Information Complexity Density and Simulation of Protocols

[Extended Abstract] *

Himanshu Tyagi
Indian Institute of Science, Bangalore
htyagi@ece.iisc.ernet.in

Shaileshh Venkatakrishnan
University of Illinois, Urbana-Champaign
bjjvnkt2@illinois.edu

Pramod Viswanath
University of Illinois, Urbana-Champaign
pramodv@illinois.edu

Shun Watanabe
Tokyo University of Agriculture and Technology
shunwata@cc.tuat.ac.jp

ABSTRACT

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

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

Categories and Subject Descriptors

F.2.0 [**Theory of Computation**]: Analysis of algorithms and problem complexity—*General*

Keywords

Information complexity, simulation of protocols, interactive protocols

1. INTRODUCTION

Two parties observing random variables X and Y seek to run an interactive protocol π with inputs X and Y. The parties have access to private as well as shared public randomness. What is the minimum number of bits that they must exchange in order to simulate π to within a fixed statistical distance ε? This question is of importance to the theoretical computer science as well as the information theory communities. On the one hand, it is related closely to the communication complexity problem [44], which in turn is an important tool for deriving lower bounds for computational complexity [24] and for space complexity of streaming algorithms [2]. On the other hand, it is a significant generalization of the classical information theoretic problem of distributed data compression [38], replacing data to be compressed with an interactive protocol and allowing interactive communication as opposed to the usual one-sided communication.

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

*A full version of this paper is available at http://eccc.hpi-web.de/report/2015/070/.

ITCS'16, January 14–16, 2016, Cambridge, MA, USA.
 ISBN 978-1-4503-4057-1/16/01 ...$15.00.
DOI: http://dx.doi.org/10.1145/2840728.2840754.

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

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.* [37, 12]). For a protocol π with communication complexity $\|\pi\|$ (the depth of the binary protocol tree), a simulation protocol requiring $\tilde{\mathcal{O}}(\sqrt{\mathtt{IC}(\pi)\|\pi\|})$ bits of communication was given in [4] and one requiring $2^{\mathcal{O}(\mathtt{IC}(\pi))}$ bits of communication was given in [7]. A general version of the simulation problem was considered in [46], but only bounded round simulation protocols were considered. Interestingly, it was shown in [8] that the amortized distributional communication complexity of simulating n copies of a protocol π for vanishing simulation error is bounded above by[2] $\mathtt{IC}(\pi)$. While a matching lower bound was also derived in [8], it is not valid in our context – [8] considered function computation and used a coordinate-wise error criterion. Nevertheless, we can readily modify the lower bound argument in [8] and use the continuity of conditional mutual information to formally obtain the required lower bound and thereby a characterization of the amortized distributional communication complexity for vanishing simulation error. Specifically, denoting by $D(\pi^n)$ the distributional communication complexity of simulating n copies of a protocol π with vanishing simulation error, we have

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

Perhaps motivated by this characterization, or a folklore version of it, the research in this area has focused on designing simulation protocols for π requiring communication of length depending on $\mathtt{IC}(\pi)$; the results cited above belong to this category as well. However, the central role of $\mathtt{IC}(\pi)$ in the distributional communication complexity of protocol simulation is far from settled and many important questions remain unanswered. For instance, (a) does $\mathtt{IC}(\pi)$ suffice to capture the dependence of distributional communication complexity on the simulation error ε? (b) Does information complexity have an operational role in simulating π^n besides being the leading asymptotic term? (c) How about the simulation of more complicated protocols such as a mixture $\pi_{\mathtt{mix}}$ 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 [37] and $H(X|Y)$ in the transmission of X^n by the first to the second party with access to Y^n [38]. The questions raised above have been addressed for these classical problems (*cf.* [19]). In this paper, we answer these questions for simulation of interactive protocols. In particular, we answer all these questions in the negative by exhibiting another quantity that plays such a fundamental role and can differ from information complexity significantly. To this end, we introduce the notion of *information complexity density* of a protocol π with inputs X and Y generated from a fixed distribution P_{XY}.

[2]Braverman and Rao actually used their general simulation protocol as a tool for deriving the amortized distributional communication complexity of function computation. This result was obtained independently by Ma and Ishwar in [26] using standard information theoretic techniques.

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 $\mathtt{IC}(\pi) = \mathbb{E}[\mathtt{ic}(\Pi; X, Y)]$. We show that it is the *ε-tail of the information complexity density* $\mathtt{ic}(\Pi; X, Y)$, *i.e.*, the supremum[3] over values of λ such that $\Pr(\mathtt{ic}(\Pi; X, Y) > \lambda) > \varepsilon$, 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

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

This condition holds, for instance, in the amortized regime considered in [8]. 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 4.3, 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 $\tilde{\mathcal{O}}(n^{-2})$.

1.1 Summary of results

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

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 4.3. On evaluating our lower bound for this protocol, for $\varepsilon = 1/n^3$ we get $D_\varepsilon(\pi) = \Omega(n)$ which is far more than $2^{\mathtt{IC}(\pi)}$ since $\mathtt{IC}(\pi) = \tilde{\mathcal{O}}(n^{-2})$. Remarkably, [18, 17] 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 $D_\varepsilon(\pi) \leq \mathcal{O}(2^{\mathtt{IC}}(\pi))$ given in [7]. Our simple example shows a much stronger

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

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 simulation protocol, which is of independent interest. For a protocol π with bounded rounds of interaction, using our proposed protocol we can show that $D_\varepsilon(\pi) \lesssim \lambda_\varepsilon$. Much as the protocol of [8], our simulation protocol simulates one round at a time, and thus, the slack in our upper bound does depend on the number of rounds.

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

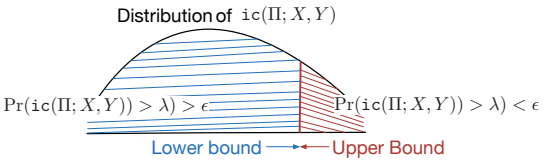

Figure 1: Illustration of lower and upper bounds for $D_\varepsilon(\pi)$

Amortized regime: second-order asymptotics. Denote by π^n the n-fold product protocol obtained by applying π to each coordinate (X_i, Y_i) for inputs X^n and Y^n. Consider the communication complexity $D_\varepsilon(\pi^n)$ of ε-simulating π^n for *independent and identically distributed* (IID) (X^n, Y^n) generated from P_{XY}^n. Using the bounds above, we can obtain the following sharpening of the results of [8]: 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[4] [8] or [46] 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 $\pi_\mathtt{h}^n$ if $\Pi_0 = \mathtt{h}$ and $\pi_\mathtt{t}^n$ if $\Pi_0 = \mathtt{t}$. What is the amortized communication complexity of simulating the mixed protocol $\pi_{\mathtt{mix,n}}$? Note that

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

[4]The proof in [8] uses the inequality $\mathtt{IC}(\pi) \leq \|\pi\|$, a multi-party extension of which is available in [13, 27].

Is it true that in the manner of [8] 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 $\mathtt{IC}(\pi_\mathtt{h}) \gg \mathtt{IC}(\pi_\mathtt{t})$.

1.2 Proof techniques

Proof for the lower bound. We present a new method for deriving lower bounds on distributional communication complexity. Our proof relies on a reduction argument that utilizes an ε-simulation to generate an information theoretically secure secret key for X and Y (for a definition of the latter, see [28, 1]). Heuristically, a protocol can be simulated using fewer bits of communication than its length because of the correlation in the observations X and Y. Due to this correlation, when simulating the protocol, the parties agree on more bits (generate more *common randomness*) than what they communicate. These extra bits can be extracted as an information theoretically secure secret key for the two parties using the *leftover hash lemma* (*cf.* [6, 36]). 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 [42, 41].

Protocol for the upper bound. We simulate a given protocol one round at a time. Simulation of each round consists of two subroutines: Interactive Slepian-Wolf compression and message reduction by public randomness. The first subroutine is an interactive version of the classical Slepian-Wolf compression [38] for sending X to an observer of Y which is of optimal instantaneous rate. The second subroutine uses an idea that appeared first in [35] (see, also, [30, 45]) and reduces the number of bits communicated in the first by realizing a portion of the required communication by the shared public randomness. This is possible since we are not required to recover a given random variable Π, but only simulate it to within a fixed statistical distance.

The proposed protocol is closely related to that proposed in [8]. However, there are some crucial differences. The protocol in [8], 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 [8] and relies on randomness extraction techniques. In particular, the protocol in [8] 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 [20] (see the textbook [19] for a detailed account). In this approach, the classical measures of information such as entropy and mutual information are viewed as expectations of the corresponding *information densities*, and the notion of "typical sets" is replaced by sets where these information densities are bounded uniformly. The set of values taken by an in-

formation density (such as $h(x) = -\log \mathrm{P}_X(x)$) is called its *spectrum*. Coding theorems of classical information theory consider IID repetitions and rely on the so-called the *asymptotic equipartition property* [11] which essentially corresponds to the concentration of spectrums on small intervals. For *single-shot* problems such concentrations are not available and we have to work with the whole span of the spectrum.

Our main technical contribution in this paper is the extension of the information spectrum method to handle interactive communication. Our results rely on the analysis of appropriately chosen information densities and, in particular, will rely on the spectrum of the information complexity density $\mathtt{ic}(\Pi; X, Y)$. As is usually the case, different components of our analysis require bounds on these information densities in different directions, which in turn renders our bounds loose and incurs a gap equal to the length of the corresponding information spectrum. To overcome this shortcoming, we use the *spectrum slicing* technique of Han [19][5] to divide the information spectrum into small portions with information densities closely bounded from both sides. While in our upper bounds spectrum slicing is used to carefully choose the parameters of the protocol, it is required in our lower bounds to identify a set of inputs where a given simulation will require a large number of bits to be communicated.

1.3 Organization

A formal statement of the problem, along with the necessary preliminaries, is given in the next section. Section **??** contains all our results. While the proofs of our general single-shot results are deferred to the full-version of the paper, proofs of the asymtotic results, derived using our single-shot results, are included in Section 4.

1.4 Notations

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

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

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

2. PROBLEM STATEMENT

Two parties observe correlated random variables X and Y, with Party 1 observing X and Party 2 observing Y, generated from a fixed distribution P_{XY} and taking values in finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. An *interactive protocol* π (for these two parties) consists of shared public randomness U, private randomness[6] $U_{\mathcal{X}}$ and $U_{\mathcal{Y}}$, and interactive communication $\Pi_1, ..., \Pi_r$. The parties communicate alternatively with Party 1 transmitting in the odd rounds and Party 2 in the even rounds. Specifically, Π_i is a string of bits determined by the previous transmissions $\Pi_1, ..., \Pi_{i-1}$ together with $(X, U_{\mathcal{X}}, U)$ for odd i and $(Y, U_{\mathcal{Y}}, U)$ for even i. For simplicity, we assume that the realizations of Π_i constitute a prefix-free code, *i.e.*, no realizations of Π_i is a prefix of another realization of Π_i. The number of rounds of communication r is a random stopping-time such that the event $\{r = t\}$ is determined by the transcript $\Pi_1, ..., \Pi_t$; we denote the overall transcript of the protocol[7] by Π. The length of a protocol π, $\|\pi\|$, is the maximum number of bits that are communicated in any execution of the protocol.

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

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

When we execute the protocol π above, the overall *view* of the parties consists of random variables $(XY\Pi\Pi)$, where the two Πs correspond to the transcript of the protocol seen by the two parties. A simulation of the protocol consists of another protocol which generates almost the same view as that of the original protocol. We are interested in the simulation of private coin protocols, using arbitrary[8] protocols; public coin protocols can be simulated by simulating for each fixed value of public randomness the resulting private coin protocol.

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

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

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

Definition 3. **Distributional communication complexity.** The ε-error distributional communication complexity $D_\varepsilon(\pi|\mathrm{P}_{XY})$ 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 $D_\varepsilon(\pi|\mathrm{P}_{XY})$ by $D_\varepsilon(\pi)$.

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

[5]The spectrum slicing technique was introduced in [19] to derive the error exponents of various problems for general sources and a rate-distortion function for general sources.

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

[7]We allow Π_i to be constant and allow it to depend only on the local observation (and not on the previous communication $\Pi_1, ..., \Pi_{i-1}$). This description of an interactive protocol is very general and is equivalent to the usual protocol-tree based description (*cf.* [4, 8]).

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

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

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

Therefore, a protocol is an ε-simulation of a deterministic protocol if and only if it computes the corresponding interactive function with probability of error less than ε. Furthermore, randomization does not help in this case, and it suffices to use deterministic simulation protocols. Thus, our results below provide tight bounds for distributional communication complexity of interactive functions and, in fact, of all functions which are *information theoretically securely computable* for the distribution P_{XY}, since computing these functions is tantamount to computing an interactive function [31] (see, also, [5, 25]).

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

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

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

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

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

3. MAIN RESULTS

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

3.1 Lower bound

We prove the following lower bound.

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

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

where the fudge parameters ε' and λ' depend on η as well as appropriately chosen information spectrums and will be described below in (4) *and* (5)*.*

The appearance of fudge parameters such as ε' and λ' in the bound above is not surprising 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 evalutation of error probability of single-party non-interactive compression problems (*cf.* [20, 19]).

When the fudge parameters ε' and λ' are negligible, the right-side of the bound above is close to ε-tail of $\mathtt{ic}(\Pi; X, Y)$. Indeed, the fudge parameters turn out to be negligible in many cases of interest. For instance, for the amortized case ε' can be chosen to be arbitrarily small. The parameter λ' is related to the length of the interval in which the underlying information densities lie with probability greater than $1-\varepsilon'$, the essential length of spectrums. For the amortized case with product protocols, by the central limit theorem the related essential spectrums are of length $\Lambda = \mathcal{O}(\sqrt{n})$ and $\lambda' = \log \Lambda$. On the other hand, λ_{ε} is $\mathcal{O}(n)$. Thus, the $\log n$ order fudge parameter λ' is negligible in this case. The same is true also for the example protocol in Appendix 4.3. Finally, it should be noted that similar fudge parameters are ubiquitous in single-shot bounds; for instance, see [19, Lemma 1.3.2].

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 (2), which is roughly equal to the largest value λ_{ε} of λ such that $\Pr\left(\mathtt{ic}(\Pi; X, Y) > \lambda\right) > \varepsilon$.

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

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

(ii) *Entropy density of (X, Y):* This density, given by $h(X, Y) = -\log \mathrm{P}_{XY}\left(X, Y\right)$, captures the randomness in the data and plays a fundamental role in the compression of the collective data of the two parties (*cf.* [19]).

(iii) *Conditional entropy density of X given $Y\Pi$:* The conditional entropy density $h(X|Y) = -\log \mathrm{P}_{X|Y}\left(X|Y\right)$ plays a fundamental role in the compression of X for an observer of Y [29, 19]. We shall use the conditional entropy density $h(X|Y\Pi)$ in our bounds.

(iv) *Sum conditional entropy density of $(X\Pi, Y\Pi)$:* The sum conditional entropy density is given by $h\left(X \triangle Y\right) = -\log \mathrm{P}_{X|Y}\left(X|Y\right) \mathrm{P}_{Y|X}\left(Y|X\right)$ has been shown recently to play a fundamental role in the communication complexity of the data exchange problem [40]. We shall use the sum conditional entropy density $h\left(X\Pi \triangle Y\Pi\right)$.

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

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

$$\Pr\left(\mathcal{E}_1\right) + \Pr\left(\mathcal{E}_2\right) + \Pr\left(\mathcal{E}_3\right) \leq \varepsilon_{\mathtt{tail}}, \quad (3)$$

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

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

and

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

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

3.2 Upper bound

We prove the following upper bound.

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

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

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

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

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

The first subroutine uses an interactive version of the classical Slepian-Wolf compression [38] (see [29] 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.* [15, 32, 43, 22, 40]).

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 [35, 30, 45].

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 [8]. However, the information theoretic form here makes it amenable to techniques such as spectrum slicing, which leads to tighter bounds than those established in [8].

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

Specifically, for a protocol π with communication complexity d, for each t, $1 \le t \le d$, we slice the essential spectrums $\left(\lambda^{\min}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}, \lambda^{\max}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right]$ and $\left(\lambda^{\min}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}, \lambda^{\max}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right]$ of $h(\Pi_t|X\Pi^{t-1})$ and $h(\Pi_t|Y\Pi^{t-1})$, respectively, into $N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$ and $N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$ slices of lengths $\Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}$ and $\Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}$. Let

$$\begin{aligned}\varepsilon_t \stackrel{\text{def}}{=} &\Pr\left(h(\Pi_t|X\Pi^{t-1}) \notin \left(\lambda^{\min}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}, \lambda^{\max}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}\right]\right)\\ &+ \Pr\left(h(\Pi_t|Y\Pi^{t-1}) \notin \left(\lambda^{\min}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}, \lambda^{\max}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}\right]\right),\end{aligned}$$

and

$$\delta_t = \begin{cases} N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + 3\log N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} \\ \qquad\qquad + \Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 3\gamma, \quad \text{odd } t, \\ N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 3\log N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + \Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} \\ \qquad\qquad + \Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + 3\gamma, \quad \text{even } t. \end{cases} \tag{6}$$

Then the fudge parameters ε' and λ' are given by

$$\varepsilon' = \sum_{t=1}^{d}\left[4\varepsilon_t + 3\left(N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} + N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} + 2\right)2^{-\gamma} + \frac{3}{N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}}} + \frac{3}{N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}}\right],$$

$$\lambda' = \sum_{t=1}^{d}\delta_t,$$

where δ_t is given by (6). Note that here

$$\Delta_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} N_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} = \lambda^{\max}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}} - \lambda^{\min}_{\mathrm{P}_{\Pi_t|X\Pi^{t-1}}},$$

and

$$\Delta_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} N_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} = \lambda^{\max}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}} - \lambda^{\min}_{\mathrm{P}_{\Pi_t|Y\Pi^{t-1}}}.$$

Thus, the optimal choice of fudge parameters ε' and δ' is roughly the sum of square roots of the lengths of essential spectrums of $h(\Pi_t|X\Pi^{t-1})$ and $h(\Pi_t|Y\Pi^{t-1})$, summed over $t = 1, ..., d$.

3.3 Amortized regime: second-order asymptotics

It was shown in [8] that information complexity of a protocol equals the amortized communication rate for simulating the protocol, *i.e.*,

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

where P_{XY}^n 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 [8] sheds no light on finer asymptotics such as the second-order term or the dependence of $D_\varepsilon(\pi^n|\mathrm{P}_{XY}^n)$ on[9] ε. On the one hand, it even remains unclear from [8] 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 $D_\varepsilon(\pi^n|\mathrm{P}_{XY}^n)$ for a finite, fixed n. A better estimate of $D_\varepsilon(\pi^n|\mathrm{P}_{XY}^n)$ 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 [39] and have received a lot of attention in information theory in recent years following [21, 33].

Our lower bound in Theorem 1 and upper bound in Theorem 2 show that the leading term in $D_\varepsilon(\pi^n|\mathrm{P}_{XY}^n)$ 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 [8]. Furthermore, the second-order asymptotic term depends on the variance $\mathtt{V}(\pi)$ of $\mathtt{ic}(\Pi; X, Y)$, *i.e.*, on

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

We have the following result.

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

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

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

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

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

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

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

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

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

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

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

A similar converse was first shown for the channel coding problem in information theory by Arimoto [3] (see [14, 34] for further refinements of this result), and has been studied for other classical information theory problems as well. To the best of our knowledge, Corollary 5 is the first instance of an Arimoto converse for a problem involving interactive communication.

In the TCS literature, such converse results have been termed *direct product theorems* and have been considered in the context of the (distributional) communication complexity problem (for computing a given function) [9, 10, 23]. Our lower bound in Theorem 1, too, yields a direct product theorem for the communication complexity problem. We state this simple result in the passing, skipping the details since they closely mimic Theorem 5. Specifically, given a function f on $\mathcal{X} \times \mathcal{Y}$, by slight abuse of notations and terminologies, let $D_\varepsilon(f) = D_\varepsilon(f|\mathrm{P}_{XY})$ be the communication complexity of computing f. As noted in Remark 3, Theorem 1 is valid for an arbitrary random variables Π, and not just an interactive protocol. Then, by following the proof of Theorem 5 with $F = f(X, Y)$ replacing Π in the application of Theorem 1, we get the following direct product theorem.

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

$$\|\pi\| \le n[H(F|X) + H(F|Y) - \delta], \tag{7}$$

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

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

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

Recall that [8, 26] showed that the first order asymptotic term in the amortized communication complexity for function computation was shown to equal the information complexity $\mathtt{IC}(f)$ of the function, namely the infimum over $\mathtt{IC}(\pi)$

[9]The lower bound in [8] gives only the *weak converse* which holds only when $\varepsilon = \varepsilon_n \to 0$ as $n \to \infty$.

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 (7) 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 [9, 10] and can be stronger in some regimes[10].

3.4 General formula for amortized communication complexity

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

The previous section, and much of the theoretical computer science literature, has focused on the case when $\mathrm{P}_{X_nY_n} = \mathrm{P}_{XY}^n$ and the same protocol π is executed on each coordinate. In this section, we identify the first-order asymptotic term in $D_{\varepsilon_n}(\pi_n|\mathrm{P}_{X_nY_n})$ for a general sequence of distributions[11] $\{\mathrm{P}_{X_nY_n}\}_{n=1}^{\infty}$ and a general sequence of protocols $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^{\infty}$. Formally, the amortized (distributional) communication complexity of $\boldsymbol{\pi}$ for $\{\mathrm{P}_{X_nY_n}\}_{n=1}^{\infty}$ is given by[12]

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

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 [20] for handling this situation (*cf.* [19]). 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 $\boldsymbol{\pi} = \{\pi_n\}_{n=1}^{\infty}$,*

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

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

For the case when $\pi_n = \pi^n$ and $\mathrm{P}_{X_nY_n} = \mathrm{P}_{XY}^n$, it follows from the law of large numbers that $\overline{\mathtt{IC}}(\boldsymbol{\pi}) = \mathtt{IC}(\pi)$ and we recover the result of [8]. However, the utility of the general formula goes far beyond this simple amortized regime. Example 1 provides one such instance. In this case, $\overline{\mathtt{IC}}(\boldsymbol{\pi})$ can be easily shown to equal $\mathtt{IC}(\pi_{\mathtt{h}})$ for any bias of the coin Π_0.

[10]The result in [9, 10] shows a direct product theorem when we communicate less than $n\mathtt{IC}(f)/\mathtt{poly}(\log n)$.

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

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

4. ASYMPTOTIC OPTIMALITY

We now present the proofs of Theorem 3, Theorem 7 and Therem 5 using single-shot bounds given in Theorem 1 and Theorem 2. Both the proofs rely on carefully choosing the slice-sizes in the lower and upper bounds.

4.1 Proof of Theorem 3

We start with the upper bound. Note that, for IID random variables (Π^n, X^n, Y^n), the spectrums of $h(\Pi_t^n|Z^n, (\Pi^{t-1})^n)$ for [13] $Z = X$ or Y have width $O(\sqrt{n})$. Therefore, the parameters Δs and Ns that appear in the fudge parameters can be chosen as $O(n^{1/4})$. Specifically, by standard measure concentration bounds (for bounded random variables), for every $\nu > 0$, there exists a constant[14] $c > 0$ such that with

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

the following bound holds:

$$\Pr\left((\Pi_t^n, (Z^n, (\Pi^{t-1})^n)) \in \mathcal{T}^{(0)}_{\mathrm{P}_{\Pi_t^n|Z^n(\Pi^{t-1})^n}}\right) \le \nu. \qquad (8)$$

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

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

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

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

for sufficiently large n. By its definition given in (6), $\delta_t = O(n^{1/4})$ for the choice of parameters above. Thus, the Berry-Esséen theorem (*cf.* [16]) 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

$$\lambda^{(1)}_{\min} = n(H(X,Y) - \delta), \quad \lambda^{(1)}_{\max} = n(H(X,Y) + \delta),$$
$$\lambda^{(2)}_{\min} = n(H(X|Y,\Pi) - \delta), \quad \lambda^{(2)}_{\max} = n(H(X|Y,\Pi) + \delta),$$
$$\lambda^{(3)}_{\min} = n(H(X\Pi\triangle Y\Pi) - \delta), \quad \lambda^{(3)}_{\max} = n(H(X\Pi\triangle Y\Pi) + \delta).$$

Then, by standard measure concentration bounds imply that the tail probability $\varepsilon_{\mathtt{tail}}$ in (3) is bounded above by $\frac{c}{n}$ for some constant $c > 0$. We also set $\eta = \frac{1}{n}$. For these choices of parameters, we note that the fudge parameter is $\lambda' = O(\log n)$. Thus, by setting

$$\lambda = \lambda_n$$

[13]We introduce Z as a placeholder for X or Y for brevity.

[14]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 (8) holds for every t.

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

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

4.2 Proof of Theorem 5

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

$$\log \|\pi_{\mathtt{sim}}\| < \lambda - \lambda', \tag{9}$$

then its simulation error must be larger than

$$\Pr\left(\mathtt{ic}\left(\Pi^n; X^n, Y^n\right) > \lambda\right) - \varepsilon'. \tag{10}$$

To compute fudge parameters, we set

$$\lambda^{(1)}_{\min} = n(H(X,Y) - \delta), \quad \lambda^{(1)}_{\max} = n(H(X,Y) + \delta),$$
$$\lambda^{(2)}_{\min} = n(H(X|Y,\Pi) - \delta), \quad \lambda^{(2)}_{\max} = n(H(X|Y,\Pi) + \delta),$$
$$\lambda^{(3)}_{\min} = n(H(X\Pi\triangle Y\Pi) - \delta), \quad \lambda^{(3)}_{\max} = n(H(X\Pi\triangle Y\Pi) + \delta).$$

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

$$\varepsilon' \le 2^{-E_1 n} + 2^{-\frac{\delta}{27}n} \tag{11}$$

and

$$\lambda' \le \frac{\delta}{3}n + O(\log n). \tag{12}$$

Finally, upon setting

$$\lambda = n\mathtt{IC}(\pi) - \frac{\delta}{3} \tag{13}$$

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

$$\Pr\left(\mathtt{ic}\left(\Pi^n; X^n, Y^n\right) > \lambda\right) \ge 1 - 2^{-E_2 n}. \tag{14}$$

The result follows upon combining (9)-(14). □

4.3 Proof of Theorem 7

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

$$\underline{H}(\mathbf{\Pi}_t|\mathbf{Z}, \mathbf{\Pi}^{t-1}) \tag{15}$$
$$= \sup\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(\Pi_{n,t}|Z_n\Pi_n^{t-1}) < \alpha\right) = 0\right\}, \tag{16}$$
$$\overline{H}(\mathbf{\Pi}_t|\mathbf{Z}, \mathbf{\Pi}^{t-1}) \tag{17}$$
$$= \inf\left\{\alpha : \lim_{n\to\infty} \Pr\left(h(\Pi_{n,t}|Z_n\Pi_n^{t-1}) > \alpha\right) = 0\right\}, \tag{18}$$

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 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

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

where δ_t is given by (6). Then, by Theorem 2, by the definition of $\overline{\mathtt{IC}}(\boldsymbol{\pi})$ and by (16) and (18), there exists a simulation protocol of length $l_{\max}$ with vanishing simulation error. Since $\delta > 0$ is arbitrary, we have the desired achievability bound.

For converse part, we fix arbitrary $\delta > 0$, and set

$$\lambda^{(1)}_{\min} = n(\underline{H}(\mathbf{X}, \mathbf{Y}) - \delta),$$
$$\lambda^{(1)}_{\max} = n(\overline{H}(\mathbf{X}, \mathbf{Y}) + \delta),$$
$$\lambda^{(2)}_{\min} = n(\underline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) - \delta),$$
$$\lambda^{(2)}_{\max} = n(\overline{H}(\mathbf{X}|\mathbf{Y}, \mathbf{\Pi}) + \delta),$$
$$\lambda^{(3)}_{\min} = n(\underline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) - \delta),$$
$$\lambda^{(3)}_{\max} = n(\overline{H}(\mathbf{X\Pi}\triangle\mathbf{Y\Pi}) + \delta),$$

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

$$\lambda = \lambda_n = n\left(\overline{\mathtt{IC}}(\boldsymbol{\pi}) + \delta\right), \tag{19}$$

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

Appendix: An example of a mixture protocol

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

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

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

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

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

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

Since

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

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

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

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

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

and

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

Thus, the ε-tail λ_ε of information complexity density is given by

$$\lambda_\varepsilon = \sup\{\lambda : \Pr\left(\mathtt{ic}(\Pi;X,Y) > \lambda\right) > \varepsilon\} = 2n. \tag{20}$$

On the other hand, we have

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

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

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

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

which in view of (20) and Theorem 1 gives $D_\varepsilon(\pi) = \Omega(2n)$.

5. REFERENCES

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

[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)*, pages 20–29, 1996.

[3] S. Arimoto. On the converse to the coding theorem for discrete memoryless channels. *IEEE Trans. Inf. Theory*, 19(3):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)*, pages 67–76, 2010.

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

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

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

[8] M. Braverman and A. Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011.

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

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

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

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

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

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

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

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

[17] 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*, pages 176–185, 2014.

[18] A. Ganor, G. Kol, and R. Raz. Exponential separation of information and communication for boolean functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:113, 2014.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[39] V. Strassen. Asymptotische abscha??tzungen in Shannon's informationstheorie. *Third Prague Conf. Inf. Theory*, pages 689–723, 1962.

[40] H. Tyagi, P. Viswanath, and S. Watanabe. Interactive communication for data exchange. *Proc. IEEE International Symposium on Information Theory*, pages 1806–1810, 2015.

[41] H. Tyagi and S. Watanabe. Converses for secret key agreement and secure computing. *IEEE Trans. Inf. Theory*, 61(9):4809–4827, September 1998.

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

[43] 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*, 56(4):1808–1824, April 2010.

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

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

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