Interactive Communication for Data Exchange

Himanshu Tyagi* Pramod Viswanath† Shun Watanabe‡

Abstract

Two parties observing correlated data seek to exchange their data using interactive communication. How many bits must they communicate? We propose a new interactive protocol for data exchange which increases the communication size in steps until the task is done. We also derive a lower bound on the minimum number of bits that is based on relating the data exchange problem to the secret key agreement problem. Our single-shot analysis applies to all discrete random variables and yields upper and lower bounds of a similar form. In fact, the bounds are asymptotically tight and lead to a characterization of the optimal rate of communication needed for data exchange for a general source sequence such as a mixture of IID random variables as well as the optimal second-order asymptotic term in the length of communication needed for data exchange for IID random variables, when the probability of error is fixed. This gives a precise characterization of the asymptotic reduction in the length of optimal communication due to interaction; in particular, two-sided Slepian-Wolf compression is strictly suboptimal.

I. Introduction

Random correlated data (X, Y) is distributed between two parties with the first observing X and the second Y. What is the optimal communication protocol for the two parties to exchange their data? We allow (randomized) interactive communication protocols and a nonzero probability of error. This basic problem was introduced by El Gamal and Orlitsky in [22] where they presented bounds on the average number of bits of communication needed by deterministic protocols for data exchange without error[1]. When interaction is not allowed, a simple solution is to apply Slepian-Wolf compression [28] for each of the two one-sided data transfer problems. The resulting protocol was shown to be of optimal rate,

*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: 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 2015 IEEE International Symposium on Information Theory.

[1] They also illustrated the advantage of using randomized protocols when error is allowed

even in comparison with interactive protocols, when the underlying observations are *independent and identically distributed* (IID) by Csiszár and Narayan in [7]. They considered a multiterminal version of this problem, namely the problem of attaining *omniscience*, and established a lower bound on the rate of communication to show that interaction does not help in improving the asymptotic rate of communication if the probability of error vanishes to 0. However, interaction is known to be beneficial in one-sided data transfer ($cf.$ [23], [35], [8]). Can interaction help to reduce the communication needed for data exchange, and if so, what is the minimum length of interactive communication needed for data exchange?

We address the data exchange problem, illustrated in Figure 1, and provide answers to the questions raised above. We provide a new approach for establishing *converse bounds* for problems with interactive communication that relates efficient communication to secret key agreement and uses the recently established conditional independence testing bound for the length of a secret key [32]. Furthermore, we propose an *interactive protocol for data exchange* which matches the performance of our lower bound in several asymptotic regimes. As a consequence of the resulting single-shot bounds, we obtain a characterization of the optimal rate of communication needed for data exchange for a general sequence (X_n, Y_n) such as a mixture of IID random variables as well as the optimal second-order asymptotic term in the length of communication needed for data exchange for the IID random variables (X^n, Y^n), first instance of such a result in source coding with interactive communication[2]. This in turn leads to a precise characterization of the gain in asymptotic length of communication due to interaction.

Related work: The role of interaction in multiparty data compression has been long recognized. For the data exchange problem, this was first studied in [22] where interaction was used to facilitate data exchange by communicating optimally few bits in a single-shot setup with zero error. In a different direction, [8], [35] showed that interaction enables a universal variable-length coding for the Slepian-Wolf problem (see, also, [9] for a related work on universal encoding). Furthermore, it was shown in [35] that the redundancy in variable-length Slepian-Wolf coding with known distribution can be improved by interaction. In fact, the first part of our protocol is essentially the same as the one in [35] (see, also, [4]) wherein the length of the communication is increased in steps until the second party can decode. In [35], the step size was chosen to be $\mathcal{O}(\sqrt{n})$ for the universal scheme and roughly $\mathcal{O}(n^{1/4})$ for the known distribution case. We recast this protocol in an information spectrum framework (in the spirit of [15]) and allow for a flexible choice of the step size. By choosing this step size appropriately, we obtain exact asymptotic results in various regimes. Specifically, the optimal choice of this step size Δ

[2]In a different context, recently [2] showed that the second-order asymptotic term in the size of good channel codes can be improved using feedback.

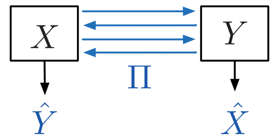

Fig. 1: The data exchange problem.

is given by the square root of the essential *length of the spectrum* of $\mathrm{P}_{X|Y}$, *i.e.*, $\Delta = \sqrt{\lambda_{\max} - \lambda_{\min}}$ where $\lambda_{\max}$ and $\lambda_{\min}$ are large probability upper and lower bounds, respectively, for the random variable $h(X|Y) = -\log \mathrm{P}_{X|Y}(X|Y)$. The $\mathcal{O}(\sqrt{n})$ choice for the universal case of [35] follows as a special case since for the universal setup with IID source $h(X^n|Y^n)$ can vary over an interval of length $\mathcal{O}(n)$. Similarly, for a given IID source, $h(X^n|Y^n)$ can essentially vary over an interval of length $\mathcal{O}(\sqrt{n})$ for which the choice of $\Delta = \mathcal{O}(n^{1/4})$ in [35] is appropriate by our general principle. While the right choice of Δ (upto the order) was identified in [35] for special cases, the optimality of this choice was not shown there. Our main contribution is a converse which shows that our achieved length of communication is optimal in several asymptotic regimes. In process, we obtain a precise characterization of gain due to interaction, one of the few such instances in literature.

Organization: The remainder of this paper is organized as follows: We formally describe the data exchange problem in Section II. Our results are summarized in Section III. Section IV contains our single-shot achievability scheme, along with the necessary prerequisites to describe it, and Section V contains our single-shot converse bound. The strong converse and second-order asymptotics for the communication length and the optimal rate of communication for general sources are obtained as a consequence of single-shot bounds in Section VI and VII, respectively. The final section contains a discussion of our results and extensions to the error exponent regime.

II. Problem formulation

Let the first and the second party, respectively, observe discrete random variables X and Y taking values in finite sets $\mathcal{X}$ and $\mathcal{Y}$. The two parties wish to know each other's observation using interactive communication over a noiseless (error-free) channel. The parties have access to local randomness (private coins) $U_{\mathcal{X}}$ and $U_{\mathcal{Y}}$ and shared randomness (public coins) U such that the random variables $U_{\mathcal{X}}, U_{\mathcal{Y}}, U$ are finite-valued and mutually independent and independent jointly of (X, Y). For simplicity, we resrict to *tree-protocols* ($cf.$ [19]). A tree-protocol π consists of a binary tree, termed the *protocol-tree*, with the vertices labeled by 1 or 2. The protocol starts at the root and proceeds towards the leaves. When the protocol is at vertex v with label $i_v \in \{1, 2\}$, party i_v communicates a bit b_v based on its local observations $(X, U_{\mathcal{X}}, U)$ for $i_v = 1$ or $(Y, U_{\mathcal{Y}}, U)$ for $i_v = 2$. The protocol proceeds to the left- or right-child of v, respectively, if b_v is 0 or 1. The protocol terminates when it reaches a leaf, at which point each party produces an output based on its local observations and the bits communicated during the protocol, namely the transcript $\Pi = \pi(X, Y, U_{\mathcal{X}}, U_{\mathcal{Y}}, U)$. Note that for tree-protocols the set of possible transcripts is prefix-free. Figure 2 shows an example of a protocol tree.

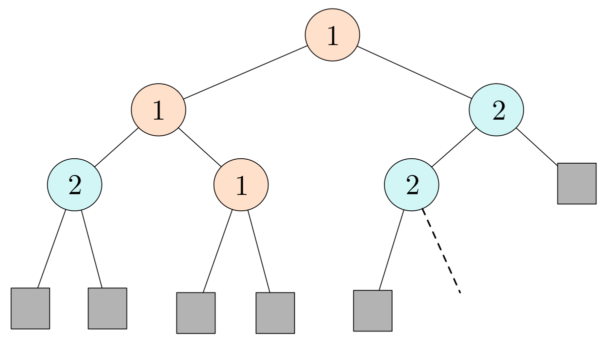

Fig. 2: A two-party protocol tree.

The *length of a protocol* π, denoted $|\pi|$, is the maximum accumulated number of bits transmitted in any realization of the protocol, namely the depth of the protocol tree.

Definition 1. For $0 \leq \varepsilon < 1$, a protocol Π attains *ε-data exchange* (ε-DE) if there exist functions $\hat{Y}$ and $\hat{X}$ of $(X, \Pi, U_{\mathcal{X}}, U)$ and $(Y, \Pi, U_{\mathcal{X}}, U)$, respectively, such that

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

The *minimum communication for ε-DE* $L_\varepsilon(X, Y)$ is the infimum of lengths of protocols[3] that attain ε-DE, *i.e.*, $L_\varepsilon(X, Y)$ is the minimum number of bits that must be communicated by the two parties in order to exchange their observed data with probability of error less than ε.

Protocols with 2 rounds of communication Π_1 and Π_2 which are functions of only X and Y, respectively, are termed *simple protocols*. We denote by $L_\varepsilon^{\mathrm{s}}(X, Y)$ the minimum communication for ε-DE by a simple protocol.

III. Summary of results

To describe our results, denote by $h(X) = -\log \mathrm{P}_X(X)$ and $h(X|Y) = -\log \mathrm{P}_{X|Y}(X|Y)$, respectively, the *entropy density* of X and the *conditional entropy density* of X given Y. Also, pivotal in our

[3]By derandomizing (1), it is easy to see that local and shared randomness does not help, and deterministic protocols attain $L_\varepsilon(X, Y)$.

results is a quantity we call the *sum conditional entropy density* of X and Y defined as

$$h(X\triangle Y) := h(X|Y) + h(Y|X).$$

An interactive data exchange protocol. Our data exchange protocol is based on an interactive version of the Slepian-Wolf protocol where the length of the communication is increased in steps until the second party decodes the data of the first. Similar protocols have been proposed earlier for distributed data compression in [9], [35], for protocol simulation in [4], and for secret key agreement in [16], [15].

In order to send X to an observer of Y, a single-shot version of the Slepian-Wolf protocol was proposed in [21] (see, also, [12, Lemma 7.2.1]). Roughly speaking, this protocol simply hashes X to as many bits as the right most point in the spectrum[4] of $\mathrm{P}_{X|Y}$. The main shortcoming of this protocol for our purpose is that it sends the same number of bits for every realization of (X, Y). However, we would like to use as few bits as possible for sending X to party 2 so that the remaining bits can be used for sending Y to party 1. Note that once X is recovered by party 2 correctly, it can send Y to Party 1 without error using, say, Shannon-Fano-Elias coding (eg. see [5, Section 5]); the length of this second communication is $\lceil h(Y|X)\rceil$ bits. Our protocol accomplishes the first part above using roughly $h(X|Y)$ bits of communication.

Specifically, in order to send X to Y we use a *spectrum slicing technique* introduced in [12] (see, also, [16], [15]). We divide the support $[\lambda_{\min}, \lambda_{\max}]$ of spectrum of $\mathrm{P}_{X|Y}$ into N slices size Δ each; see Figure 3 for an illustration.

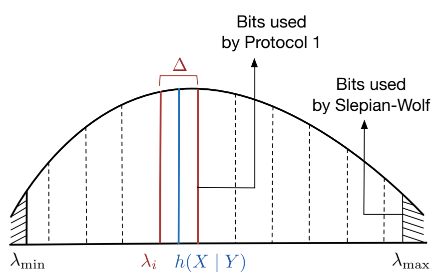

Fig. 3: Spectrum slicing in Protocol 1.

The protocol begins with the left most slice and party 1 sends $\lambda_{\min} + \Delta$ hash bits to party 2. If party 2 can find a unique x that is compatible with the received hash bits, it sends back an ACK and the protocol stops. Else, party 2 sends back a NACK and the protocol now moves to the next round, in which Party

[4]Spectrum of a distribution P_X refers, loosely, to the distribution of the random variable $-\log \mathrm{P}_X$.

1 sends additional Δ hash bits. The parties keep on moving to the next slice until either party 2 sends an ACK or all slices are covered. It is easy to show that this protocol is reliable and uses no more than $h(X|Y) + \Delta + N$ bits of communication for each realization of (X, Y). As mentioned above, once party 2 gets X, it sends back Y using $h(Y|X) + 1$ bits, thereby resulting in an overall communication of $h(X \triangle Y) + \Delta + N + 1$ bits. In our applications, we shall choose N and Δ to be of negligible order in comparison with the tail bounds for $h(X \triangle Y)$. Thus, we have the following upper bound on $L_\varepsilon(X, Y)$. (The statement here is rough; see Theorem 2 below for a precise version.)

Result 1 (Rough statement of the single-shot upper bound). *For every* $0 < \varepsilon < 1$,

$$L_\varepsilon(X, Y) \lesssim \inf\{\gamma : \mathbb{P}\left(h(X \triangle Y) > \gamma\right) \leq \varepsilon\}.$$

A converse bound. Our next result, which is perhaps the main contribution of this paper, is a lower bound on $L_\varepsilon(X, Y)$. This bound is derived by connecting the data exchange problem to the two-party secret key agreement problem. For an illustration of our approach in the case of IID random variables X^n and Y^n, note that the optimal rate of a secret key that can be generated is given by $I(X \wedge Y)$, the mutual information between X and Y [20], [1]. Also, using a privacy amplification argument ($cf.$ [3], [26]), it can be shown that a data exchange protocol using nR bits can yield roughly $n(H(XY) - R)$ bits of secret key. Therefore, $I(X \wedge Y)$ exceeds $H(XY) - R$, which further gives

$$R \geq H(X|Y) + H(Y|X).$$

This connection between secret key agreement and data exchange was noted first in [7] where it was used for designing an optimal rate secret key agreement protocol. Our converse proof is, in effect, a single-shot version of this argument.

Specifically, the "excess" randomness generated when the parties observing X and Y share a communication Π can be extracted as a secret key independent of Π using the *leftover hash lemma* [18], [27]. Thus, denoting by $S_\varepsilon(X, Y)$ the maximum length of secret key and by H the length of the common randomness ($cf.$ [1]) generated by the two parties during the protocol, we get

$$H - L_\varepsilon(X, Y) \leq S_\varepsilon(X, Y).$$

Next, we apply the recently established *conditional independence testing* upper bound for $S_\varepsilon(X, Y)$ [32], [33], which follows by reducing a binary hypothesis testing problem to secret key agreement. However, the resulting lower bound on $L_\varepsilon(X, Y)$ is good only when the spectrum of P_{XY} is concentrated. Heuristically, this slack in the lower bound arises since we are lower bounding the worst-case communication complexity of the protocol for data exchange – the resulting lower bound need not apply for every (X, Y) but only for a few realizations of (X, Y) with probability greater than ε. To remedy this shortcoming, we once again take recourse to spectrum slicing and show that there exists a slice of the spectrum of P_{XY} where the protocol requires sufficiently large number of bits; Figure 4 illustrates this approach. The resulting lower bound on $L_\varepsilon(X, Y)$ is stated below roughly, and a precise statement is given in Theorem 4.

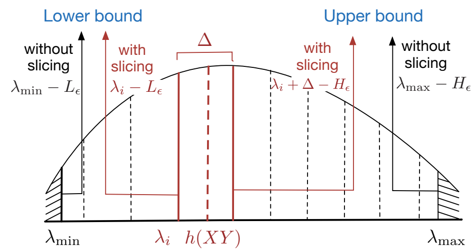

Fig. 4: Bounds on secret key length leading to the converse. Here L_ε abbreviates $L_\varepsilon(X, Y)$ and H_ε denotes the ε-tail of $h(X \triangle Y)$.

Result 2 (Rough statement of the single-shot lower bound). *For every* $0 < \varepsilon < 1$,

$$L_\varepsilon(X, Y) \gtrsim \inf\{\gamma : \mathbb{P}\left(h(X \triangle Y) > \gamma\right) \le \varepsilon\}.$$

Note that the upper and the lower bounds for $L_\varepsilon(X, Y)$ in the two results above appear to be of the same form (upon ignoring a few error terms). In fact, the displayed term dominates asymptotically and leads to tight bounds in several asympotitic regimes. Thus, the imprecise forms above capture the spirit of our bounds.

Asymptotic optimality. The single-shot bounds stated above are asymptotically tight up to the first order term for any sequence of random variables (X_n, Y_n), and up to the second order term for a sequence of IID random variables (X^n, Y^n).

Specifically, consider a general source sequence $(\mathbf{X}, \mathbf{Y}) = \{(X_n, Y_n)\}_{n=1}^\infty$. We are interested in characterizing the minimum asymptotic rate of communication for asymptotically error-free data exchange, and seek its comparison with the minimum rate possible using simple protocols.

Definition 2. The minimum rate of communication for data exchange R^* is defined as

$$R^*(\mathbf{X}, \mathbf{Y}) = \inf_{\varepsilon_n} \limsup_{n \to \infty} \frac{1}{n} L_{\varepsilon_n}(X_n, Y_n),$$

where the infimum is over all $\varepsilon_n \to 0$ as $n \to \infty$. The corresponding minimum rate for simple protocols is denoted by R_s^*.

Denote by $\overline{H}(\mathbf{X}\triangle\mathbf{Y})$, $\overline{H}(\mathbf{X}|\mathbf{Y})$, and $\overline{H}(\mathbf{Y}|\mathbf{X})$, respectively, the $\limsup$ in probability of random variables $h(X_n\triangle Y_n)$, $h(X_n|Y_n)$, and $h(Y_n|X_n)$. The quantity $\overline{H}(\mathbf{X}|\mathbf{Y})$ is standard in information spectrum method [13], [12] and corresponds to the asymptotically minimum rate of communication needed to send X_n to an observer of Y_n [21] (see, also, [12, Lemma 7.2.1]). Thus, a simple communication protocol of rate $\overline{H}(\mathbf{X}|\mathbf{Y})+\overline{H}(\mathbf{Y}|\mathbf{X})$ can be used to accomplish data exchange. In fact, a standard converse argument can be used to show the optimality of this rate for simple communication. Therefore, when we restrict ourselves to simple protocols, the asymptotically minimum rate of communication needed is

$$R_s^*(\mathbf{X},\mathbf{Y}) = \overline{H}(\mathbf{X}|\mathbf{Y}) + \overline{H}(\mathbf{Y}|\mathbf{X}).$$

As an illustration, consider the case when (X_n, Y_n) are generated by a mixture of two n-fold IID distributions $\mathrm{P}^{(1)}_{X^nY^n}$ and $\mathrm{P}^{(2)}_{X^nY^n}$. For this case, the right-side above equals ($cf.$ [12])

$$\max\{H(X^{(1)} \mid Y^{(1)}), H(X^{(2)} \mid Y^{(2)})\} + \max\{H(Y^{(1)} \mid X^{(1)}), H(Y^{(2)} \mid X^{(2)})\}.$$

Can we improve this rate by using interactive communication? Using our single-shot bounds for $L_\varepsilon(X,Y)$, we answer this question in the affirmative.

Result 3 (Min rate of communication for data exchange). *For a sequence of sources* $(\mathbf{X},\mathbf{Y}) = \{(X_n,Y_n)\}_{n=1}^\infty$,

$$R^*(\mathbf{X},\mathbf{Y}) = \overline{H}(\mathbf{X}\triangle\mathbf{Y}).$$

For the mixture of IID example above,

$$\overline{H}(\mathbf{X}\triangle\mathbf{Y}) = \max\{H(X^{(1)} \mid Y^{(1)}) + H(Y^{(1)} \mid X^{(1)}), H(X^{(2)} \mid Y^{(2)}) + H(Y^{(2)} \mid X^{(2)})\},$$

and therefore, simple protocols are strictly suboptimal in general. Note that while the standard information spectrum techniques suffice to prove the converse when we restrict to simple protocols, their extension to interactive protocols is unclear and our single-shot converse above is needed.

Turning now to the case of IID random variables, $i.e.$ when $X_n = X^n = (X_1, ..., X_n)$ and $Y_n = Y^n = (Y_1, ..., Y_n)$ are n-IID repetitions of random variables (X,Y). For brevity, denote by $R^*(X,Y)$ the corresponding minimum rate of communication for data exchange, and by $H(X\triangle Y)$ and V, respectively, the mean and the variance of $h(X\triangle Y)$. Earlier, Csiszár and Narayan [7] showed that $R^*(X,Y) =$

$H(X\triangle Y)$. We are interested in a finer asymptotic analysis than this first order characterization.

In particular, we are interested in characterizing the asymptotic behavior of $L_\varepsilon(X^n, Y^n)$ up to the second-order term, for every fixed ε in (0,1). We need the following notation:

$$R^*_\varepsilon(X,Y) = \lim_{n\to\infty} \frac{1}{n} L_\varepsilon(X^n, Y^n), \quad 0 < \varepsilon < 1.$$

Note that $R^*(X,Y) = \sup_{\varepsilon\in(0,1)} R^*_\varepsilon(X,Y)$. Our next result shows that $R^*_\varepsilon(X,Y)$ does not depend on ε and constitutes a *strong converse* for the result in [7].

Result 4 (Strong converse). *For every* $0 < \varepsilon < 1$*,*

$$R^*_\varepsilon(X,Y) = H(X\triangle Y).$$

In fact, this result follows from a general result characterizing the second-order asymptotic term[5].

Result 5 (Second-order asymptotic behavior). *For every* $0 < \varepsilon < 1$*,*

$$L_\varepsilon\left(X^n, Y^n\right) = nH(X\triangle Y) + \sqrt{nV} Q^{-1}(\varepsilon) + o(\sqrt{n}),$$

where $Q(a)$ *is the tail probability of the standard Gaussian distribution and* V *is the variance of the sum conditional entropy density* $h(X\triangle Y)$*.*

While simple protocols are optimal for the first-order term for IID observations, Example 1 in Section VII exhibits the strict suboptimality of simple protocols for the second-order term.

IV. A SINGLE-SHOT DATA EXCHANGE PROTOCOL

We present a single-shot scheme for two parties to exchange random observations X and Y. As a preparation for our protocol, we consider the restricted problem where only the second party observing Y seeks to know the observation X of the first party. This basic problem was introduced in the seminal work of Slepian and Wolf [28] for the case where the underlying data is IID where a scheme with optimal rate was given. A single-shot version of the Slepian-Wolf scheme was given in [21] (see, also, [12, Lemma 7.2.1]). which we describe below.

Using the standard "random binning" and "typical set" decoding argument, it follows that there exists

[5]Following the pioneering work of Strassen [29], study of these second-order terms in coding theorems has been revived recently by Hayashi [14], [17] and Polyanskiy, Poor, and Verdú [25].

an l-bit communication $\Pi_1 = \Pi_1(X)$ and a function $\hat{X}$ of (Π_1, Y) such that

$$\mathbb{P}\left(X \neq \hat{X}\right) \leq \mathbb{P}\left(h(X|Y) > l - \eta\right) + 2^{-\eta}. \tag{2}$$

In essence, the result of [21] shows that we can send X to Party 2 with a probability of error less than ε using roughly as many bits as the ε-tail of $h(X|Y)$. However, the proposed scheme uses the same number of bits for every realization of (X, Y). In contrast, we present an interactive scheme that achieves the same goal and uses roughly $h(X|Y)$ bits when the underlying observations are (X, Y)

While the bound in (2) can be used to establish the asymptotic rate optimality of the Slepian-Wolf scheme even for general sources, the number of bits communicated can be reduced for specific realizations of X, Y. This improvement is achieved using an interactive protocol with an $\mathrm{ACK} - \mathrm{NACK}$ feedback which halts as soon as the second party decodes first's observation; this protocol is described in the next subsection. A similar scheme was introduced by Feder and Shulman in [9], a variant of which was shown to be of least *average-case complexity* for stationary sources by Yang and He in [35], requiring $H(X \mid Y)$ bits on average. Another variant of this scheme has been used recently in [15] to generate secret keys of optimal asymptotic length upto the second-order term.

A. Interactive Slepian-Wolf Compression Protocol

We begin with an interactive scheme for sending X to an observer of Y, which hashes (bins) X into a few values as in the scheme of [21], but unlike that scheme, increases the hash-size gradually, starting with $\lambda_1 = \lambda_{\min}$ and increasing the size Δ-bits at a time until either X is recovered or $\lambda_{\max}$ bits have been sent. After each transmission, Party 2 sends either an ACK-NACK feedback signal; the protocol stops when an ACK symbol is received.

As mentioned in the introduction, we rely on spectrum slicing. Our protocol focuses on the "essential spectrum" of $h(X|Y)$, *i.e.*, those values of (X, Y) for which $h(X|Y) \in (\lambda_{\min}, \lambda_{\max})$. For $\lambda_{\min}, \lambda_{\max}, \Delta > 0$ with $\lambda_{\max} > \lambda_{\min}$, let

$$N = \frac{\lambda_{\max} - \lambda_{\min}}{\Delta}, \tag{3}$$

and

$$\lambda_i = \lambda_{\min} + (i-1)\Delta, \quad 1 \leq i \leq N. \tag{4}$$

Further, let

$$\mathcal{T}_0 = \Big\{(x,y) : h_{\mathrm{P}_{X|Y}}(x|y) \geq \lambda_{\max} \text{ or } h_{\mathrm{P}_{X|Y}}(x|y) < \lambda_{\min}\Big\}, \tag{5}$$

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

$$\mathcal{T}_i = \Big\{(x,y) : \lambda_i \leq h_{\mathrm{P}_{X|Y}}(x|y) < \lambda_i + \Delta\Big\}. \tag{6}$$

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

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

Protocol 1: Interactive Slepian-Wolf compression

```
Input: Observations X and Y, uniform public randomness U, and parameters l, Δ
Output: Estimate X̂ of X at party 2
Both parties use U to select h_1 uniformly from H_l(X)
Party 1 sends Π_1 = h_1(X)
if Party 2 finds a unique x ∈ T_1 with hash value h_1(x) = Π_1 then
    set X̂ = x
    send back Π_2 = ACK
else
    send back Π_2 = NACK
while 2 ≤ i ≤ N and party 2 did not send an ACK do
    Both parties use U to select h_i uniformly from H_Δ(X), independent of h_1, ..., h_{i-1}
    Party 1 sends Π_{2i-1} = h_i(X)
    if Party 2 finds a unique x ∈ T_i with hash value h_j(x) = Π_{2j-1}, ∀ 1 ≤ j ≤ i then
        set X̂ = x
        send back Π_{2i} = ACK
    else
        if More than one such x found then
            protocol declares an error
        else
            send back Π_{2i} = NACK
    Reset i → i + 1
if No X̂ found at party 2 then
    Protocol declares an error
```

Theorem 1 (Interactive Slepian-Wolf). *Protocol 1 with $l = \lambda_{\min} + \Delta + \eta$ sends at most $(h\,(X|Y) +$*

$\Delta + N + \eta)$ *bits when the observations are* $(X, Y) \notin \mathcal{T}_0$ *and has probability of error less than*

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

Note that when $\mathcal{T}_0$ is chosen to be of small probability, Protocol 1 sends essentially the same number of bits in the worst-case as the Slepian-Wolf protocol.

B. Interactive protocol for data exchange

Returning to the data exchange problem, our protocol for data exchange builds upon Protocol 1 and uses it to first transmit X to the second party (observing Y). Once Party 2 has recovered X correctly, it sends Y to Party 1 without error using, say, Shannon-Fano-Elias coding (eg. see [5, Section 5]); the length of this second communication is $\lceil h(Y|X) \rceil$ bits. When the accumulated number of bits communicated in the protocol exceeds a prescribed length $l_{\max}$, the parties abort the protocol and declare an error.[6] Using Theorem 1, the probability of error of the combined protocol is bounded above as follows.

Theorem 2 (**Interactive data exchange protocol**). *Given* $\lambda_{\min}, \lambda_{\max}, \Delta, \eta > 0$ *and for* N *in* (3)*, there exists a protocol for data exchange of length* $l_{\max}$ *such that*

$$\mathbb{P}\left(X \neq \hat{X} \text{ or } Y \neq \hat{Y}\right) \leq \mathbb{P}\left(h(X \triangle Y) + \Delta + N + \eta + 1 > l_{\max}\right) + \mathrm{P}_{XY}\left(\mathcal{T}_0\right) + N2^{-\eta}.$$

Thus, we attain ε-DE using a protocol of length

$$l_{\max} = \lambda_\varepsilon + \Delta + N + \eta + 1,$$

where λ_ε is the ε-tail of $h(X \triangle Y)$. Note that using the noninteractive Slepian-Wolf protocol on both sides will require roughly as many bits of communication as the sum of ε-tails of $h(X|Y)$ and $h(Y|X)$, which, in general, is more than the ε-tail of $h(X|Y) + h(Y|X)$.

C. Proof of Theorem 1

The theorem follows as a corollary of the following observation.

Lemma 3 (**Performance of Protocol 1**). *For* $(x, y) \in \mathcal{T}_i$*,* $1 \leq i \leq N$*, denoting by* $\hat{X} = \hat{X}(x, y)$ *the estimate of* x *at Party 2 at the end of the protocol (with the convention that* $\hat{X} = \emptyset$ *if an error is declared),*

[6]Alternatively, we can use the (noninteractive) Slepian-Wolf coding by setting the size of hash as $l_{\max} - (h(X|Y) + \Delta + N + \eta)$.

Protocol 1 sends at most $(l + (i-1)\Delta + i)$ *bits and has probability of error bounded above as follows:*

$$\mathbb{P}\left(\hat{X} \neq x \mid X = x, Y = y\right) \leq i 2^{\lambda_{\min} + \Delta - l}.$$

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

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

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

V. CONVERSE BOUND

Our converse bound, while heuristically simple, is technically involved. We first state the formal statement and provide the high level ideas underlying the proof; the formal proof will be provided later.

Our converse proof, too, relies on spectrum slicing to find the part of the spectrum of P_{XY} where the protocol communicates large number of bits. As in the achievability part, we shall focus on the "essential spectrum" of $h(XY)$.

Given $\lambda_{\max}$, $\lambda_{\min}$, and $\Delta > 0$, let N be as in (3) and the set $\mathcal{T}_0$ be as in (5), with $h_{\mathrm{P}_{X|Y}}(x|y)$ replaced by $h_{\mathrm{P}_{XY}}(xy)$ in those definitions.

Theorem 4. *For* $0 \leq \varepsilon < 1$, $0 < \eta < 1 - \varepsilon$, *and parameters* Δ, N *as above, the following lower bound on* $L_\varepsilon(X, Y)$ *holds for every* $\gamma > 0$*:*

$$L_\varepsilon(X, Y) \geq \gamma + 3 \log \left(\mathrm{P}_\gamma - \varepsilon - \mathrm{P}_{XY}(\mathcal{T}_0) - \frac{1}{N}\right)_+ + \log(1 - 2\eta) - \Delta - 6 \log N - 4 \log \frac{1}{\eta} - 1,$$

where $\mathrm{P}_\gamma := \mathrm{P}_{XY}\left(h(X \triangle Y) > \gamma\right)$.

Thus, a protocol attaining ε-DE must communicate roughly as many bits as ε-tail of $h(X \triangle Y)$.

The main idea is to relate data exchange to secret key agreement, which is done in the following two steps:

1) Given a protocol Π for ε-DE of length l, use the leftover hash lemma to extract an ε-secret key of length roughly $\lambda_{\min} - l$.
2) The length of the secret key that has been generated is bounded above by $S_\varepsilon(X, Y)$, the maximum possible length of an ε-secret key. Use the conditional independence testing bound in [32], [33] to further upper bound $S_\varepsilon(X, Y)$, thereby obtaining a lower bound for l.

This approach leads to a loss of $\lambda_{\max} - \lambda_{\min}$, the length of the spectrum of P_{XY}. However, since we are lower bounding the worse-case communication complexity, we can divide the spectrum into small slices of length Δ, and show that there is a slice where the communication is high enough by applying the steps above to the conditional distribution given that (X, Y) lie in a given slice. This reduces the loss from $\lambda_{\max} - \lambda_{\min}$ to Δ.

A. Review of two party secret key agreement

Consider two parties with the first and the second party, respectively, observing the random variable X and Y. Using an interactive protocol Π and their local observations, the parties agree on a secret key. A random variable K constitutes a secret key if the two parties form estimates that agree with K with probability close to 1 and K is concealed, in effect, from an eavesdropper with access to communication Π. Formally, let K_x and K_y, respectively, be randomized functions of (X, Π) and (Y, Π). Such random variables K_x and K_y with common range $\mathcal{K}$ constitute an *ε-secret key* (ε-SK) if the following condition is satisfied:

$$\frac{1}{2}\left\|\mathrm{P}_{K_x K_y \Pi} - \mathrm{P}^{(2)}_{\mathtt{unif}} \times \mathrm{P}_{\Pi}\right\| \leq \varepsilon,$$

where

$$\mathrm{P}^{(2)}_{\mathtt{unif}}\left(k_x, k_y\right) = \frac{\mathbb{1}(k_x = k_y)}{|\mathcal{K}|},$$

and $\|\cdot\|$ is the variational distance. The condition above ensures both reliable *recovery*, requiring $\mathbb{P}\left(K_x \neq K_y\right)$ to be small, and information theoretic *secrecy*, requiring the distribution of K_x (or K_y) to be almost independent of the communication Π and to be almost uniform. See [32] for a discussion on connections between the combined condition above and the usual separate conditions for recovery and secrecy.

Definition 3. Given $0 \le \varepsilon < 1$, the supremum over lengths $\log |\mathcal{K}|$ of an ε-SK is denoted by $S_\varepsilon(X, Y)$.

A key tool for generating secret keys is the *leftover hash lemma* [18], [27] which, given a random variable X and an l-bit eavesdropper's observation Z, allows us to extract roughly $H_{\min}(\mathrm{P}_X) - l$ bits of uniform bits, independent of Z. Here $H_{\min}$ denotes the *min-entropy* and is given by

$$H_{\min}\left(\mathrm{P}_X\right) = \sup_x - \log \mathrm{P}_X\left(x\right).$$

Formally, 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

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

Lemma 5 (Leftover Hash). *Consider random variables X and Z taking values in a countable set $\mathcal{X}$ and a finite set $\mathcal{Z}$, respectively. Let S be a random seed such that f_S is uniformly distributed over a 2-universal family $\mathcal{F}$. Then, for $K = f_S(X)$*

$$\|\mathrm{P}_{KZS} - \mathrm{P}_{\mathtt{unif}}\mathrm{P}_{ZS}\| \le \sqrt{|\mathcal{K}||\mathcal{Z}|2^{-H_{\min}(\mathrm{P}_X)}},$$

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

The version above is a straightforward modification of the leftover hash lemma in, for instance, [26] and can be derived in a similar manner (see Appendix B of [15]).

Next, we recall the *conditional independence testing* upper bound on $S_\varepsilon(X, Y)$, which was established in [32], [33]. In fact, the general upper bound in [32], [33] is a single-shot upper bound on the secret key length for a multiparty secret key agreement problem with side information at the eavesdropper. Below, we recall a specialization of the general result for the two party case with no side information at the eavesdropper. In order to state the result, we need the following concept from binary hypothesis testing.

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

$$\beta_\varepsilon(\mathrm{P}, \mathrm{Q}) := \inf_{\mathrm{T}\,:\,\mathrm{P}[\mathrm{T}] \geq 1-\varepsilon} \mathrm{Q}[\mathrm{T}], \tag{7}$$

where

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

The following upper bound for $S_\varepsilon(X, Y)$ was established in [32], [33].

Theorem 6 (**Conditional independence testing bound**). *Given $0 \leq \varepsilon < 1$, $0 < \eta < 1 - \varepsilon$, the following bound holds:*

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

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

We close by noting a further upper bound for $\beta_\varepsilon(\mathrm{P}, \mathrm{Q})$, which is easy to derive ($cf.$ [24]).

Lemma 7. *For every $0 \leq \varepsilon \leq 1$ and λ,*

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

where $(x)_+ = \max\{0, x\}$.

B. Converse bound for almost uniform distribution

First, we consider a converse bound under the almost uniformity assumption. Suppose that there exist $\lambda_{\min}$ and $\lambda_{\max}$ such that

$$\lambda_{\min} \leq -\log \mathrm{P}_{XY}(x, y) \leq \lambda_{\max}, \quad \forall (x, y) \in \mathrm{supp}(\mathrm{P}_{XY}), \tag{8}$$

where $\mathrm{supp}(\mathrm{P}_{XY})$ denotes the support of P_{XY}. We call such a distribution P_{XY} an almost uniform distribution with margin $\Delta = (\lambda_{\max} - \lambda_{\min})$.

Theorem 8. *Let* P_{XY} *be almost uniform with margin* Δ*. Given* $0 \le \varepsilon < 1$*, for every* $0 < \eta < 1 - \varepsilon$*, and all distributions* Q_X *and* Q_Y*, it holds that*

$$L_\varepsilon(X,Y) \ge \gamma + \log \left(\mathbb{P}\left(-\log \frac{\mathrm{P}_{XY}\left(X,Y\right)^2}{\mathrm{Q}_X\left(X\right)\mathrm{Q}_Y\left(Y\right)} \ge \gamma \right) - \varepsilon - 2\eta \right)_+ - \Delta - 4\log\frac{1}{\eta} - 1.$$

Remark 1. If $\Delta \approx 0$ (the almost uniform case), the bound above yields Result 2 upon choosing $\mathrm{Q}_X = \mathrm{P}_X$ and $\mathrm{Q}_Y = \mathrm{P}_Y$.

Proof. Given a protocol Π of length l that attains ε-DE, Lemma 5 implies that there exists an $(\varepsilon+\eta)$-SK taking values in $\mathcal{K}$ with

$$\log|\mathcal{K}| \ge \lambda_{\min} - l - 2\log(1/\eta) - 1.$$

Also, by Theorem 6

$$\log|\mathcal{K}| \le -\log\beta_{\varepsilon+2\eta}(\mathrm{P}_{XY}, \mathrm{Q}_X\mathrm{Q}_Y) + 2\log(1/\eta),$$

which along with the inequality above and Lemma 7 yields

$$l \ge \lambda_{\min} + \log\left(\mathbb{P}\left(\log\frac{\mathrm{P}_{XY}\left(X,Y\right)}{\mathrm{Q}_X\left(X\right)\mathrm{Q}_Y\left(Y\right)} < \lambda\right) - \varepsilon - 2\eta\right)_+ - \lambda - 4\log(1/\eta) - 1.$$

The claimed bound follows upon choosing $\lambda = \lambda_{\max} - \gamma$ and using assumption (8). □

C. Converse bound for all distributions

The shortcoming of Theorem 8 is the Δ-loss, which is negligible only if $\lambda_{\max} \approx \lambda_{\min}$. To circumvent this loss, we divide the spectrum of P_{XY} into slices such that, conditioned on any slice, the distribution is almost uniform with a small margin Δ. To lower bound the worst-case communication complexity of a given protocol, we identify a particular slice where appropriately many bits are communicated; the required slice is selected using Lemma 9 below.

Given $\lambda_{\max}$, $\lambda_{\min}$, and $\Delta > 0$, let N be as in (3), $\mathcal{T}_0$ be as in (5), and λ_i and $\mathcal{T}_i$, too, be as defined there, with $h_{\mathrm{P}_{X|Y}}(x|y)$ replaced by $h_{\mathrm{P}_{XY}}(xy)$ in those definitions. Let random variable J take the value j when $\{(X,Y) \in \mathcal{T}_j\}$. For a protocol Π attaining ε-DE, denote

$$\begin{aligned} \mathcal{E}_{\mathtt{correct}} &:= \{X = \hat{X}, Y = \hat{Y}\}, \\ \mathcal{E}_\gamma &:= \{h(X \triangle Y) \ge \gamma\}, \\ \mathcal{E}_j &:= \mathcal{E}_{\mathtt{correct}} \cap \mathcal{T}_0^c \cap \mathcal{E}_\gamma \cap \{J = j\}, \quad 1 \le j \le N, \\ \mathrm{P}_\gamma &:= \mathrm{P}_{XY}\left(\mathcal{E}_\gamma\right). \end{aligned} \tag{9}$$

Lemma 9. *There exists an index* $1 \leq j \leq N$ *such that* $\mathrm{P}_J(j) > 1/N^2$ *and*

$$\mathrm{P}_{XY|J}(\mathcal{E}_j \mid j) \geq \left(\mathrm{P}_\gamma - \varepsilon - \mathrm{P}_{XY}(\mathcal{T}_0) - \frac{1}{N}\right).$$

Proof. Let $\mathcal{J}_1$ be the set of indices $1 \leq j \leq N$ such that $\mathrm{P}_J(j) > 1/N^2$, and let $\mathcal{J}_2 = \{1, ..., N\} \setminus \mathcal{J}_1$. Note that $\mathrm{P}_J(\mathcal{J}_2) \leq 1/N$. Therefore,

$$\begin{aligned}
\mathrm{P}_\gamma - \varepsilon - \mathrm{P}_{XY}(\mathcal{T}_0) &\leq \mathbb{P}\left(\mathcal{E}_{\texttt{correct}} \cap \mathcal{T}_0^c \cap \mathcal{E}_\gamma\right) \\
&\leq \sum_{j \in \mathcal{J}_1} \mathrm{P}_J(j)\, \mathrm{P}_{XY|J}(\mathcal{E}_j \mid j) + \mathrm{P}_J(\mathcal{J}_2) \\
&\leq \max_{j \in \mathcal{J}_1} \mathrm{P}_{XY|J}(\mathcal{E}_j \mid j) + \frac{1}{N}.
\end{aligned}$$

Thus, the maximizing $j \in \mathcal{J}_1$ on the right satisfies the claimed properties. $\square$

We now state our main converse bound.

Theorem 10 (Single-shot converse)**.** *For* $0 \leq \varepsilon < 1$, $0 < \eta < 1 - \varepsilon$, *and parameters* Δ, N *as above, the following lower bound on* $L_\varepsilon(X, Y)$ *holds:*

$$L_\varepsilon(X, Y) \geq \gamma + 3\log\left(\mathrm{P}_\gamma - \varepsilon - \mathrm{P}_{XY}(\mathcal{T}_0) - \frac{1}{N}\right)_+ + \log(1 - 2\eta) - \Delta - 6\log N - 4\log\frac{1}{\eta} - 1.$$

Proof. Let j satisfy the properties stated in Lemma 9. The basic idea is to apply Theorem 8 to $\mathrm{P}_{XY|\mathcal{E}_j}$, where $\mathrm{P}_{XY|\mathcal{E}_j}$ denotes the conditional distributions on X, Y given the event $\mathcal{E}_j$.

First, we have

$$\mathrm{P}_{XY|\mathcal{E}_j}(x, y) \geq \mathrm{P}_{XY}(x, y). \tag{10}$$

Furthermore, denoting $\alpha = \mathrm{P}_\gamma - \varepsilon - \mathrm{P}_{XY}(\mathcal{T}_0) - 1/N$ and noting $\mathrm{P}_J(j) > 1/N^2$, we have for all $(x, y) \in \mathcal{E}_j$ that

$$\begin{aligned}
\mathrm{P}_{XY|\mathcal{E}_j}(x, y) &\leq \frac{1}{\alpha}\mathrm{P}_{XY|J=j}(x, y) &(11)\\
&\leq \frac{N^2}{\alpha}\mathrm{P}_{XY}(x, y), &(12)
\end{aligned}$$

where $\mathrm{P}_{XY|J=j}$ denotes the conditional distributions on X, Y given $\{J = j\}$. Thus, (10) and (12) together imply, for all $(x, y) \in \mathcal{E}_j$,

$$\lambda_j + \log\alpha - 2\log N \leq -\log \mathrm{P}_{XY|\mathcal{E}_j}(x, y) \leq \lambda_j + \Delta,$$

$i.e.$, $\mathrm{P}_{XY|\mathcal{E}_j}$ is almost uniform with margin $\Delta - \log \alpha + 2\log N$ ($cf.$ (8)). Also, note that (12) implies

$$\begin{aligned}
&\mathrm{P}_{XY|\mathcal{E}_j}\left(-\log \frac{\mathrm{P}_{XY|\mathcal{E}_i}(X,Y)^2}{\mathrm{P}_X(X)\,\mathrm{P}_Y(Y)} \geq \gamma + 2\log \alpha - 4\log N\right)\\
&\geq \mathrm{P}_{XY|\mathcal{E}_j}\left(-\log \frac{\mathrm{P}_{XY}(X,Y)^2}{\mathrm{P}_X(X)\,\mathrm{P}_Y(Y)} \geq \gamma\right)\\
&= \mathrm{P}_{XY|\mathcal{E}_i}(\mathcal{E}_\gamma)\\
&= 1,
\end{aligned}$$

where the final equality holds by the definition of $\mathcal{E}_\gamma$ in (9). Moreover,

$$\mathrm{P}_{XY|\mathcal{E}_j}\left(X = \hat{X}, Y = \hat{Y}\right) = 1.$$

Thus, the proof is completed by applying Theorem 8 to $\mathrm{P}_{XY|\mathcal{E}_j}$ with $\mathrm{Q}_X = \mathrm{P}_X$ and $\mathrm{Q}_Y = \mathrm{P}_Y$, and $\Delta - \log \alpha + 2\log N$ in place of Δ. □

D. Converse bound for simple communication protocol

We close by noting a lower bound for the length of communication when we restrict to simple communication. For simplicity assume that the joint distribution P_{XY} is indecomposable, $i.e.$, the *maximum common function* of X and Y is a constant (see [11]) and the parties can't agree on even a single bit without communicating ($cf.$ [34]). The following bound holds by a standard converse argument using the information spectral method ($cf.$ [12, Lemma 7.2.2]).

Proposition 11. *For $0 \leq \varepsilon < 1$, we have*

$$L_\varepsilon^{\mathrm{s}}(X,Y) \geq \inf\left\{l_1 + l_2 : \forall \delta > 0,\quad \mathbb{P}\Big(h(X|Y) > l_1 + \delta \text{ or } h(Y|X) > l_2 + \delta\Big) \leq \varepsilon + 2\cdot 2^{-\delta}\right\}.$$

Proof: Since randomization (local or shared) does not help in improving the length of communication ($cf.$ [19, Chapter 3]) we can restrict to deterministic protocols. Then, since P_{XY} is indecomposible, both parties have to predetermine the lengths of messages they send; letl_1 and l_2, respectively, be the length of message sent by the first and the second party. For $\delta > 0$, let

$$\begin{aligned}
\mathcal{T}_1 &:= \Big\{(x,y) : -\log \mathrm{P}_{X|Y}(x|y) \leq l_1 + \delta\Big\},\\
\mathcal{T}_2 &:= \Big\{(x,y) : -\log \mathrm{P}_{Y|X}(y|x) \leq l_2 + \delta\Big\},
\end{aligned}$$

and $\mathcal{T} := \mathcal{T}_1 \cap \mathcal{T}_2$. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be the set of all (x,y) such that party 2 and party 1 correctly recover

x and y, respectively, and let $\mathcal{A} := \mathcal{A}_1 \cap \mathcal{A}_2$. Then, for any simple communication protocol that attains ε-DE, we have

$$\begin{aligned}
\mathrm{P}_{XY}\left(\mathcal{T}^c\right) &= \mathrm{P}_{XY}\left(\mathcal{T}^c \cap \mathcal{A}^c\right) + \mathrm{P}_{XY}\left(\mathcal{T}^c \cap \mathcal{A}\right) \\
&\leq \mathrm{P}_{XY}\left(\mathcal{A}^c\right) + \mathrm{P}_{XY}\left(\mathcal{T}_1^c \cap \mathcal{A}\right) + \mathrm{P}_{XY}\left(\mathcal{T}_2^c \cap \mathcal{A}\right) \\
&\leq \varepsilon + \mathrm{P}_{XY}\left(\mathcal{T}_1^c \cap \mathcal{A}_1\right) + \mathrm{P}_{XY}\left(\mathcal{T}_2^c \cap \mathcal{A}_2\right) \\
&\leq \varepsilon + 2 \cdot 2^{-\delta},
\end{aligned}$$

where the last inequality follows by a standard argument ($cf.$ [12, Lemma 7.2.2]) as follows:

$$\begin{aligned}
\mathrm{P}_{XY}\left(\mathcal{T}_1^c \cap \mathcal{A}_1\right) &\leq \sum_y \mathrm{P}_Y\left(y\right) \mathrm{P}_{X|Y}\left(\mathcal{T}_1^c \cap \mathcal{A}_1 | y\right) \\
&\leq \sum_y \mathrm{P}_Y\left(y\right) |\{x : (x,y) \in \mathcal{A}_1\}| 2^{-l_1-\delta} \\
&\leq \sum_y \mathrm{P}_Y\left(y\right) |\{x : (x,y) \in \mathcal{A}_1\}| 2^{-l_1-\delta} \\
&\leq \sum_y \mathrm{P}_Y\left(y\right) 2^{-\delta} \\
&= 2^{-\delta},
\end{aligned}$$

and similarly for $\mathrm{P}_{XY}\left(\mathcal{T}_2^c \cap \mathcal{A}_2\right)$; the desired bound follows. ■

VI. General sources

While the best rate of communication required for two parties to exchange their data is known [7], and it can be attained by simple (noninteractive) Slepian-Wolf compression on both sides, the problem remains unexplored for general sources. In fact, the answer is completely different in general and, as pointed-out in the Introduction, simple Slepian-Wolf compression is suboptimal.

Formally, let (X_n, Y_n) with joint distribution[7] $\mathrm{P}_{X_nY_n}$ be a sequence of sources. We need the following concepts from the information spectrum method; see [12] for a detailed account. For random variables $(\mathbf{X}, \mathbf{Y}) = \{(X_n, Y_n)\}_{n=1}^{\infty}$, the the *inf entropy rate* $\underline{H}(\mathbf{XY})$ and the *sup entropy rate* $\overline{H}(\mathbf{XY})$ are defined as follows:

$$\underline{H}(\mathbf{XY}) = \sup\left\{\alpha \mid \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n} h(X_nY_n) < \alpha\right) = 0\right\},$$

[7]The distributions $\mathrm{P}_{X_nY_n}$ need not satisfy the consistency conditions.

$$\overline{H}(\mathbf{XY}) = \inf\left\{\alpha \mid \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}h(X_nY_n) > \alpha\right) = 0\right\};$$

the *sup-conditional entropy rate* $\overline{H}(\mathbf{X}|\mathbf{Y})$ is defined analogously by replacing $h(X_nY_n)$ with $h(X_n|Y_n)$. To state our result, we also need another quantity defined by a limit-superior in probability, namely the *sup sum conditional entropy rate*, given by

$$\begin{aligned}&\overline{H}(\mathbf{X}\triangle\mathbf{Y})\\&= \inf\left\{\alpha \mid \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}h(X_n\triangle Y_n) > \alpha\right) = 0\right\}.\end{aligned}$$

The result below characterizes $R^*(\mathbf{X},\mathbf{Y})$ (see Definition 2).

Theorem 12. *For a sequence of sources* $(\mathbf{X},\mathbf{Y}) = \{(X_n,Y_n)\}_{n=1}^{\infty}$,

$$R^*(\mathbf{X},\mathbf{Y}) = \overline{H}(\mathbf{X}\triangle\mathbf{Y}).$$

Proof. The claim follows from Theorems 2 and 10 on choosing the spectrum slicing parameters $\lambda_{\min}, \lambda_{\max}$, and Δ appropriately.

Specifically, using Theorem 2 with

$$\begin{aligned}\lambda_{\min} &= n(\underline{H}(\mathbf{X},\mathbf{Y}) - \delta),\\ \lambda_{\max} &= n(\overline{H}(\mathbf{X},\mathbf{Y}) + \delta),\\ \Delta &= \sqrt{\lambda_{\max} - \lambda_{\min}}\\ &= N\\ \eta &= \Delta,\\ l_{\max} &= n(\overline{H}(\mathbf{X}\triangle\mathbf{Y}) + \delta) + 3\Delta + 1\\ &= n(\overline{H}(\mathbf{X}\triangle\mathbf{Y}) + \delta) + O(\sqrt{n}),\end{aligned}$$

where $\delta > 0$ is arbitrary, we get a communication protocol of rate $\overline{H}(\mathbf{X}\triangle\mathbf{Y}) + \delta + O(n^{-1/2})$ attaining ε_n-DE with $\varepsilon_n \to 0$. Since $\delta > 0$ is arbitrary, $R^*(\mathbf{X},\mathbf{Y}) \leq \overline{H}(\mathbf{X}\triangle\mathbf{Y})$.

For the other direction, given a sequence of protocols attaining ε_n-DE with $\varepsilon_n \to 0$. Let

$$\begin{aligned}\lambda_{\min} &= n(\underline{H}(\mathbf{X},\mathbf{Y}) - \Delta),\\ \lambda_{\max} &= n(\overline{H}(\mathbf{X},\mathbf{Y}) + \Delta),\end{aligned}$$

and so, $N = O(n)$. Using Theorem 10 with

$$\gamma = n(\overline{H}(\mathbf{X}\triangle\mathbf{Y}) - \delta)$$

for arbitrarily fixed $\delta > 0$, we get for n sufficiently large that

$$L_{\varepsilon_n}(X_n, Y_n) \geq n(\overline{H}(\mathbf{X}\triangle\mathbf{Y}) - \delta) + o(n).$$

Since $\delta > 0$ is arbitrary, the proof is complete. $\square$

VII. Strong converse and second-order asymptotics

We now turn to the case of IID observations (X^n, Y^n) and establish the second-order asymptotic term in $L_\varepsilon(X^n, Y^n)$.

Theorem 13. *For every* $0 < \varepsilon < 1$*,*

$$L_\varepsilon\left(X^n, Y^n\right) = nH(X\triangle Y) + \sqrt{nV}Q^{-1}(\varepsilon) + o(\sqrt{n}).$$

Proof. As before, we only need to choose appropriate parameters in Theorems 2 and 10. Let T denote the third central moment of the random variable $h(X\triangle Y)$.

For the achievability part, note that for IID random variables (X^n, Y^n) the spectrum of $P_{X^nY^n}$ has width $O(\sqrt{n})$. Therefore, the parameters Δ and N can be $O(n^{1/4})$. Specifically, by standard measure concentration bounds (for bounded random variables), for every $\delta > 0$ there exists a constant c such that with $\lambda_{\max} = nH(XY) + c\sqrt{n}$ and $\lambda_{\min} = nH(XY) - c\sqrt{n}$,

$$\mathbb{P}\left((X^n, Y^n) \in \mathcal{T}_0\right) \leq \delta.$$

For

$$\lambda_n = nH(X\triangle Y) + \sqrt{nV}Q^{-1}\left(\varepsilon - 2\delta - \frac{T^3}{2V^{3/2}\sqrt{n}}\right),$$

choosing $\Delta = N = \eta = \sqrt{2c}n^{1/4}$, and $l_{\max} = \lambda_n + 3\Delta + 1$ in Theorem 2, we get a protocol of length $l_{\max}$ satisfying

$$\mathbb{P}\left(X \neq \hat{X},\ \text{or}\ Y \neq \hat{Y}\right) \leq \mathbb{P}\left(\sum_{i=1}^n h(X_i\triangle Y_i) > \lambda_n\right) + 2\delta,$$

for n sufficiently large. Thus, the Berry-Esséen theorem (*cf.* [10]) and the observation above gives a protocol of length $l_{\max}$ attaining ε-DE. Therefore, using the Taylor approximation of $Q(\cdot)$ yields the achievability of the claimed protocol length; we skip the details of this by-now-standard argument (see, for instance, [25]).

Similarly, the converse follows by Theorem 10 and the Berry-Esséen theorem upon choosing $\lambda_{\max}$, $\lambda_{\min}$, and N as in the proof of converse part of Theorem 12 when λ_n is chosen to be

$$\begin{aligned}\lambda_n &= nH(X\triangle Y) + \sqrt{nV}Q^{-1}\left(\varepsilon - 2\frac{1}{N} - \frac{T^3}{2V^{3/2}\sqrt{n}}\right)\\ &= nH(X\triangle Y) + \sqrt{nV}Q^{-1}\left(\varepsilon\right) + O(\log n),\end{aligned}$$

where the final equality is by the Taylor approximation of $Q(\cdot)$. □

In the previous section, we saw that interaction is necessary to attain the optimal first order asymptotic term in $L_\varepsilon(X_n, Y_n)$ for a mixture of IID random variables. In fact, even for IID random variables interaction is need to attain the correct second order asymptotic term in $L_\varepsilon(X^n, Y^n)$, as shown by the following example.

Example 1. Consider random variables X and Y with an indecomposable joint distribution P_{XY} such that the matrix

$$\mathbf{V} = \mathrm{Cov}([-\log \mathrm{P}_{X|Y}\left(X|Y\right), -\log \mathrm{P}_{Y|X}\left(Y|X\right)])$$

$\mathbf{V}$ is nonsingular. For IID random variables (X^n, Y^n) with common distribution P_{XY}, using Proposition 11 and a multidimensional Berry-Esséen theorem ($cf.$ [30]), we get that the second-order asymptotic term for the minimum length of simple communication for ε-DE is given by[8]

$$L_\varepsilon^{\mathrm{s}}(X^n, Y^n) = nH(X\triangle Y) + \sqrt{n}D_\varepsilon + o(\sqrt{n}),$$

where

$$D_\varepsilon := \inf\left\{r_1 + r_2 : \mathbb{P}\left(Z_1 \le r_1, Z_2 \le r_2\right) \ge 1 - \varepsilon\right\},$$

for Gaussian vector $\mathbf{Z} = [Z_1, Z_2]$ with mean $[0, 0]$ and covariance matrix $\mathbf{V}$. Since $\mathbf{V}$ is nonsingular,[9] we have

$$\sqrt{V}Q^{-1}(\varepsilon) = \inf\left\{r : \mathbb{P}\Big(Z_1 + Z_2 \le r\Big) \ge 1 - \varepsilon\right\}$$

[8]The achievability part can be derived by a slight modification of the arguments in [21],[12, Lemma 7.2.1].

[9]For instance, when X is uniform random variable on $\{0, 1\}$ and Y is connected to X via a binary symmetric channel, the covariance matrix $\mathbf{V}$ is singular and interaction does not help.

$$< D_\varepsilon.$$

Therefore, $L_\varepsilon(X^n, Y^n)$ has strictly smaller second order term than $L_\varepsilon^s(X^n, Y^n)$, and interaction is necessary for attaining the optimal second order term in $L_\varepsilon(X^n, Y^n)$.

VIII. Discussion

We have presented an interactive data exchange protocol and a converse bound which shows that, in a single-shot setup, the parties can exchange data using roughly $h(X\Delta Y)$ bits when the parties observe X and Y. Our analysis is based on the information spectrum approach. In particular, we extend this approach to enable handling of interactive communication. A key step is the *spectrum slicing* technique which allows us to split a nonuniform distribution into almost uniform "spectrum slices". Another distinguishing feature of this work is our converse technique which is based on extracting a secret key from the exchanged data and using an upper bound for the rate of this secret key. In effect, this falls under the broader umbrella of *common randomness decomposition* methodology presented in [31]. As a consequence, we obtain both the optimal rate for data exchange for general sources as well as the precise second-order asymptotic term for IID observations (which in turn implies a strong converse). Interestingly, none of these optimal results can be obtained by simple communication and interaction is necessary, in general.

Another asymptotic regime, which was not considered in this paper, is the error exponent regime where we seek to characterize the largest possible rate of exponential decay of error probability with blocklength for IID observations. Specifically, denoting by $\mathrm{P}_{\mathtt{err}}\left(l|X,Y\right)$ the least probability of error ε that can be attained for data exchange by communicating less than l bits, $i.e.$,

$$\mathrm{P}_{\mathtt{err}}\left(l|X,Y\right) := \inf\{\varepsilon : L_\varepsilon(X,Y) \leq l\},$$

we seek to characterize the limit of

$$-\frac{1}{n}\log \mathrm{P}_{\mathtt{err}}\left(2^{nR}|X^n, Y^n\right).$$

The following result is obtained using a slight modification of our single-shot protocol for data exchange where the slices of the spectrum $\mathcal{T}_i$ in (6) are replaced with type classes. For a fixed rate $R \geq 0$, our modified protocol enables data exchange, with large probability, for every $(\mathbf{x}, \mathbf{y})$ with joint type $\mathrm{P}_{\overline{X}\,\overline{Y}}$ such that (roughly)

$$R > H(\overline{X} \triangle \overline{Y}).$$

The converse part follows from the strong converse of Result 4, together with a standard measure change argument ($cf.$ [6]). The formal proof is very similar to other arguments in this paper and is omitted.

Result 6 (Error Exponent Behaviour). *For a given rate $R > H(X \triangle Y)$, define*

$$E_{\mathtt{r}}(R) := \min_{\mathrm{Q}_{\overline{X}\,\overline{Y}}} \left[D(\mathrm{Q}_{\overline{X}\,\overline{Y}} \| \mathrm{P}_{XY}) + |R - H(\overline{X} \triangle \overline{Y})|^{+} \right]$$

and

$$E_{\mathtt{sp}}(R) := \inf_{\mathrm{Q}_{\overline{X}\,\overline{Y}} \in \mathcal{Q}(R)} D(\mathrm{Q}_{\overline{X}\,\overline{Y}} \| \mathrm{P}_{XY}),$$

where $|a|^{+} = \max\{a, 0\}$ and

$$\mathcal{Q}(R) := \left\{ \mathrm{Q}_{\overline{X}\,\overline{Y}} : R < H(\overline{X} \triangle \overline{Y}) \right\}.$$

Then, it holds that

$$\liminf_{n \to \infty} -\frac{1}{n} \log \mathrm{P}_{\mathtt{err}} \left(2^{nR} | X^n, Y^n \right) \geq E_{\mathtt{r}}(R)$$

and that

$$\limsup_{n \to \infty} -\frac{1}{n} \log \mathrm{P}_{\mathtt{err}} \left(2^{nR} | X^n, Y^n \right) \leq E_{\mathtt{sp}}(R).$$

$E_{\mathtt{r}}(R)$ and $E_{\mathtt{sp}}(R)$, termed the *random coding exponent* and the *sphere-packing exponent*, may not match in general. However, when R is sufficiently close to $H(X \triangle Y)$, the two bounds can be shown to coincide. In fact, in the Appendix we exhibit an example where the optimal error exponent attained by interactive protocols is strictly larger than that attained by simple communication. Thus, in the error exponent regime, too, interaction is strictly necessary.

APPENDIX

Consider the following source: $\mathcal{X}$ and $\mathcal{Y}$ are both binary, and P_{XY} is given by

$$\mathrm{P}_{XY}(0,0) = \mathrm{P}_{XY}(1,0) = \mathrm{P}_{XY}(1,1) = \frac{1}{3},$$

that is, X and Y are connected by a Z-channel. To evaluate $E_{\mathtt{sp}}(R)$, without loss of generality, we can assume that $\mathrm{Q}_{\overline{X}\,\overline{Y}}(0,1) = 0$ (since otherwise $D(\mathrm{Q}_{\overline{X}\,\overline{Y}} \| \mathrm{P}_{XY}) = \infty$). Let us consider the following parametrization:

$$\mathrm{Q}_{\overline{X}\,\overline{Y}}(0,0) = u, \;\; \mathrm{Q}_{\overline{X}\,\overline{Y}}(1,0) = 1 - u - v, \;\; \mathrm{Q}_{\overline{X}\,\overline{Y}}(1,1) = v,$$

where $0 \le u, v \le 1$. Then, we have

$$D(\mathrm{Q}_{\overline{X}\,\overline{Y}} \| \mathrm{P}_{XY}) = \log 3 - H(\{u, 1-u-v, v\}) \tag{13}$$

and

$$\begin{aligned} H(\overline{X}|\overline{Y}) + H(\overline{Y}|\overline{X}) &= \kappa(u,v) \\ &:= (1-v)h\left(\frac{u}{1-v}\right) + (1-u)h\left(\frac{v}{1-u}\right). \end{aligned}$$

When the rate R is sufficiently close to $H(X \triangle Y) = \kappa(1/3, 1/3) = 4/3$, the set $\mathcal{Q}(R)$ is not empty.[10] Since (13) and $\kappa(u,v)$ are both symmetric with respect to u and v and (13) and $\mathcal{Q}(R)$ are convex function and convex set, respectively, the optimal solution (u^*, v^*) in the infimum of $E_{\mathtt{sp}}(R)$ satisfies $u^* = v^*$. Furthermore, since $R > \kappa(1/3, 1/3)$, we also have $u^* = v^* \neq 1/3$.

Note that for R sufficiently close to $H(X \triangle Y)$, $E_{\mathtt{sp}}(R)$ can be shown to equal $E_{\mathtt{r}}(R)$. Thus, to show that a simple communication is strictly suboptimal for error exponent, it suffices to show that $E_{\mathtt{sp}}(R) > E^{\mathrm{s}}_{\mathtt{sp}}(R)$, where the latter quantity $E^{\mathrm{s}}_{\mathtt{sp}}(R)$ corresponds to the sphere packing bound for error exponent using simple communication and is given by

$$E^{\mathrm{s}}_{\mathtt{sp}}(R) := \max_{\substack{(R_1,R_2): \\ R_1+R_2 \le R}} \inf_{\mathrm{Q}_{\overline{X}\,\overline{Y}} \in \mathcal{Q}^{\mathrm{s}}(R_1,R_2)} D(\mathrm{Q}_{\overline{X}\,\overline{Y}} \| \mathrm{P}_{XY})$$

and

$$\mathcal{Q}^{\mathrm{s}}(R_1, R_2) := \left\{ \mathrm{Q}_{\overline{X}\,\overline{Y}} : R_1 < H(\overline{X}|\overline{Y}) \text{ or } R_2 < H(\overline{Y}|\overline{X}) \right\}.$$

Since the source is symmetric with respect to X and Y, for evaluating $E^{\mathrm{s}}_{\mathtt{sp}}(R)$ we can assume without loss of generality that $R_1 \ge R_2$. Let $u^\dagger := u^*$ and $v^\dagger := \frac{1-u^\dagger}{2}$ so that $\frac{v^\dagger}{1-u^\dagger} = \frac{1}{2}$. Let $\mathrm{Q}^\dagger_{\overline{X}\,\overline{Y}}$ be the distribution that corresponds to $(u^\dagger, v^\dagger)$. Note that $\mathrm{Q}^\dagger_{\overline{X}\,\overline{Y}}$ satisfies

$$\begin{aligned} H(\overline{Y}|\overline{X}) &= (1-u^\dagger)h\left(\frac{v^\dagger}{1-u^\dagger}\right) \\ &> (1-u^*)h\left(\frac{v^*}{1-u^*}\right) \\ &\ge \frac{R}{2} \\ &\ge R_2, \end{aligned}$$

[10]In fact, we can check that $\left.\frac{\kappa(z,z)}{dz}\right|_{z=1/3} = -1$, and thus the function $\kappa(z,z)$ takes its maximum away from $(1/3, 1/3)$.

and so, $\mathrm{Q}^{\dagger}_{\overline{X}\,\overline{Y}} \in \mathcal{Q}^{\mathrm{s}}(R_1, R_2)$. For this choice of $\mathrm{Q}^{\dagger}_{\overline{X}\,\overline{Y}}$, we have

$$\begin{aligned} D(\mathrm{Q}^{*}_{\overline{X}\,\overline{Y}} \| \mathrm{P}_{XY}) &= \log 3 - H(\{u^*, 1-u^*-v^*, v^*\}) \\ &= \log 3 - h(1-u^*) - (1-u^*)h\left(\frac{v^*}{1-u^*}\right) \\ &> \log 3 - h(1-u^\dagger) - (1-u^\dagger)h\left(\frac{v^\dagger}{1-u^\dagger}\right) \\ &= D(\mathrm{Q}^{\dagger}_{\overline{X}\,\overline{Y}} \| \mathrm{P}_{XY}), \end{aligned}$$

which implies $E_{\mathtt{sp}}(R) > E^{\mathrm{s}}_{\mathtt{sp}}(R)$.

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] Y. Altuğ and A. B. Wagner, "Feedback can improve the second-order coding performance in discrete memoryless channels," *ISIT*, 2014.

[3] 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.

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

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory. 2nd edition*. John Wiley & Sons Inc., 2006.

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

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

[8] S. C. Draper, "Universal incremental slepian-wolf coding," *Proc. Conference on Communication, Control, and Computing (Allerton)*, October 2004.

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

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

[11] 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.

[12] T. S. Han, *Information-Spectrum Methods in Information Theory [English Translation]*. Springer, 2003.

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

[14] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4619–4637, Oct 2008.

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

[16] ——, “Secret key agreement: General capacity and second-order asymptotics,” *ISIT*, 2014.

[17] 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.

[18] R. Impagliazzo, L. A. Levin, and M. Luby, “Pseudo-random generation from one-way functions,” *Proc. Annual Symposium on Theory of Computing*, pp. 12–24, 1989.

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

[20] 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.

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

[22] A. Orlitsky and A. El Gamal, “Communication with secrecy constraints,” *STOC*, pp. 217–224, 1984.

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

[24] Y. Polyanskiy, “Channel coding: non-asymptotic fundamental limits,” *Ph. D. Dissertation, Princeton University*, 2010.

[25] 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.

[26] R. Renner, “Security of quantum key distribution,” *Ph. D. Dissertation, ETH Zurich*, 2005.

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

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

[29] 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.

[30] V. Y. F. Tan and O. Kosut, “On the dispersions of three network information theory problems,” *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, 2014.

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

[32] 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.

[33] ——, “Converses for secret key agreement and secure computing,” *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4809–4827, Sept 2015.

[34] H. S. Witsenhausen, “On sequences of pairs of dependent random variables,” *Siam J. Appl. Math.*, vol. 28, no. 1, pp. 100–113, January 1975.

[35] 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.