal construction appeared first in [34], [1]. Consider two

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

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 an interactive protocol 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* of length $\log|\mathcal{K}|$ if the following condition is satisfied:

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

where

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

The condition above ensures both reliable *recovery*, requiring $\Pr\left(K_{\mathcal{X}} \neq K_{\mathcal{Y}}\right)$ 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 [51] for a discussion.

Definition 4. Given $0 \leq \varepsilon < 1$, the maximum length 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 a private-coin 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

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

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 $(Y, V_{\mathcal{Y}})$.

1) Leftover hash lemma: A key tool for generating secret keys is the *leftover hash lemma* which, given a random variable X and an eavesdropper's l-bit 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

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

We define the *conditional min-entropy* of X given Z as

$$H_{\min}\left(\mathrm{P}_{XZ} \mid Z\right) \stackrel{\text{def}}{=} \sup_{\mathrm{Q}_Z \,:\, \mathtt{supp}(\mathrm{P}_Z) \subset \mathtt{supp}(\mathrm{Q}_Z)} H_{\min}\left(\mathrm{P}_{XZ} \mid \mathrm{Q}_Z\right). \tag{4}$$

An alternative operational form for conditional min-entropy was derived in [29] (see, also, [26, Theorem 2(ii)]), showing that $H_{\min}(\mathrm{P}_{XY}|Y)$ corresponds to the $-\log$ of the *average conditional guessing probability* for X given Y, i.e.,

$$H_{\min}\left(\mathrm{P}_{XZ} \mid Z\right) = -\log \sum_y \mathrm{P}_Y\left(y\right) \max_x \mathrm{P}_{X|Y}\left(x|y\right).$$

However, the variational form in (4) yields useful bounds by appropriately fixing Q_Z and is more more suited for our purpose.

Next, 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 of leftover hash lemma above was given in [25] and followed readily from [43].

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.

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 [51], [52]. In fact, the general upper bound in [51], [52] 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 fact, we consider a slightly relaxed version of the bound (cf. [52, Eq. (7)]), which is summarized in the following lemma.

Lemma 11. *For every* $0 \leq \varepsilon < 1$*,* $\eta > 0$ *and* λ*,*

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

for all distributions Q_X *and* Q_Y*, where* $(x)_+ = \max\{0, x\}$.

B. 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 [38], is in effect a two-party symmetric version of the Slepian-Wolf compression [46] (see [15] for a multiparty version). In a recent work [50], 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 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 [50] can be recovered as a special case of our simulation protocol in Section VI. The next result paraphrases [50, Theorem 2] and can also be recovered as a special case of Lemma 21.

We paraphrase the result form [50] in a form that is more suited for our application here. The data exchange protocol proposed in [50] relies on slicing the spectrum of $h(X|Y)$ (or $h(Y|X)$). Let $\mathcal{E}_{\mathtt{tail}}$ denote the tail event $h(X|Y) \notin [\lambda'_{\min}, \lambda'_{\max}]$. The protocol entails slicing an essential spectrum

$[\lambda'_{\min}, \lambda'_{\max}]$ into N parts of length $\Delta = \frac{\lambda'_{\max} - \lambda'_{\min}}{N}$ each.

Theorem 12 ([50, 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 $\mathcal{E}_{\texttt{error}} = \{X \neq \hat{X} \text{ or } Y \neq \hat{Y}\}$ the error event, it holds that*

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

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

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

where $h(X \triangle Y) = -\log \mathrm{P}_{X|Y}(X|Y)\,\mathrm{P}_{Y|X}(Y|X)$;

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

(iii) *for every (X, Y) such that $\lambda'_{\min} < h(X|Y) < \lambda'_{\max}$, 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 [50, 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 $\lambda'_{\min} < h(X|Y) < \lambda'_{\max}$, $h\left(X \triangle Y\right) + \Delta + N + \xi$ bits are communicated to interactively generate the transcript, the number of (variable length) transcripts is no more than[14] $2^{h(X \triangle Y)+\Delta+N+\xi}$. Property (ii) above was crucial to establish the communication complexity results of [50]; property (iii) was not relevant in the context of that work. On the other hand, here we shall use the protocol of Theorem 12 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; two to the power of the number of bits actually communicated in the interactive protocol is a loose upper bound on the former quantity.

Interestingly, our simulation protocol given in Section VI is used both in our upper bound to compress a given protocol and in our lower bound to complete the reduction argument.

V. General Lower Bound

In this section, we describe our general lower bound which yields all the lower bounds reported in Section III as special cases. Formally, given a private-coin protocol π, let $\pi_{\texttt{sim}}$ 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. Our

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

result involves the lengths of essential spectrums of information densities $\zeta_1 = h(X,Y)$, $\zeta_2 = h(X|Y\Pi)$, and $\zeta_3 = h\left(X\Pi \triangle Y\Pi\right)$. Let the tail events $\mathcal{E}_i \stackrel{\text{def}}{=} \{\zeta_i \notin [\lambda_{\min}^{(i)}, \lambda_{\max}^{(i)}]\}$, $i = 1,2,3$, satisfy

$$\Pr\left(\mathcal{E}_1\right) + \Pr\left(\mathcal{E}_2\right) + \Pr\left(\mathcal{E}_3\right) \leq \varepsilon_{\mathtt{tail}}, \tag{5}$$

where $\varepsilon_{\mathtt{tail}}$ can be chosen to be appropriately small, i.e., for $i = 1,2,3$, $[\lambda_{\min}^{(i)}, \lambda_{\max}^{(i)}]$ is an essential spectrum of ζ_i. Further, let $\Lambda_i = \lambda_{\max}^{(i)} - \lambda_{\min}^{(i)}$, $i = 1,2,3$.

Theorem 13. *Let* $0 \leq \lambda_{\min}^{(i)} \leq \lambda_{\max}^{(i)}$, $i = 1,2,3$ *and* $0 < \varepsilon_{\mathtt{tail}} < 1$ *satisfy* (5)*. Given* $0 \leq \varepsilon < 1$ *and a private-coin protocol* π*, for every* $0 < \eta < 1/3$

$$D_\varepsilon\left(\pi\right) \geq \sup\{\lambda : \Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) \geq \varepsilon + \varepsilon'\} - \lambda', \tag{6}$$

where $\varepsilon' = \varepsilon_{\mathtt{tail}} + 2\eta$ *and, letting*

$$\Lambda_i = \begin{cases} \lambda_{\max}^{(i)} - \lambda_{\min}^{(i)}, & \textit{if } \lambda_{\max}^{(i)} > \lambda_{\min}^{(i)}, \\ 1, & \textit{if } \lambda_{\max}^{(i)} = \lambda_{\min}^{(i)}, \end{cases}$$

$$\lambda' = 2\log \Lambda_1 \Lambda_3 + \log \Lambda_2 + \log \frac{1}{1 - 3\eta} + 9 \log \frac{1}{\eta} + 2.$$

Remark 3. The result above does not rely on the interactive nature of Π and is valid for simulation of any random variable Π. 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 (6).

The appearance of fudge parameters such as ε' and λ' in the bound above is typical since the techniques to bound the tail probability of random variables invariably entail such parameters, which are tuned based on the specific scenario being studied. For instance, the Chernoff bound has a parameter that is tuned with respect to the moment generating function of the random variable of interest. More relevant to the problem studied here, such fudge parameters also show up in the evaluation of error probability of single-party non-interactive compression problems ($cf.$ [23], [22]).

When the fudge parameters ε' and λ' are negligible, the right-side of the bound above is close to the ε-tail of $\mathtt{ic}(\Pi; X, Y)$. Indeed, the fudge parameters turn out to be negligible in the cases reported in Section III. For instance, for the amortized case ε' can be chosen to be arbitrarily small. The parameter λ' is related to the smallest length of an essential spectrum Λ, which, by the central limit theorem, is $\mathcal{O}(\sqrt{n})$; thus, $\lambda' = \mathcal{O}(\log n)$. On the other hand, the ε-tail of $\mathtt{ic}(\Pi; X, Y)$ 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.

In the remainder of this section, we provide a proof of Theorem 13. As described in the introduction, the main component in the proof of our lower bound is a reduction argument which uses a given simulation protocol to generate a secret key for X and Y. 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 [50] 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 [49] to characterize a class of securely computable functions.

Second, our methodology described above requires both side bounds for various information densities. 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 fix these gaps using careful spectrum slicing arguments. To make the exposition clear, we have divided the proof into four steps:

A) *From simulation to probability of error:* In the first step, we use a coupling argument to replace the variational distance based error criterion of simulation to a more tractable probability of error criterion.

B) *From partial knowledge to omniscience:* Next, as an intermediate step towards generating a secret key, once the parties execute the simulation protocol, we use the aforementioned data exchange protocol to enable omniscience. This yields a tractable form of common randomness which in turn yields tractable bounds for the rate of secret key generated.

C) *From original to conditional probabilities:* The next step is technical and uses a spectrum slicing argument to identify an appropriate critical event conditioned on which we get the desired tight bounds.

D) *From simulation to secret keys:* Finally, we complete the proof of reduction by combining all the previous steps.

A. 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 (see [48] for a general version of this result).

Lemma 14 (**Maximal Coupling Lemma**). *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*

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

Given the random transcript of the protocol Π and its estimates $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Y}}$ produced by the simulation, by the maximal coupling lemma, for each x, y there exists a joint distribution $\mathrm{P}_{\Pi\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}|X=x,Y=y}$ such that the marginal distributions $\mathrm{P}_{\Pi|X=x,Y=y}$ and $\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}|X=x,Y=y}$ are the same as that of the original random variables Π, $\Pi_{\mathcal{X}}$, and $\Pi_{\mathcal{Y}}$ and

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

where, with an abuse of notation, we use the same symbol Π for the random transcript as well the coupled marginal defined here[15].

Consequently,

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

As pointed in footnote 8, it suffices to consider 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

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

[15]It should be noted that the coupled marginal differs from the original random transcript to be simulated and may not even be a function of $(X, Y, U_{\mathcal{X}}, U_{\mathcal{Y}})$.

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.

It should also be noted that, while (7) resembles the condition for compression given in Remark 2, simulation does not imply compression in general. For instance, in the example given in Remark 2, the original transcript Π is a function of private coins $U_\mathcal{X}$ and $U_\mathcal{Y}$. However, the public coin U constitutes a simulation with estimates $\Pi_\mathcal{X} = \Pi_\mathcal{Y} = U$ since both Π and U are independent of X, Y and the marginal distribution of U is the same as that of Π. Therefore, the required coupling is obtained with U in the role of Π. But the coupled marginal Π is independent of the original transcript which is a function of $(U_\mathcal{X}, U_\mathcal{Y})$. Thus, the coupled Π does not constitute a compression of the protocol; compression mandates the reproduction of exactly the same trascript with large probability.

B. From partial knowledge to omniscience

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 12 to make all the data available to both parties, which was termed *attaining omniscience*[16] in [15]. 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 parties. Once both 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 IV-B, let $\pi_{\mathtt{DE}}$ be the data exchange protocol of Theorem 12 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 $\lambda = \lambda^{(3)}_{\max}$, $\lambda'_{\min} = \lambda^{(2)}_{\min}$, $\lambda'_{\max} = \lambda^{(2)}_{\max}$. Then, denoting by $\mathcal{E}_{\mathtt{error}}$ the error event for the protocol $\pi_{\mathtt{DE}}$ Theorem 12(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 (5). 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

[16]Csiszár and Narayan considered a multiterminal version of the data exchange problem in [15] and connected the minimum (amortized) rate of communication needed to the maximum (amortized) secret key rate.

the data exchange protocol $\pi_{\texttt{DE}}$ 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 $\pi_{\texttt{DE}}$. Note that $\pi_{\texttt{DE}}$ 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

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

which is the same as $\mathcal{E}_{\texttt{error}}^c$ for $\mathcal{E}_{\texttt{error}}$ in (9). Then, we have

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

where the last inequality follows from (7) and (9) since $\mathcal{B}^c = \mathcal{E}_{\texttt{error}}$.

C. 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 12 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 [51] shows that $S_\eta(X, Y)$ is bounded above, roughly, by the mutual information density $i(X \wedge Y) = \log \mathrm{P}_{XY}(X, Y) / \mathrm{P}_X(X)\mathrm{P}_Y(Y)$ (cf. Lemma 11), 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 $(\pi_{\mathtt{sim}}, \pi_{\mathtt{DE}})$ 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 $\pi_{\mathtt{DE}}$ 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 $(\pi_{\mathtt{sim}}, \pi_{\mathtt{DE}})$, which allows both parties to recover (X, Y), communicates no more than $|\pi_{\mathtt{sim}}| + h\left(X\Pi \triangle Y\Pi\right)$ bits for every fixed realization (X, Y, Π). Using the leftover hash lemma, we can extract a secret key of rate roughly $h(X, Y) - |\pi_{\mathtt{sim}}| - h\left(X\Pi \triangle Y\Pi\right)$.

The following approximate inequalities summarize our reduction:

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

where the first inequality is by Lemma 8 and the the second by Lemma 9. Note that the idea of generating secret keys from data exchange was first proposed in [15] in an amortized, IID setup and was shown to yield a secret key of asymptotically optimal rate.

From (12) and (13) it follows that

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

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 $D_\varepsilon\left(\pi\right)$. 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.

D. From original to conditional probabilities: A Spectrum slicing argument

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 entropy spectrum and the sum conditional entropy spectrum described in Section I-B. For the combined protocol $(\pi_{\mathtt{sim}}, \pi_{\mathtt{DE}})$ and the estimates $(\hat{X}, \hat{Y})$ as above, let

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

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 (5). Finally, define the event $\mathcal{E}_{ij}$ as follows:

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

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*

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

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

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

where the second inequality uses (11) and and the third uses (5). The proof is completed upon noting that $\{\mathcal{E}_{ij}\}_{i,j}$ constitutes a partition of $\mathcal{E}_{\texttt{sim}} \cap \mathcal{E}_{\texttt{DE}} \cap \mathcal{E}_{\lambda} \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c$ with $N_1 N_3$ parts.

E. 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 $\Pr\left(\mathcal{E}\right) \geq \alpha$.

Our proof essentially formalizes the steps outlined in Section V-C, 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 11 with $\mathrm{Q}_X = \mathrm{P}_X$ and $\mathrm{Q}_Y = \mathrm{P}_Y$, we get the following bound in place of (12):

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

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

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

To replace (13), note that 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, $h\left(X\Pi\triangle Y\Pi\right) \leq \lambda_{\min}^{(3)} + j\Delta_3$. It follows by Lemma 10 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, in the previous inequality, we used the leftover hash lemma (Lemma 9) by setting X as (X,Y), Z as U, and V as constant. Furthermore, by using

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

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

$$\begin{aligned} H_{\min}(\mathrm{P}_{XYU|\mathcal{E}} \mid U) &\geq \min_{x,y,u} -\log \frac{\mathrm{P}_{XYU|\mathcal{E}}(x,y,u)}{\mathrm{P}_U(u)} \\ &\geq \min_{x,y,u} -\log \frac{\mathrm{P}_{XYU}(x,y,u)\,\mathbb{1}(\mathrm{P}_{XYU|\mathcal{E}}(x,y,u) > 0)}{\alpha \mathrm{P}_U(u)} \\ &= \min_{x,y \in \mathcal{E}_i^{(1)}} h_{\mathrm{P}_{XY}}(x,y) + \log \alpha \\ &\geq \lambda_{\min}^{(1)} + (i-1)\Delta_1 + \log \alpha. \end{aligned} \tag{19}$$

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

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

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 $\mathcal{E}$

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

it holds that

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

On choosing

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

it follows from (15) that

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

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

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

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

$$\begin{aligned}|\pi_{\mathtt{sim}}| &\geq \lambda + 2\log\alpha - \Delta_1 - \Delta_2 - \Delta_3 - \xi - 6\log(1/\eta) + \log(1-3\eta) + 4 \\ &= \lambda + 2\log(\Pr(\mathcal{E}_\lambda) - \varepsilon - \varepsilon_{\mathtt{tail}} - \eta) - 2\log N_1 N_3 - (\Delta_1 + \Delta_2 + \Delta_3) - \log N_2 \\ &\qquad - 7\log(1/\eta) + \log(1-3\eta) + 4,\end{aligned}$$

where the equality holds for $\xi = -\log\eta + \log N_2$. Note that N_i and Δ_i in the right-side above can be chosen arbitrarily under the constraint $N_i\Delta_i = \Lambda_i$, $i = 1, 2, 3$; we set $N_i = \lfloor \Lambda_i \rfloor$, which implies $\Delta_i \leq 2$, $i = 1, 2, 3$. Substituting this choice of parameters, we get

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

where the final inequality holds for every λ such that $\Pr(\mathcal{E}_\lambda) \geq \varepsilon + \varepsilon_{\mathtt{tail}} + 2\eta$; Theorem 13 follows upon maximizing the right side-over all such λ.

VI. Simulation Protocol and the Upper Bound

In this section, we formally present an ε-simulation of a given interactive protocol π with the maximum number of rounds $r_{\max} < \infty$. 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 classic Slepian-Wolf compression [46] (see [35] 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. [17], [39], [53], [25], [50]).

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 [42], [36], [55].

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 [9]. However, the information theoretic form here makes it amenable to techniques such as spectrum slicing, which leads to tighter bounds than those established in [9].

For clarity, we build the simulation protocol in steps.

A. *Sending X using one-sided communication*

We start with the well-known Slepian-Wolf compression problem [46] 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 II 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 $\mathcal{F}_l(\mathcal{X})$ of mappings with l-bits output. The parameter l will correspond to the number of bits transmitted. The actual choice of l depends on the allowed probability of error ε, and can be chosen using Lemma 16 below. For the decoder, we use a slight modification of the standard joint typical decoder [13], [22]. 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 $\gamma > 0$. The formal description of the protocol is given in Protocol 1.

The following result is from [35], [22, Lemma 7.2.1] (see, also, [32]).

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

Protocol 1: Slepian-Wolf compression

Input: Observations X and Y, uniform public randomness U_{hash}, and a parameter l
Output: Estimate $\hat{X}$ of X at party 2
Both parties use U_{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

than ε using roughly as many bits as the ε-tail of $h_{\mathrm{P}_{X|Y}}(X|Y)$, namely the infimum over l such that $\mathrm{P}_{XY}\left(h_{\mathrm{P}_{X|Y}}(X|Y) > l\right)$ is less than ε.

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 allows us to choose the free parameter Q_{XY} as per our convenience and simplifies our performance analysis of the more involved protocols in the following sections.

B. *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 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 ACK or NACK feedback signal. The protocol stops when an ACK symbol is received or an error is declared. Note that the protocol of this section is essentially the same as the one in [50]. However, instead of executing the protocol for the original input, we apply it to the input generated by an appropriately chosen conditional distribution, which in turn is analyzed by choosing the free parameter Q_{XY} appropriately.

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

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

and

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

Further, let

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

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

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

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

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

Lemma 17 (Performance of Protocol 2). *For $(x,y) \in \mathcal{T}_{\mathrm{Q}_{X|Y}}^{(i)}$, $1 \le i \le N_{\mathrm{Q}_{X|Y}}$, 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:*

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

Proof: Since $(x,y) \in \mathcal{T}_{\mathrm{Q}_{X|Y}}^{(i)}$, an error occurs if there exists a $\hat{x} \ne x$ such that $(\hat{x}, y) \in \mathcal{T}_{\mathrm{Q}_{X|Y}}^{(j)}$ 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

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

Protocol 2: Interactive Slepian-Wolf compression

Input: Observations X and Y with distribution P_{XY}, uniform public randomness U_{hash}, auxiliary distribution Q_{XY}, and parameters γ, $\lambda^{\min}_{\mathrm{Q}_{X|Y}}$, $\Delta_{\mathrm{Q}_{X|Y}}$, $N_{\mathrm{Q}_{X|Y}}$, and l

Output: Estimate $\hat{X}$ of X at party 2

Both parties use U_{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

  if *More than one such x found* **then**

    protocol declares an error and terminates

  else

    send back $\Pi_{\mathsf{sim},2} = \mathrm{NACK}$

while $2 \le i \le N_{\mathrm{Q}_{X|Y}}$ *and party 2 did not send an ACK* **do**

  Both parties use U_{hash} to select f_i from $\mathcal{F}_{\Delta_{\mathrm{Q}_{X|Y}}}(\mathcal{X})$, independent of $f_1, ..., f_{i-1}$

  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* $f_j(x) = \Pi_{\mathsf{sim},2j-1}, \forall 1 \le j \le i$ **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 and terminates

    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 and terminates

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

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

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

Note that the protocol sends l bits in the first transmission, and $\Delta_{\mathrm{Q}_{X|Y}}$ bits and 1-bit feedback in every subsequent 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*

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

C. Simulation of Π_1 *using interactive communication*

We now proceed to simulate 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 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 public coin $U = (U_{\mathsf{hash}}, U_{\mathsf{sim}})$ 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. Since this portion is extracted from shared public randomness, it need not be communicated, which reduces the overall communication requirement. Note that since U_{sim} is independent of (X,Y), this portion of communication must as well be almost independent of (X,Y). The existence of 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 to guarantee the independence of X and Π_1 instead of (X,Y) and Π_1.

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

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

Party 1locally generates a sample $\Pi_{1\mathcal{X}}$ using $\mathrm{P}_{\Pi_1|Xf(\Pi_1)}(\cdot|X, U_{\mathsf{sim}})$

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

[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.

$N_{\mathrm{Q}_{\Pi_1|Y}}$ be defined analogously to the corresponding quantities in Section VI-B 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*

$$d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{1\mathcal{X}}\Pi_{1\mathcal{Y}}XY}, \mathrm{P}_{\Pi_1\Pi_1XY}\right) \le \Pr\left((\Pi_1, Y) \in \mathcal{T}^{(0)}_{\mathrm{Q}_{\Pi_1|Y}}\right) + N_{\mathrm{Q}_{\Pi_1|Y}} 2^{-\gamma} + \frac{1}{2}\sqrt{2^{k - H_{\min}(\mathrm{P}_{\Pi_1X}|\mathrm{Q}_X)}}$$

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

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 Q_{Π_1Y}, 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]

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

Since

$$d_{\mathtt{var}}(\mathrm{P}, \mathrm{Q}) = \mathrm{Q}\left(\{v : \mathrm{Q}(v) \ge \mathrm{P}(v)\}\right) - \mathrm{P}\left(\{v : \mathrm{Q}(v) \ge \mathrm{P}(v)\}\right)$$

and

$$\begin{aligned}
&\left\{(m, m', \tau, \hat{\tau}, x, y) : \mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\Pi_1XY}\left(m, m', \tau, \hat{\tau}, x, y\right) \ge \mathrm{P}_{f(\Pi_1)f'(\Pi_1)\Pi_1\hat{\Pi}_1XY}\left(m, m', \tau, \hat{\tau}, x, y\right)\right\} \\
&= \{(m, m', \tau, \hat{\tau}, x, y) : \tau = \hat{\tau}\},
\end{aligned}$$

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

we have

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

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

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

Therefore, the simulation error for Protocol 3 is bounded as

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

where the first inequality is by the monotonicity of $d_{\mathtt{var}}(\cdot, \cdot)$, 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

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

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. ■

D. Improved simulation of Π_1

In Protocol 3 we were able to reduce the communication by roughly $H_{\min}(\mathrm{P}_{\Pi_1X}|\mathrm{Q}_X)$ 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 worst-case reduction in communication we can obtain, and a higher gain is possible for specific realizations. 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 $j \in \{0, 1, \ldots, N_{\mathrm{P}_{\Pi_1|X}}\}$ 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

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

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 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 γ
Output: Estimates $\Pi_{1\mathcal{X}}$ and $\Pi_{1\mathcal{Y}}$ of Π_1
Party 1generates $J \sim \mathrm{P}_{J|X}(\cdot|X)$, and sends 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 γ, $\lambda^{\min}_{\mathrm{P}_{\Pi_1|Y}}$, $\Delta_{\mathrm{P}_{\Pi_1|Y}}$, $N_{\mathrm{P}_{\Pi_1|Y}}$, 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 and terminates

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*

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

Proof: First, we have

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

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

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

Thus, we have the desired bound on simulation error for our choice of k.

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

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

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

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

for each $j \in \mathcal{J}_{\mathsf{g}}$. We have the claimed bound by setting $\mathrm{Q}_{\Pi_1|Y} = \mathrm{P}_{\Pi_1|Y}$. ■

E. 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.

Protocol 5: Simulation of Π

Input: Observations X and Y with distribution P_{XY}, uniform public randomness $U = (U_{t,\mathsf{hash}}, U_{t,\mathsf{sim}} : t = 1, \dots, d)$, 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 $t = 1, \dots, d$ and γ.

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

while *Total communication is less than* $l_{\max}$ *bits, and simulation is not complete* **do**

- Party 1 and Party 2, respectively, use estimates $\Pi^{t-1}_{\mathcal{X}}$ and $\Pi^{t-1}_{\mathcal{Y}}$ for Π^{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 γ ;
- Update $t \to t+1$

if *Total communication exceeds* $l_{\max}$ *bits* **then**

- Declare an error

The following lemma provides a bound on the simulation 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

$$
\delta_t = \begin{cases} N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + 3\log N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 3\gamma & \textit{odd } t \\ N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 3\log N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + 3\gamma & \textit{even } t \end{cases}. \tag{28}
$$

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

$$\begin{aligned}
& d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right) \\
& \leq d_{\mathtt{var}}\left(\mathrm{P}_{\Pi_{\mathcal{X}}\Pi_{\mathcal{Y}}XY}, \mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}\right) + d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right) \\
& \leq \Pr\left((\Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}}) \neq (\bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}})\right) + d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY}, \mathrm{P}_{\Pi\Pi XY}\right). \qquad (29)
\end{aligned}$$

First, we bound the second term of (29). 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_dXY}\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:\text{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:\text{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 (30)
\end{aligned}$$

Denote

$$l(X, Y, \bar{\Pi}_{\mathcal{X}}, \bar{\Pi}_{\mathcal{Y}}) \stackrel{\text{def}}{=} \sum_{t:\text{odd}} h_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{X}}|Y, \bar{\Pi}^{t-1}_{\mathcal{Y}}) - h_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{X}}|X, \bar{\Pi}^{t-1}_{\mathcal{X}})$$

$$+\sum_{t:\text{even}} h_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{Y}}|X,\bar{\Pi}_{\mathcal{X}}^{t-1}) - h_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}(\bar{\Pi}_{t\mathcal{Y}}|Y,\bar{\Pi}_{\mathcal{Y}}^{t-1}).$$

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 $(\bar{\Pi}_{t\mathcal{X}},(Y,\bar{\Pi}_{\mathcal{Y}}^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$ or $(\bar{\Pi}_{t\mathcal{Y}},(X,\bar{\Pi}_{\mathcal{X}}^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$, we have

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

Since

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

for any event $\mathcal{E}$, it follows from (31) that

$$\begin{aligned}
&\Pr\left((\Pi_{\mathcal{X}},\Pi_{\mathcal{Y}}) \neq (\bar{\Pi}_{\mathcal{X}},\bar{\Pi}_{\mathcal{Y}})\right)\\
&\le \Pr\left(l(X,Y,\Pi,\Pi) + \sum_{t=1}^{d}\delta_t > l_{\max}\right)\\
&\quad + \Pr\left(\bigcup_{t:\text{odd}}(\Pi_t,(Y,\Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} \text{ or } \bigcup_{t:\text{even}}(\Pi_t,(X,\Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)\\
&\quad + 2d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY},\mathrm{P}_{\Pi\Pi XY}\right)\\
&\le \Pr\left(l(X,Y,\Pi,\Pi) + \sum_{t=1}^{d}\delta_t > l_{\max}\right)\\
&\quad + \sum_{t=1}^{d}\left[\Pr\left((\Pi_t,(Y,\Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right) + \Pr\left((\Pi_t,(X,\Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)\right]\\
&\quad + 2d_{\mathtt{var}}\left(\mathrm{P}_{\bar{\Pi}_{\mathcal{X}}\bar{\Pi}_{\mathcal{Y}}XY},\mathrm{P}_{\Pi\Pi XY}\right).
\end{aligned} \tag{32}$$

Thus, by combining this bound with (29) and (30), and by noting

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

we have the desired bound on simulation error. ■

We have proved the following general upper bound.

Theorem 22. *Consider a protocol π with the maximum number of rounds $r_{\max} < \infty$ and $0 < \eta < 1$. Then,*

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

where, with δ_t given by (28)*, $\lambda' = \sum_{t=1}^{r_{\max}} \delta_t$ and*

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

VII. Proofs of Results of Section III

We now apply the general lower bound in Theorem 13 and the upper bound in Theorem 22 to obtain the proofs of Theorem 1, 2, 3, 5, and 7. All proofs rely on carefully choosing the slice-sizes in the general lower and upper bounds.

A. Proofs of Theorem 1 and 2

We use the following simple observation to bound the minimum length of an essential spectrum.

Lemma 23. *For $0 < \varepsilon < 1$ and random variables X and Y, the conditional entropy density $h(x|y) = -\log \mathrm{P}_{X|Y}(x|y)$ satisfies*

$$\Pr\left(0 \le h(X|Y) \le \log\frac{|\mathcal{X}|}{\varepsilon}\right) \ge 1 - \varepsilon.$$

Proof: Since $h(X|Y)$ is nonnegative with probability 1, it suffices to show that

$$\Pr\left(h(X|Y) > \log\frac{|\mathcal{X}|}{\varepsilon}\right) \le \varepsilon.$$

Indeed,

$$\begin{aligned}
\Pr\left(h(X|Y) > \log\frac{|\mathcal{X}|}{\varepsilon}\right) &= \sum_y \mathrm{P}_Y(y) \sum_{x: h(x|y) > \log\frac{|\mathcal{X}|}{\varepsilon}} \mathrm{P}_{X|Y}(x|y) \\
&\le \sum_y \mathrm{P}_Y(y) \sum_{x: h(x|y) > \log\frac{|\mathcal{X}|}{\varepsilon}} 2^{-\log\frac{|\mathcal{X}|}{\varepsilon}} \\
&\le \sum_y \mathrm{P}_Y(y) \, |\mathcal{X}| \cdot \frac{\varepsilon}{|\mathcal{X}|}
\end{aligned}$$

$$= \varepsilon.$$

■

Proof of Theorem 1. Fix $\lambda_{\min}^{(1)} = \lambda_{\min}^{(2)} = \lambda_{\min}^{(3)} = 0$ and

$$\lambda_{\max}^{(1)} = \log |\mathcal{X}||\mathcal{Y}| + \log \frac{3}{\eta},$$
$$\lambda_{\max}^{(2)} = \log |\mathcal{X}| + \log \frac{3}{\eta},$$
$$\lambda_{\max}^{(3)} = \log |\mathcal{X}| + \log \frac{6}{\eta} + \log |\mathcal{Y}| + \log \frac{6}{\eta}.$$

Then, by Lemma 23, the events $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ in (5) each have probability less than $\eta/3$ and (5) holds with $\varepsilon_{\texttt{tail}} = \eta$. Thus, the conditions of Theorem 13 hold and the claimed bound follows since

$$\begin{aligned} 2\log \Lambda_1\Lambda_3 + \log \Lambda_2 &\leq 5 \log\left(\log |\mathcal{X}||\mathcal{Y}| + 2\log \frac{6}{\eta}\right) \\ &\leq 5 + 5\log\log|\mathcal{X}||\mathcal{Y}| + 5\log 2\log\frac{6}{\eta}, \end{aligned}$$

where the last inequality uses $\log(a+b) \leq \log 2\max\{a,b\} \leq 1 + \log a + \log b$. ■

Proof of Theorem 2. For $1 \leq t \leq r_{\max}$, fix $\lambda^{\min}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} = \lambda^{\min}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} = 0$,

$$\lambda^{\max}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} = \lambda^{\max}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} = \Lambda \stackrel{\text{def}}{=} |\pi| + \log\frac{11\, r_{\max}}{\eta},$$

$N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} = \sqrt{\lambda^{\max}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}}$, $N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} = \lambda^{\max}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$ for odd t and $N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} = \lambda^{\max}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$, $N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} = \sqrt{\lambda^{\max}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}}$ for even t, and

$$\gamma = 1 + \log \Lambda + \log \frac{11\, r_{\max}}{\eta}.$$

Then, by Lemma 23, $\Pr\left((\Pi_t, (Y, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right)$ and $\Pr\left((\Pi_t, (X, \Pi^{t-1})) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right)$ are bounded above by $\eta/(11\, r_{\max})$. Then, the parameters ε' and λ' of Theorem 22 are bounded above by

$$\begin{aligned} \varepsilon' &\leq \eta + r_{\max}\left[\frac{6\eta}{22\, r_{\max}\Lambda} + 3\left(\frac{1}{\sqrt{\Lambda}} + \frac{1}{\Lambda}\right)\right] \\ &\leq \eta + \frac{9\, r_{\max}}{\sqrt{\Lambda}} \\ &\leq \eta + \frac{9\, r_{\max}}{\sqrt{|\pi|}}, \end{aligned}$$

and

$$\lambda' \leq r_{\max} \cdot \left(2\sqrt{\Lambda} + 6\log\Lambda + 4 + 3\log\frac{11 r_{\max}}{\eta}\right)$$

$$\leq 12\, r_{\max}\sqrt{\Lambda} + 3\log\frac{11 r_{\max}}{\eta}.$$

The claimed bound follows by Theorem 22. ■

B. Proof of Theorem 3

We start with the upper bound. Note that, for IID random variables (Π^n, X^n, Y^n), the Chebyshev inequality implies that 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 $O(n^{1/4})$. More specifically, 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:

$$\Pr\left((\Pi_t^n,(Z^n,(\Pi^{t-1})^n)) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t^n|Z^n(\Pi^{t-1})^n}}\right) \leq \nu. \tag{33}$$

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 22, we get a protocol of length $l_{\max}$ and satisfying

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

for sufficiently large n. By its definition given in (28), $\delta_t = O(n^{1/4})$ for the choice of parameters above. Thus, the Berry-Esséen theorem (cf. [18]) 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 the Chernoff bound the tail probability $\varepsilon_{\mathtt{tail}}$

[20] We use this notation throughout this section for brevity.

[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 (33) holds for every t.

in (5) can be seen to be 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

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

where the final equality is by the Tailor approximation, an application of the Berry-Esséen theorem to the bound in (6) gives the desired lower bound on the protocol length. ■

C. Proof of Theorem 5

Theorem 13 implies that if a protocol $\pi_{\mathtt{sim}}$ is such that

$$\log|\pi_{\mathtt{sim}}| < \lambda - \lambda', \tag{34}$$

then its simulation error must be larger than

$$\Pr\left(\mathtt{ic}\left(\Pi^n; X^n, Y^n\right) > \lambda\right) - \varepsilon'. \tag{35}$$

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

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

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

$$\varepsilon' \le 2^{-E_1 n} + 2^{-\frac{\delta}{27}n} \tag{36}$$

and

$$\lambda' \le \frac{\delta}{3}n + O(\log n). \tag{37}$$

Finally, upon setting

$$\lambda = n\mathtt{IC}(\pi) - \frac{\delta}{3} \tag{38}$$

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

$$\Pr\left(\mathtt{ic}\left(\Pi^n; X^n, Y^n\right) > \lambda\right) \ge 1 - 2^{-E_2 n}. \tag{39}$$

The result follows upon combining (34)-(39). ■

D. 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

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

$$\overline{H}(\mathbf{\Pi}_t|\mathbf{Z}, \mathbf{\Pi}^{t-1}) = \inf\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(\Pi_{n,t}|Z_n\Pi_n^{t-1}) > \alpha\right) = 0\right\}, \tag{41}$$

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

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

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

$$\begin{aligned} l_{\max} &= n\left(\overline{\mathtt{IC}}(\boldsymbol{\pi}) + \delta\right) + \sum_{t=1}^{d} \delta_t \\ &= n\left(\overline{\mathtt{IC}}(\boldsymbol{\pi}) + \delta\right) + O(\sqrt{n}), \end{aligned}$$

where δ_t is given by (28). Then, by Theorem 22, the definition of $\overline{\mathtt{IC}}(\boldsymbol{\pi})$, (40), and (41), 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 the converse part, we fix arbitrary $\delta > 0$, and set $\lambda^{(1)}_{\min} = n(\underline{H}(\mathbf{X}, \mathbf{Y}) - \delta)$, $\lambda^{(1)}_{\max} = n(\overline{H}(\mathbf{X}, \mathbf{Y}) + \delta)$, $\lambda^{(2)}_{\min} = n(\underline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) - \delta)$, $\lambda^{(2)}_{\max} = n(\overline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) + \delta)$, $\lambda^{(3)}_{\min} = n(\underline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) - \delta)$, $\lambda^{(3)}_{\max} = n(\overline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) + \delta)$, respectively, where

$$\underline{H}(\mathbf{X}, \mathbf{Y}) = \sup\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(X_nY_n) < \alpha\right) = 0\right\},$$

$$\overline{H}(\mathbf{X}, \mathbf{Y}) = \inf\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(X_nY_n) > \alpha\right) = 0\right\},$$

$$\underline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) = \sup\left\{\alpha : \Pr\left(h(X_n|Y_n\Pi_n) < \alpha\right) = 0\right\},$$

$$\overline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) = \inf\left\{\alpha : \Pr\left(h(X_n|Y_n\Pi_n) > \alpha\right) = 0\right\},$$

$$\underline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) = \sup\left\{\alpha : \Pr\left(-h(X_n\Pi_n\triangle Y_n\Pi_n) < \alpha\right) = 0\right\},$$

$$\overline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) = \inf\left\{\alpha : \Pr\left(-h(X_n\Pi_n\triangle Y_n\Pi_n) > \alpha\right) = 0\right\}.$$

Then, by definition of the quantities involved, the tail probability $\varepsilon_{\texttt{tail}}$ in (5) converges to 0. Setting $\eta = (1/n)$, we note that the fudge parameter is $\lambda' = O(\log n)$. Thus, by using the bound in (6) for

$$\lambda = \lambda_n = n \left(\overline{\texttt{IC}}(\boldsymbol{\pi}) + \delta \right), \tag{42}$$

upon letting $\delta \to 0$, we have the desired converse bound. ■

VIII. Conclusion

We have proposed a *common randomness decomposition* based approach (cf. [49]) to derive a lower bound on communication complexity of protocol simulation by relating the protocol simulation problem to the secret key agreement. A key step in our approach is identifying the amount of common randomness generated through protocol simulation. Our estimate for the amount of common randomness does not rely on the structure of the function to be computed. This is in contrast to most of the existing lower bounds on communication complexity for function computation, such as the partition bound or the discrepancy bound, where the structure of the computed function plays an important role. In particular, a comparison of our approach with other existing approaches for specific functions is not available. An important future research agenda for us is to incorporate the structure of functions in our bound; the case of functions with a small range such as Boolean functions is of particular interest.

Appendix

A. Example Protocol

To illustrate the utility of our lower bound, we consider a deterministic 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 $\mathcal{X} = \mathcal{Y} = \{1, \ldots, 2^n\}$ and let π be a deterministic protocol such that the transcript $\tau(x, y)$ for (x, y) is given by

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

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 $\lambda < 2n$,

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

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

$$\lambda_\varepsilon = 2n. \tag{43}$$

On the other hand, we have

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

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

Also, to evaluate the lower bound of Theorem 13, we bound the fudge parameters in that bound. To that end, we fix $\varepsilon_{\mathtt{tail}} = 0$ and bound the spectrum lengths $\Lambda_1, \Lambda_2, \Lambda_3$. Since (X, Y) is uniform,

$h(X,Y) = 2n$ and so, $\Lambda_1 = 0$. Note that with probability 1 the conditional entropy density $h(X|\Pi, Y)$ is 0, $\log(\delta 2^n)$, or $\log((1-\delta)2^n)$, 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 (43) and Theorem 13 gives $D_\varepsilon(\pi) = \Omega(2n)$. ■

B. Proof of Lemma 10

Lemma. *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 $\mathrm{Q}_{K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' ZV}$ denote the distribution $\mathrm{P}'^{(2)}_{\mathtt{unif}}\mathrm{P}_{\Pi' Z}\mathrm{P}_{V|K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' Z}$, where $\mathrm{P}'^{(2)}_{\mathtt{unif}}$ denotes the distribution

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

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

$$\begin{aligned} d_{\mathtt{var}}\left(\mathrm{P}_{K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' ZV}, \mathrm{Q}_{K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' ZV}\right) &= d_{\mathtt{var}}\left(\mathrm{P}_{K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' Z}\mathrm{P}_{V|K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' Z}, \mathrm{P}'^{(2)}_{\mathtt{unif}}\mathrm{P}_{\Pi' Z}\mathrm{P}_{V|K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' Z}\right) \\ &= d_{\mathtt{var}}\left(\mathrm{P}_{K'_{\mathcal{X}}K'_{\mathcal{Y}}\Pi' Z}, \mathrm{P}'^{(2)}_{\mathtt{unif}}\mathrm{P}_{\Pi' Z}\right) \\ &\leq \varepsilon. \end{aligned} \tag{44}$$

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

$$d_{\mathtt{var}}\left(\mathrm{Q}_{K_{\mathcal{X}}\Pi' ZV}, \mathrm{P}_{\mathtt{unif}}\mathrm{Q}_{\Pi' ZV}\right) \leq \varepsilon, \tag{45}$$

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

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

$$\leq 2\varepsilon,$$

where the first inequality is by (44) and the second by (45), 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

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

[2] N. Alon, Y. Matias, and M. Szegedy, "The space complexity of approximating the frequency moments," in *Proc. ACM Symposium on Theory of Computing (STOC)*, 1996, pp. 20–29.

[3] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 357–359, May 1973.

[4] B. Barak, M. Braverman, X. Chen, and A. Rao, "How to compress interactive communication," in *Proc. ACM Symposium on Theory of Computing (STOC)*, 2010, pp. 67–76.

[5] B. Bauer, S. Moran, and A. Yehudayoff, "Internal Compression of Protocols to Entropy," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, vol. 40, 2015, pp. 481–496.

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

[7] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1915–1923, November 1995.

[8] C. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM J. Comput.*, vol. 17, no. 2, pp. 210–229, 1988.

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

[10] M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff, "Direct products in communication complexity," in *FOCS*, 2013, pp. 746–755.

[11] M. Braverman, "Interactive information complexity," in *Proc. ACM Symposium on Theory of Computing Conference (STOC)*, 2012, pp. 505–524.

[12] M. Braverman and O. Weinstein, "An interactive information odometer and applications," in *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, ser. STOC '15. ACM, 2015, pp. 341–350.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

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

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

[16] G. Dueck and J. Korner, "Reliability function of a discrete memoryless channel at rates above capacity (corresp.)," *Information Theory, IEEE Transactions on*, vol. 25, no. 1, pp. 82–85, Jan 1979.

[17] M. Feder and N. Shulman, "Source broadcasting with unknown amount of receiver side information," in *ITW*, Oct 2002, pp. 127–130.

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

[19] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[20] A. Ganor, G. Kol, and R. 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*, 2014, pp. 176–185.

[21] ——, "Exponential separation of information and communication for Boolean functions," in *Proc. ACM Symposium on Theory of Computing (STOC)*, 2015, pp. 557–566.

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

[23] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[24] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Novemeber 2009.

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

[26] M. Iwamoto and J. Shikata, "Information theoretic security for encryption based on conditional Rényi entropies," in *Information Theoretic Security*. Springer International Publishing, 2014, pp. 103–121.

[27] R. Jain, A. Pereszlenyi, and P. Yao, "A direct product theorem for the two-party bounded-round public-coin communication complexity," in *FOCS*, 2012, pp. 167–176.

[28] M. Karchmer and A. Wigderson, "Monotone circuits for connectivity require super-logarithmic depth," in *Proc. Symposium on Theory of Computing (STOC)*, 1988, pp. 539–550.

[29] R. König, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4337–4347, Sept 2009.

[30] E. Kushilevitz, "Privacy and communication complexity," *SIAM Journal on Math*, vol. 5, no. 2, pp. 273–284, 1992.

[31] E. Kushilevitz and N. Nisan, *Communication Complexity*. New York, NY, USA: Cambridge University Press, 1997.

[32] S. Kuzuoka, "On the redundancy of variable-rate Slepian-Wolf coding," *Proc. International Symposium on Information Theory and its Applications (ISITA)*, pp. 155–159, 2012.

[33] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6180–6195, September 2011.

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

[35] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IIEICE Trans. Fundamental*, vol. E78-A, no. 9, pp. 1063–1070, September 1995.

[36] J. Muramatsu, "Channel coding and lossy source coding using a generator of constrained random numbers," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2667–2686, May 2014.

[37] P. Narayan, H. Tyagi, and S. Watanabe, "Common randomness for secure computing," *Proc. IEEE International Symposium on Information Theory*, pp. 949–953, 2015.

[38] A. Orlitsky and A. E. Gamal, "Communication with secrecy constraints," in *Proc. ACM Symposium on Theory of Computing (STOC)*, 1984, pp. 217–224.

[39] A. Orlitsky, "Worst-case interactive communication I: Two messages are almost optimal," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1111–1126, 1990.

[40] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

[42] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7377–7385, November 2011.

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

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

[45] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[46] D. Slepian and J. Wolf, "Noiseless coding of correlated information source," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[47] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," *Third Prague Conf. Inf. Theory*, pp. 689–723, 1962, English translation: http://www.math.cornell.edu/ pmlut/strassen.pdf.

[48] ——, "The existence of probability measures with given marginals," *Ann. Math. Statist.*, vol. 36, no. 2, pp. 423–439, 1965.

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

[50] H. Tyagi, P. Viswanath, and S. Watanabe, "Interactive communication for data exchange," *Proc. IEEE International Symposium on Information Theory*, pp. 1806–1810, 2015, arXiv:1601.03617.

[51] H. Tyagi and S. Watanabe, "A bound for multiparty secret key agreement and implications for a problem of secure computing," in *EUROCRYPT*, 2014, pp. 369–386.

[52] ——, "Converses for secret key agreement and secure computing," *IEEE Trans. Inf. Theory*, vol. 61, pp. 4809–4827, 2015.

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

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

[55] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6760–6786, November 2014.

[56] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," in *Proc. IEEE Symposium on Information Theory (ISIT)*, 2012, pp. 1049–1053.