Strong Converse using Change of Measure Arguments

Himanshu Tyagi[†] and Shun Watanabe[‡]

Abstract

The strong converse for a coding theorem shows that the optimal asymptotic rate possible with vanishing error cannot be improved by allowing a fixed error. Building on a method introduced by Gu and Effros for centralized coding problems, we develop a general and simple recipe for proving strong converse that is applicable for distributed problems as well. Heuristically, our proof of strong converse mimics the standard steps for proving a weak converse, except that we apply those steps to a modified distribution obtained by conditioning the original distribution on the event that no error occurs. A key component of our recipe is the replacement of the hard Markov constraints implied by the distributed nature of the problem with a soft information cost using a variational formula introduced by Oohama. We illustrate our method by providing a short proof of the strong converse for the Wyner-Ziv problem and strong converse theorems for interactive function computation, common randomness and secret key agreement, and the wiretap channel; the latter three strong converse problems were open prior to this work.

I. Introduction

A coding theorem in information theory characterizes the optimal rate such that there exists a code of that rate for the problem studied. Often, the first version of such theorems are proved assuming a vanishing probability of error criterion. This criterion facilitates a simple proof relying on chain rules and Fano's inequality. The strong converse holds for a coding theorem if the optimal rate claimed by the theorem cannot be improved even if a fixed error is allowed. The first strong converse was shown for the point-to-point channel coding theorem and source coding theorem by Wolfowitz (see [48]). A

A preliminary version of this work was presented at the IEEE International Symposium on Information Theory, Vail, USA, 2018.

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

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

general method for proving strong converse for coding theorems in multiterminal information theory was introduced in [4]. This method uses a strong converse for the image-size characterization problem, which is in turn shown using the blowing-up lemma; see [12] for a comprehensive treatment. The approach based on blowing-up lemma entails, in essence, changing the code to a list-code with a list-size of vanishing rate. Related recent works that involve a change in the underlying code, too, but use modern tools from functional inequalities and measure concentration literature include[1] [15] and[2] [30].

In this work, we present a simple method for proving strong converses for multiterminal problems that uses very similar steps as the weak converse proofs. Our method consists of two steps, both building on techniques available in the literature. The first step is a *change of measure argument*[3] due to Gu and Effros [19], [20]. The key idea is to evaluate the performance of a given code not under the original product measure, but under another modified measure which depends on the code and under which the code is error-free. Thus, when the standard rate bounds are applied along with Fano's inequality, we get a bound involving information quantities for the tilted measure, but without the Fano correction term for the error.

In [19], [20], Gu and Effros applied the change of measure argument for proving strong converse for source coding problems where there exists a terminal that observes all the random variables involved; a particular example is the Gray-Wyner (GW) problem [18]. A difficulty in extending this approach to other distributed source coding problems is the Markov chain constraints among random variables implied by the information structure of the communication. Specifically, these Markov chain constraints might be violated when the measure is switched. This technical difficulty was circumvented in [45] for the Wyner-Ahlswede-Körner (WAK) problem [5], [49], i.e., the problem of lossless source coding with coded side information, by relating the WAK problem to an extreme case of the GW problem. In this paper, we develop a more direct and general recipe for applying the change of measure argument to various distributed coding problems.

The second step of our recipe is the replacement of the hard Markov chain and functional constraints by soft information cost penalties using variational formulae introduced by Oohama in a series of papers including [35], [34]. These variational formulae involve optimization over a nonnegative Lagrange multiplier, with the optimum corresponding to the form with Markov constraints. In fact, when the change of measure step is applied some of the distributions that need to preserved, such as the channel transition probabilities, may change. These, too, can be accommodated by a KL-divergence cost constraint. At a

[1] For another use of the Gaussian-Poincaré inequality in information theory, see [39].

[2] The approach in [30] was extended in [29] to derive a dispersion converse bound for the Wyner-Ahlswede-Körner network.

[3] Our argument differs from the change of measure argument used to prove sphere-packing bounds (cf. [12], [22]).

high level, we replace all "hard" information constraints by "soft" divergence costs and complete the proof of strong converse by establishing super- or sub-additivity of the resulting penalized rate functions.

As an illustration of this approach, consider the lossless source coding problem; even though this problem does not involve any Markov chain constraint, it illustrates the essential ideas involved in our approach. Suppose that an independent and identically distributed (i.i.d.) source Z^n is compressed to $\varphi(Z^n)$ such that there exists a function ψ satisfying $\mathbb{P}\left(\psi(\varphi(Z^n)) = Z^n\right) \geq 1-\varepsilon$. Let $\mathcal{C}$ denote the set $\{z^n : \psi(\varphi(z^n)) = z^n\}$ of sequences where no error occurs. The strong converse for the lossless source coding theorem will be obtained upon showing that the rate of the code is bounded below by entropy $H(Z)$ asymptotically, irrespective of the value of $0 < \varepsilon < 1$. To show this, we change the probability measure to $\mathrm{P}_{\tilde{Z}^n}$ defined by[4]

$$\mathrm{P}_{\tilde{Z}^n}\left(z^n\right) = \mathbb{P}\left(Z^n = z^n | Z^n \in \mathcal{C}\right). \tag{1}$$

This measure is not too far from the original measure under KL-divergence. Indeed,[5]

$$D(\mathrm{P}_{\tilde{Z}^n} \| \mathrm{P}_{Z^n}) \leq \log \frac{1}{1-\varepsilon}.$$

On the other hand, under $\mathrm{P}_{\tilde{Z}^n}$, the error probability of the code (φ, ψ) is exactly zero. Thus, by mimicking the standard weak converse arguments, we have

$$\log |\mathcal{C}| \geq H(\tilde{Z}^n).$$

The next step is to single-letterize $H(\tilde{Z}^n)$, which now does not correspond to a product measure and may not be super-additive on its own. We circumvent this difficulty by adding a divergence cost to get

$$\begin{aligned}\frac{1}{n}\log|\mathcal{C}| &\geq \frac{1}{n}H(\tilde{Z}^n) + \frac{\alpha}{n}\Big[D(\mathrm{P}_{\tilde{Z}^n}\|\mathrm{P}_{Z^n}) - \log(1/(1-\varepsilon))\Big] \\ &\geq \min_{\mathrm{P}_{\tilde{Z}}}\left[H(\tilde{Z}) + \alpha D(\mathrm{P}_{\tilde{Z}}\|\mathrm{P}_Z)\right] - \frac{\alpha\log(1/(1-\varepsilon))}{n},\end{aligned}$$

for any $\alpha > 0$. The second inequality uses a simple super-additivity property that we show in Proposition 1 for conditional entropy. The proof of strong converse is completed by using the following variational formula for entropy:

$$H(Z) = \sup_{\alpha>0}\min_{\mathrm{P}_{\tilde{Z}}}\left[H(\tilde{Z}) + \alpha D(\mathrm{P}_{\tilde{Z}}\|\mathrm{P}_Z)\right].$$

[4] In [19], [20], the new distribution had a more complicated form.

[5] This simple, but important, observation was used in [32] to provide a simple proof of the blowing-up lemma.

Using our recipe, we can obtain simple proof for some known strong converse results and can, in fact, obtain several new strong converse results, including for problems involving interactive communication. The first result we present is the lossy source coding with side information problem, also known as the Wyner-Ziv (WZ) problem [51]. The strong converse for the WZ problem was proved only recently in [34]. We use our general recipe for proving a strong converse to give a more compact proof for the WZ strong converse which, we believe, is more accessible than the original proof of [34].[6] The second problem we consider is the interactive function computation problem (cf. [37], [31], [8]). Prior to our work, a strong converse for this well-studied problem was unavailable. A technical difficulty in showing such a result arises from the multiple auxiliary random variables and Markov chain constraints that appear in the optimal sum-rate. The strong converse for the interactive function computation problem has attracted attention in the theoretical computer science community as well, in the context of direct product theorems in communication complexity. A version of the strong converse result was shown in [9] in a slightly different setting, but the basic strong converse itself has been open. Furthermore, the information odometer approach used in [9] is technically much more involved than our simple change of measure argument.

In addition to the two source coding problems mentioned above, we also apply our recipe for problems of generating common randomness and secret key with interactive communication [3], [42]. The strong converse for these problems with interactive communication were unavailable prior to our work; see [24] and the extended version of [30] for partial results. Since these problems involve a total variation distance constraint, we need some additional tricks for changing measure. In particular, we seek a replacement for the correctly decoded set of sequences $\mathcal{C}$. We illustrate the essential idea using a simple random number generation problem, which is also known as the intrinsic randomness problem (cf. [21]). Suppose that an i.i.d. source Z^n is converted to $K = \varphi(Z^n)$ such that the total variation distance criterion $d(\mathrm{P}_K, \mathrm{P}_{\mathtt{unif}}) \le \delta$ is satisfied, where $\mathrm{P}_{\mathtt{unif}}$ is the uniform distribution of the range $\mathcal{K}$ of K. Consider the set

$$\mathcal{C} = \left\{ z^n : \log \frac{1}{\mathrm{P}_K\left(\varphi(z^n)\right)} \ge \log|\mathcal{K}| - \log(2/(1-\delta)) \right\} \tag{2}$$

comprising elements z^n mapped to high entropy density realizations of K. It can be seen from our analysis in Section V that

$$\mathbb{P}\left(Z^n \in \mathcal{C}\right) \ge \frac{1-\delta}{2}.$$

[6] The proof in [34] provides a stronger result in form of an explicit lower bound on the exponent of the probability of correctness.

Thus, by changing the measure to $\mathrm{P}_{\tilde{Z}^n}$ given in (1) but using the set $\mathcal{C}$ of (2), we have $D(\mathrm{P}_{\tilde{Z}^n}\|\mathrm{P}_{Z^n}) \leq \log(2/(1-\delta))$. Furthermore, for this changed measure, the random variable $\tilde{K} = \varphi(\tilde{Z}^n)$ has the min-entropy at least $\log|\mathcal{K}| - 2\log(2/(1-\delta))$, which implies

$$\begin{aligned}\log|\mathcal{K}| &\leq H_{\min}(\tilde{K}) + 2\log(2/(1-\delta))\\ &\leq H(\tilde{K}) + 2\log(2/(1-\delta))\\ &\leq H(\tilde{Z}^n) + 2\log(2/(1-\delta)).\end{aligned}$$

The entropy term on the right-side can be bounded using the sub-additivity of entropy. However, the resulting single-letterized measure may deviate from the original P_Z, which needs to be retained. To that end, we add a divergence cost to get

$$\begin{aligned}\frac{1}{n}\log|\mathcal{K}| &\leq \frac{1}{n}H(\tilde{Z}^n) - \frac{\alpha}{n}\left[D\left(\mathrm{P}_{\tilde{Z}^n}\|\mathrm{P}_{Z^n}\right) - \log(2/(1-\delta))\right] + \frac{2\log(2/(1-\delta))}{n}\\ &\leq \max_{\mathrm{P}_{\tilde{Z}}}\left[H(\tilde{Z}) - \alpha D(\mathrm{P}_{\tilde{Z}}\|\mathrm{P}_Z)\right] + \frac{(\alpha+2)\log(2/(1-\delta))}{n},\end{aligned}$$

for any $\alpha > 0$. The strong converse for the random number generation problem follows from the variational formula

$$H(Z) = \inf_{\alpha>0}\max_{\mathrm{P}_{\tilde{Z}}}\left[H(\tilde{Z}) - \alpha D(\mathrm{P}_{\tilde{Z}}\|\mathrm{P}_Z)\right].$$

The final setting we consider is the wiretap channel [50], [10]. The strong converse theorem for degraded wiretap channel was proved in [23] (see [41] for a partial strong converse). However, its extension to general wiretap channel has remained open.[7] By using our general recipe, we provide a proof for the strong converse theorem for the general wiretap channel. Compared to other problems mentioned above, this problem is more involved, and requires a few more tricks including the expurgation of messages to replace average guarantees with worst-case guarantees and the construction of the changed measure using a set with bounded log-likelihood ratio of wiretappers observation probability and its probability given the message. Nevertheless, given the technical difficulties in prior attempts, this is a relatively simple proof.

Overall, our main message in this work is that strong converses can be proven using similar techniques as those used for proving weak converses, applied after an appropriate change of measure. However, we need to work with new variational forms of capacity formulae where the hard information constraints

[7] The argument in [47] has a technical flaw, and we are unable to verify the technically involved proof-sketch in the conference paper [17]; a full-version of [17] has not been published so far.

are replaced with soft KL-divergence costs.

A conceptually related approach for proving strong converse was recently proposed by Kosut and Kliewer in [28]. In their approach, the strong converse for a given network is reduced first to the weak converse by adding an extra edge of vanishing rate to the network, which allows negligible cooperation among users. Then, the strong converse will follow if the so-called edge removal property holds, namely the capacity is not changed when the extra edge is removed. Since Markov chain constraints in multiterminal problems stem from distributed nature of the problems, the replacement of those Markov chain constraints with soft KL-divergence costs in our recipe is, at high-level, similar to adding a "soft edge" to increase cooperation among the terminals. However, the soft divergence cost seems to be a more versatile tool; in particular, it allows us to handle even interactive communication.

Another related recent work is that of Jose and Kulkarni [26], [27]. Their approach considers the performance of the optimal code for a coding problem and poses it as an optimization problem, which is further bounded by the value of a linear program obtained by relaxing some constraints. Even though this approach provides tight converse bounds implying strong converse for some problems, applicability of this approach to problems involving auxiliary random variables is unclear.

In a slightly different direction, Fong and Tan [16] proved strong converse theorems for multi-message networks with tight cut-set bound, such as the degraded relay channel and relay channel with orthogonal components, for both discrete and Gaussian channels. These results are inspired by the result for the reliability function of a DMC with feedback above capacity [11], and are different in nature than our setting. In particular, these results do not include multiple auxiliary random variables, which is a significant difficulty we overcome in this paper.

The remainder of the paper is organized as follows. We begin by reviewing a few simple results in the next section, which will be used throughout the paper. The strong converse for the WZ problem is given in Section III and for the function computation problem in Section IV. The next two sections contain problems involving total variation constraints, with the common randomness generation and secret key agreement in Section V, and the wiretap channel problem in Section VI. We conclude with discussions on exponential strong converse and extensions in the final section.

Notation: Throughout the paper, we restrict to discrete random variables taking finitely many values and denote the random variable with a capital letter, for instance X, its range-set with the corresponding calligraphic, e.g. $\mathcal{X}$, and each realization with a small letter, e.g. x. For information measures, we follow the standard notations in [12]: The entropy, the KL divergence, and the mutual information are denoted by $H(X)$, $D(\mathrm{P}\|\mathrm{Q})$, and $I(X \wedge Y)$, respectively. The total variation distance between two distributions P and Q is denoted by $d(\mathrm{P}, \mathrm{Q}) := \frac{1}{2}\sum_x |\mathrm{P}\left(x\right) - \mathrm{Q}\left(x\right)|$. For a sequence $X^n = (X_1, \cdots, X_n)$ of random variables, we denote $X_j^- = (X_1, \dots, X_{j-1})$ and $X_j^+ = (X_{j+1}, \dots, X_n)$, where X_1^- and X_n^+ are regarded as the empty string. The indicator function is denoted by $\mathbb{1}[\cdot]$. Other notations will be introduced when necessary, but are standard notations used in the multiterminal information theory literature.

II. Technical Tools

We begin by assembling the simple tools that we will use repeatedly in our proofs. The first is perhaps a new observation; the other two are standard.

Typically, we use additivity of (conditional) entropy for independent random variables for proving converse bounds. However, in our proofs, once we change the measure, the resulting random variables need not be independent. Nevertheless, the following simple result fills the gap and shows that if we add a divergence cost for change of measure, the sum is super-additive.

Proposition 1. *For i.i.d.* $\mathrm{P}_{X^nY^n}$ *with common distribution* P_{XY} *and any* $\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}$*, we have*

$$H(\tilde{X}^n|\tilde{Y}^n) + D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}_{X^nY^n}) \geq n\big[H(\tilde{X}_J|\tilde{Y}_J) + D(\mathrm{P}_{\tilde{X}_J\tilde{Y}_J}\|\mathrm{P}_{XY})\big],$$

where $J \sim \mathtt{unif}(\{1, ..., n\})$ *is the time-sharing random variable and is assumed to be independent of all the other random variables involved.*

Proof. The left-side can be expressed as

$$H(\tilde{X}^n|\tilde{Y}^n) + D(\mathrm{P}_{\tilde{X}^n|\tilde{Y}^n}\|\mathrm{P}_{X^n|Y^n}|\mathrm{P}_{\tilde{Y}^n}) + D(\mathrm{P}_{\tilde{Y}^n}\|\mathrm{P}_{Y^n}).$$

The sum of the first two terms satisfy

$$\begin{aligned}
H(\tilde{X}^n|\tilde{Y}^n) + D(\mathrm{P}_{\tilde{X}^n|\tilde{Y}^n}\|\mathrm{P}_{X^n|Y^n}|\mathrm{P}_{\tilde{Y}^n}) &= \sum_{x^n, y^n} \mathrm{P}_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) \log \frac{1}{\mathrm{P}_{X^n|Y^n}(x^n|y^n)} \\
&= \sum_{j=1}^n \sum_{x,y} \mathrm{P}_{\tilde{X}_j\tilde{Y}_j}(x, y) \log \frac{1}{\mathrm{P}_{X|Y}(x|y)} \\
&= n \sum_{x,y} \mathrm{P}_{\tilde{X}_J\tilde{Y}_J}(x, y) \log \frac{1}{\mathrm{P}_{X|Y}(x|y)} \\
&= nH(\tilde{X}_J|\tilde{Y}_J) + nD(\mathrm{P}_{\tilde{X}_J|\tilde{Y}_J}\|\mathrm{P}_{X|Y}|\mathrm{P}_{\tilde{Y}_J}),
\end{aligned}$$

and the third satisfies

$$D(\mathrm{P}_{\tilde{Y}^n}\|\mathrm{P}_{Y^n}) = \sum_{j=1}^n D(\mathrm{P}_{\tilde{Y}_j|\tilde{Y}_j^-}\|\mathrm{P}_Y|\mathrm{P}_{\tilde{Y}_j^-})$$

$$\geq \sum_{j=1}^{n} D(\mathrm{P}_{\tilde{Y}_j} \| \mathrm{P}_Y)$$

$$\geq nD(\mathrm{P}_{\tilde{Y}_J} \| \mathrm{P}_Y),$$

which completes the proof. □

The next tool we present is essential for handling the distributed settings we consider. It allows us to replace the "hard" Markov chain and function constraints in our bounds with "soft" costs using a variational formula introduced by Oohama (cf. [34]) in this context. This is important since these hard constraints may not hold once we change the measure. We describe this approach in an abstract form below; proofs for specific variants needed for our results are similar and have been relegated to the appendix.

Let $G(\mathrm{P}_{Z_1Z_2})$ be a bounded continuous function of $\mathrm{P}_{Z_1Z_2}$. Define[8]

$$\overline{G}(\mathrm{P}_{Z_1Z_2}) = \inf_{\mathrm{P}_{U|Z_1Z_2} : \mathrm{P}_{U|Z_1Z_2} = \mathrm{P}_{U|Z_1}} \mathbb{E}\left[G(\mathrm{P}_{Z_1Z_2|U})\right].$$

Note that by the support lemma [12], it suffices to restrict the infimum to U with $|\mathcal{U}| \leq |\mathcal{Z}_1|$, and thereby the $\inf$ can be replaced by $\min$ using compactness of the finite dimensional probability simplex. The next result we present is a variational formula for $\overline{G}(\mathrm{P}_{Z_1Z_2})$ that allows us to replace the minimization over U satisfying the Markov chain condition $U \multimap Z_1 \multimap Z_2$ to that over all $\mathrm{P}_{U|Z_1Z_2}$.

Proposition 2. *Let $G(\mathrm{P}_{Z_1Z_2})$ be a bounded continuous function over the probability simplex $\mathcal{P}(\mathcal{Z}_1 \times \mathcal{Z}_2)$. Then, the function $\overline{G}(\mathrm{P}_{Z_1Z_2})$ satisfies*

$$\overline{G}(\mathrm{P}_{Z_1Z_2}) = \sup_{\alpha>0} \min_{\mathrm{P}_{U|Z_1Z_2}} \left[\mathbb{E}\left[G(\mathrm{P}_{Z_1Z_2|U})\right] + \alpha I(U \wedge Z_2|Z_1)\right]. \tag{3}$$

Proof. The left-side is greater than or equal to the right-side since, for every $\alpha > 0$, the left-side is obtained by restricting the inner minimization on the right to the distribution satisfying $U \multimap Z_1 \multimap Z_2$. To prove the other direction, first note that $I(U \wedge Z_1|Z_2)$ can be written as $D(\mathrm{P}_{UZ_1Z_2} \| \mathrm{P}_{U|Z_1}\mathrm{P}_{Z_1}\mathrm{P}_{Z_2|Z_1})$. Given $\alpha > 0$, let $\mathrm{P}^{\alpha}_{U|Z_1Z_2}$ attain the inner minimum in (3) for α. Since the function $G(\cdot)$ is bounded, say it lies in an interval $[a, b]$, the same holds for the function $\overline{G}(\cdot)$. Therefore, it must hold that $D(\mathrm{P}^{\alpha}_{UZ_1Z_2} \| \mathrm{P}^{\alpha}_{U|Z_1}\mathrm{P}_{Z_1Z_2}) \leq (b-a)/\alpha$. Let $\tilde{\mathrm{P}}_{UZ_1Z_2} = \mathrm{P}^{\alpha}_{U|Z_1}\mathrm{P}_{Z_1Z_2}$. Since $G(\cdot)$ is continuous[9] on a compact

[8] We abbreviate $G(\mathrm{P}_{Z_1Z_2|U}(\cdot|U))$ as $G(\mathrm{P}_{Z_1Z_2|U})$.

[9] We are assuming $G(\cdot)$ is continuous with respect to the total variation distance. Then, it is also continuous with respect to the KL divergence using Pinsker's inequality.

domain, it is also uniformly continuous. Therefore, there exists a function $\Delta(t)$ satisfying $\Delta(t) \to 0$ as $t \to 0$ such that

$$\begin{aligned}\mathbb{E}\left[G(\mathrm{P}^{\alpha}_{Z_1Z_2|U})\right] &\geq \mathbb{E}\left[G(\tilde{\mathrm{P}}_{Z_1Z_2|U})\right] - \Delta((b-a)/\alpha)\\ &\geq \overline{G}(\mathrm{P}_{Z_1Z_2}) - \Delta((b-a)/\alpha).\end{aligned}$$

Thus, we obtain the required inequality by taking $\alpha \to \infty$, which completes the proof. □

The variational form above can be used to handle even multiple Markov relations by adding a similar cost for each constraint. Furthermore, we can even handle functional constraints such as $H(Z_1|U, Z_2) = 0$ by adding an additional cost $\alpha H(Z_1|U, Z_2)$. These extensions of Proposition 2 will be used in our proofs.

Additionally, we also need a cost to account for the deviation from the underlying fixed source and channel distributions that occur when we apply our change of measure arguments. The following alternative variational formula for $\overline{G}(\mathrm{P}_{Z_1Z_2})$ will be handy:

Proposition 3. *Let $G(\mathrm{P}_{Z_1Z_2})$ be a bounded continuous function over the probability simplex $\mathcal{P}(\mathcal{Z}_1 \times \mathcal{Z}_2)$. Then, we have*

$$\overline{G}(\mathrm{P}_{Z_1Z_2}) = \sup_{\alpha>0} \min_{\mathrm{P}_{\tilde{U}\tilde{Z}_1\tilde{Z}_2}} \left[\mathbb{E}\left[G(\mathrm{P}_{\tilde{Z}_1\tilde{Z}_2|\tilde{U}})\right] + \alpha\big(D(\mathrm{P}_{\tilde{Z}_1\tilde{Z}_2}\|\mathrm{P}_{Z_1Z_2}) + I(U \wedge Z_2|Z_1)\big)\right].$$

The proof is similar to that of Proposition 2; instead of proving this meta-result, we will prove our specific variational formulae in the appendix.

The final result we recall is a standard tool for single-letterization from [12, pg. 314]– its power lies in its validity for arbitrary distributions. For random variables X^n, Y^n, U with an arbitrary joint distribution $\mathrm{P}_{X^nY^nU}$, it holds that

$$H(X^n|U) - H(Y^n|U) = \sum_{i=1}^{n} H(X_i|X_i^-, Y_i^+, U) - H(Y_i|X_i^-, Y_i^+, U). \tag{4}$$

III. Lossy source coding with side-information

In the lossy source coding problem with side-information, the goal is to compress a source sequence to enable its recovery within a prespecified distortion at a receiver with side-information. Formally, for a given source P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, a lossy source code with side-information consists of an encoder $\varphi : \mathcal{X}^n \to \mathcal{M}$ and a decoder $\psi : \mathcal{M} \times \mathcal{Y}^n \to \mathcal{Z}^n$, where $\mathcal{Z}$ is the reproduction alphabet. Consider a distortion measure $d : \mathcal{X} \times \mathcal{Z} \to [0, D_{\max}]$ and its n-fold extension $d(x^n, z^n) = \sum_{i=1}^{n} d(x_i, z_i)$. A rate-distortion pair (R, D) is ε-achievable if, for every sufficiently large n, there exists a code (φ, ψ) such that

$$\mathbb{P}\left(d(X^n, \psi(\varphi(X^n), Y^n)) > nD\right) \leq \varepsilon \tag{5}$$

and

$$\frac{1}{n} \log |\mathcal{M}| \leq R. \tag{6}$$

Let $\mathcal{R}_{\mathtt{WZ}}(\varepsilon|\mathrm{P}_{XY})$ be the closure of the set of all ε-achievable rate-distortion pairs. Define

$$\mathcal{R}_{\mathtt{WZ}}(\mathrm{P}_{XY}) := \bigcap_{0<\varepsilon<1} \mathcal{R}_{\mathtt{WZ}}(\varepsilon|\mathrm{P}_{XY}).$$

The following characterization[10] of $\mathcal{R}_{\mathtt{WZ}}(\mathrm{P}_{XY})$ was given in [51]:

$$\begin{aligned}\mathcal{R}_{\mathtt{WZ}}(\mathrm{P}_{XY}) = \{(R, D) : \exists\, (U, Z) \text{ s.t. } & |\mathcal{U}| \leq |\mathcal{X}| + 1, \\ & U -\!\!\circ\!\!- X -\!\!\circ\!\!- Y, Z -\!\!\circ\!\!- (U, Y) -\!\!\circ\!\!- X, \\ & R \geq I(U \wedge X|Y), \mathbb{E}[d(X, Z)] \leq D\}.\end{aligned}$$

The set $\mathcal{R}_{\mathtt{WZ}}(\mathrm{P}_{XY})$ is closed and convex and can be expressed alternatively using tangent lines as follows:

$$\mathcal{R}_{\mathtt{WZ}}(\mathrm{P}_{XY}) = \bigcap_{\mu \geq 0} \{(R, D) : R + \mu D \geq R^{\mu}_{\mathtt{WZ}}(\mathrm{P}_{XY})\},$$

where

$$\begin{aligned}R^{\mu}_{\mathtt{WZ}}(\mathrm{P}_{XY}) := \min \big\{ & I(U \wedge X|Y) + \mu \mathbb{E}[d(X, Z)] : \\ & \exists\, (U, Z) \text{ s.t. } |\mathcal{U}| \leq |\mathcal{X}|, U -\!\!\circ\!\!- X -\!\!\circ\!\!- Y, Z -\!\!\circ\!\!- (U, Y) -\!\!\circ\!\!- X \big\}.\end{aligned}$$

The optimal rate region above involves Markov relations, which will become intractable once we change the measure. Furthermore, once we change the measure and obtain a single-letter bound, the source distribution may deviate from P_{XY}. To circumvent these difficulties, we switch to the following variational form of $R^{\mu}_{\mathtt{WZ}}(\mathrm{P}_{XY})$, which will be proved in Appendix A:

$$R^{\mu}_{\mathtt{WZ}}(\mathrm{P}_{XY}) = \sup_{\alpha > 0} R^{\mu,\alpha}_{\mathtt{WZ}}(\mathrm{P}_{XY}), \tag{7}$$

[10] In fact, we can restrict Z to be a function of (U, Y).

where

$$
\begin{aligned}
R_{\mathtt{WZ}}^{\mu,\alpha}(\mathrm{P}_{XY}) &:= \min_{\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}}} \left[I(\tilde{U} \wedge \tilde{X}|\tilde{Y}) + \mu\mathbb{E}[d(\tilde{X},\tilde{Z})] + \alpha D(\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}} \| \mathrm{Q}_{\tilde{U}XY\tilde{Z}}) + D(\mathrm{P}_{\tilde{X}\tilde{Y}} \| \mathrm{P}_{XY}) \right] \qquad (8) \\
&= \min_{\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}}} \Big[I(\tilde{U} \wedge \tilde{X}|\tilde{Y}) + \mu\mathbb{E}[d(\tilde{X},\tilde{Z})] \\
&\quad + \big((\alpha+1) D(\mathrm{P}_{\tilde{X}\tilde{Y}} \| \mathrm{P}_{XY}) + \alpha I(\tilde{U} \wedge \tilde{Y}|\tilde{X}) + \alpha I(\tilde{Z} \wedge \tilde{X}|\tilde{U},\tilde{Y}) \big) \Big] \qquad (9)
\end{aligned}
$$

and $\mathrm{Q}_{\tilde{U}XY\tilde{Z}} = \mathrm{P}_{\tilde{Z}|\tilde{U}\tilde{Y}}\mathrm{P}_{\tilde{U}|\tilde{X}}\mathrm{P}_{XY}$ is the distribution induced from each $\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}}$. Note that this $\mathrm{Q}_{\tilde{U}XY\tilde{Z}}$ respects the information structure of the coding problem; we will use this convention in our usage of notation Q throughout. By the support lemma [12], the range $\mathcal{U}$ of $\tilde{U}$ can be restricted to $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}||\mathcal{Z}|$.

Remark 1. In effect, we have replaced the "hard constraints" imposed by the requirements of preserving the input source distribution and the Markov relations between the communication sent, the source and the reconstructed estimate with "soft" divergence penalties which are amenable to single-letterization using standard chain rules. The factor $(\alpha + 1)$ instead of α is only to enable a technical manipulation in the proof of Theorem 4 below. However, semantically, the bound can be understood by just considering an extra $\alpha D(\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}} \| \mathrm{Q}_{\tilde{U}XY\tilde{Z}})$ cost which captures all the aforementioned constraints. In fact, a factor in the form of any function $f(\alpha)$ of α that blows-up to infinity as α tends to infinity will work, since we take $\alpha \to \infty$ at the end. In the definition of $R_{\mathtt{WZ}}^{\mu,\alpha}(\mathrm{P}_{XY})$, the divergence form (8) is heuristically appealing and affords a simple proof of the variational formula (7) (see Appendix A); on the other hand, the mutual information form (9) is amenable to single-letterization in the proof of Theorem 4 below.

We are now in a position to prove the strong converse. The main step is to show the following result, which is obtained simply by using the super-additivity of the lower bound obtained after change of measure.

Theorem 4. *For every $n \in \mathbb{N}$, $\mu \geq 0$, and $\alpha > 0$, we have*

$$
R_{\mathtt{WZ}}^{\mu,\alpha}(\mathrm{P}_{XY}^n) \geq n R_{\mathtt{WZ}}^{\mu,\alpha}(\mathrm{P}_{XY}).
$$

As a corollary, we obtain the strong converse for the lossy source coding with side-information problem, which was shown in [34] using a different, more complicated method.

Corollary 5. *For every $0 < \varepsilon < 1$, we have $\mathcal{R}_{\mathtt{WZ}}(\varepsilon|\mathrm{P}_{XY}) = \mathcal{R}_{\mathtt{WZ}}(\mathrm{P}_{XY})$.*

Proof of Corollary 5: For a given code (φ, ψ) satisfying (5) and (6), define

$$\mathcal{D} = \{(x^n, y^n) : d(x^n, \psi(\varphi(x^n), y^n)) \le nD\}.$$

Further, let $\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}$ be defined by

$$\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) := \frac{\mathrm{P}^n_{XY}(x^n, y^n)\mathbb{1}[(x^n, y^n) \in \mathcal{D}]}{\mathrm{P}^n_{XY}(\mathcal{D})}.$$

Then, the excess distortion probability of the same code (φ, ψ) for the source $(\tilde{X}^n, \tilde{Y}^n)$ is exactly 0, which implies $\tilde{Z}^n = \psi(\varphi(\tilde{X}^n), \tilde{Y}^n)$ satisfies $\mathbb{E}[d(\tilde{X}^n, \tilde{Z}^n)] \le nD$. Thus, by mimicking the standard weak converse proof, we have

$$n(R + \mu D) \ge I(\tilde{S} \wedge \tilde{X}^n|\tilde{Y}^n) + \mu\mathbb{E}[d(\tilde{X}^n, \tilde{Z}^n)],$$

where $\tilde{S} = \varphi(\tilde{X}^n)$. Also,

$$D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}^n_{XY}) = \log\frac{1}{\mathrm{P}^n_{XY}(\mathcal{D})} \le \log\frac{1}{1-\varepsilon}.$$

Thus, by noting that costs $I(\tilde{S} \wedge \tilde{Y}^n|\tilde{X}^n)$ and $I(\tilde{Z}^n \wedge \tilde{X}^n|\tilde{S}, \tilde{Y}^n)$ are both 0, we have

$$\begin{aligned} n(R + \mu D) &\ge I(\tilde{S} \wedge \tilde{X}^n|\tilde{Y}^n) + \mu\mathbb{E}[d(\tilde{X}^n, \tilde{Z}^n)] + \big((\alpha+1)D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}^n_{XY}) \\ &\qquad + \alpha I(\tilde{S} \wedge \tilde{Y}^n|\tilde{X}^n) + \alpha I(\tilde{Z}^n \wedge \tilde{X}^n|\tilde{S}, \tilde{Y}^n)\big) - (\alpha+1)\log\frac{1}{1-\varepsilon} \\ &\ge R^{\mu,\alpha}_{\mathtt{WZ}}(\mathrm{P}^n_{XY}) - (\alpha+1)\log\frac{1}{1-\varepsilon}. \end{aligned}$$

Therefore, by Theorem 4, we have

$$R + \mu D \ge R^{\mu,\alpha}_{\mathtt{WZ}}(\mathrm{P}_{XY}) - \frac{(\alpha+1)}{n}\log\frac{1}{1-\varepsilon} \tag{10}$$

for every $\mu \ge 0$ and $\alpha > 0$, whereby the corollary follows from (7). □

Proof of Theorem 4: By setting

$$\begin{aligned} G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}) &:= H(\tilde{X}^n|\tilde{Y}^n) + \alpha H(\tilde{Y}^n|\tilde{X}^n) + (\alpha+1)D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}^n_{XY}), \\ G_2(\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}) &:= -H(\tilde{X}^n|\tilde{U}, \tilde{Y}^n) + \mu\mathbb{E}[d(\tilde{X}^n, \tilde{Z}^n)] + \alpha\big(-H(\tilde{Y}^n|\tilde{U}, \tilde{X}^n) + I(\tilde{Z}^n \wedge \tilde{X}^n|\tilde{U}, \tilde{Y}^n)\big), \end{aligned}$$

for given $\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}$, we can write

$$R^{\mu,\alpha}_{\mathtt{WZ}}(\mathrm{P}^n_{XY}) = \min_{\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}} \big[G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}) + G_2(\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n})\big].$$

Fix arbitrary $\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}$. By Proposition 1, $G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n})$ can be lower bounded as[11]

$$G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}) \geq nG_1(\mathrm{P}_{\tilde{X}_J\tilde{Y}_J}). \tag{11}$$

For $G_2(\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n})$, note that

$$\begin{aligned} -H(\tilde{X}^n|\tilde{U},\tilde{Y}^n) &= -\sum_{j=1}^{n} H(\tilde{X}_j|\tilde{U},\tilde{X}_j^-,\tilde{Y}^n) \\ &\geq -\sum_{j=1}^{n} H(\tilde{X}_j|\tilde{U},\tilde{X}_j^-,\tilde{Y}_j^+,Y_j), \\ &= -nH(\tilde{X}_J|\tilde{U}_J,J,\tilde{Y}_J), \end{aligned}$$

where $\tilde{U}_j = (\tilde{U},\tilde{X}_j^-,\tilde{Y}_j^+)$. Also, $\mathbb{E}[d(\tilde{X}^n,\tilde{Z}^n)] = n\mathbb{E}[d(\tilde{X}_J,\tilde{Z}_J)]$. For the remaining terms in G_2, we have

$$\begin{aligned} &-H(\tilde{Y}^n|\tilde{U},\tilde{X}^n) + I(\tilde{Z}^n \wedge \tilde{X}^n|\tilde{U},\tilde{Y}^n) \\ &= -H(\tilde{X}^n|\tilde{U},\tilde{Y}^n,\tilde{Z}^n) + H(\tilde{X}^n|\tilde{U}) - H(\tilde{Y}^n|\tilde{U}) \\ &= \sum_{j=1}^{n} \big[- H(\tilde{X}_j|\tilde{U},\tilde{X}_j^-,\tilde{Y}^n,\tilde{Z}^n) + H(\tilde{X}_j|\tilde{U},\tilde{X}_j^-,\tilde{Y}_j^+) - H(\tilde{Y}_j|\tilde{U},\tilde{X}_j^-,\tilde{Y}_j^+)\big] \\ &\geq n\big[- H(\tilde{X}_J|\tilde{U}_J,J,\tilde{Y}_J,\tilde{Z}_J) + H(\tilde{X}_J|\tilde{U}_J,J) - H(\tilde{Y}_J|\tilde{U}_J,J)\big] \\ &= n\big[- H(\tilde{Y}_J|\tilde{U}_J,J,\tilde{X}_J) + I(\tilde{Z}_J \wedge \tilde{X}_J|\tilde{U}_J,J,\tilde{Y}_J)\big], \end{aligned}$$

where the second identity uses (4). Upon combining the observations above, we get

$$G_2(\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}) \geq nG_2(\mathrm{P}_{\tilde{U}_JJ\tilde{X}_J\tilde{Y}_J\tilde{Z}_J}). \tag{12}$$

Since (11) and (12) hold for an arbitrary $\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}$, the proof is complete. □

IV. Interactive Function Computation Problem

The second problem we consider entails the computation of a function f of (X,Y) using interactive communication. For the ease of presentation, we limit ourselves to protocols with 2-rounds of communication, but our analysis extends to protocols with bounded (independent of n) rounds.

For a given source P_{XY} on a finite alphabet $\mathcal{X}\times\mathcal{Y}$, an (2-round) interactive communication protocol π with inputs (X^n,Y^n) consists of mappings $\varphi_1 : \mathcal{X}^n \to \{0,1\}^{l_1}$ and $\varphi_2 : \mathcal{Y}^n \times \{0,1\}^{l_1} \to \{0,1\}^{l_2}$; the

[11] By a slight abuse of notation, $G_1(\mathrm{P}_{\tilde{X}_J\tilde{Y}_J})$ is defined by replacing $\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}$ and P_{XY}^n with $\mathrm{P}_{\tilde{X}_J\tilde{Y}_J}$ and P_{XY} in the definition of $G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n})$.

length of such a protocol π is $l_1 + l_2$. The random transcript of the protocol is denoted by $\Pi = (\Pi_1, \Pi_2)$ where $\Pi_1 = \varphi_1(X^n)$, and $\Pi_2 = \varphi_2(Y^n, \Pi_1)$.

A protocol π ε-computes a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ if we can form estimates $F_1^n = \psi_1(X^n, \Pi_2)$ and $F_2^n = \psi_2(Y^n, \Pi_1)$ such that

$$\mathbb{P}\left(f(X_i, Y_i) = F_{1i} = F_{2i},\, \forall i \in [n]\right) \geq 1 - \varepsilon.$$

A rate $R > 0$ is an ε-achievable communication rate for f if, for all n sufficiently large, there exists an interactive communication protocol π of length $|\pi|$ less than nR that ε-computes f. The infimum over all ε-achievable communication rates for f is denoted by $R_f(\varepsilon|\mathrm{P}_{XY})$. The supremum over $\varepsilon \in (0,1)$ of all ε-achievable communication rates for f is denoted by $R_f(\mathrm{P}_{XY})$.

The following characterization of $\mathcal{R}_f(\mathrm{P}_{XY})$ was given in [37]:[12]

$$R_f(\mathrm{P}_{XY}) = \min I(U, V \wedge X|Y) + I(U, V \wedge Y|X), \tag{13}$$

where the minimum is over all U, V satisfying $U \mathrel{-\!\!\circ\!\!-} X \mathrel{-\!\!\circ\!\!-} Y$, $V \mathrel{-\!\!\circ\!\!-} (Y, U) \mathrel{-\!\!\circ\!\!-} X$; $H(f(X,Y)|Y,U) = H(f(X,Y)|X,U,V) = 0$; and $|\mathcal{U}| \leq |\mathcal{X}|, |\mathcal{V}| \leq |\mathcal{Y}||\mathcal{X}|$.

Remark 2. The right-side of (13) is referred to as the *intrinsic information complexity* of f (using 2-round communication protocols) in the computer science literature (cf. [8]). By noting the Markov relations $U \mathrel{-\!\!\circ\!\!-} X \mathrel{-\!\!\circ\!\!-} Y$ and $V \mathrel{-\!\!\circ\!\!-} (Y, U) \mathrel{-\!\!\circ\!\!-} X$, we can obtain the following equivalent expression for it:

$$I(U \wedge X|Y) + I(V \wedge Y|U, X),$$

which is perhaps more commonly used in the information theory literature (cf. [37], [31]). In a similar vein, we use the *extrinsic information complexity* $I(U, V \wedge X, Y)$ to describe the results in Section V, which, too, can be expressed alternatively using the aforementioned Markov relations.

In the manner of Proposition 3, we can replace the "hard" Markov chain and functional constraints with "soft" divergence penalties to get (see Appendix B for the proof)

$$R_f(\mathrm{P}_{XY}) = \sup_{\alpha > 0} R_f^{\alpha}(\mathrm{P}_{XY}), \tag{14}$$

where

$$R_f^{\alpha}(\mathrm{P}_{XY}) := \min_{\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}} \big[I(\tilde{U}, \tilde{V} \wedge \tilde{X}|\tilde{Y}) + I(\tilde{U}, \tilde{V} \wedge \tilde{Y}|\tilde{X}) + \alpha D(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}} \| \mathrm{Q}_{\tilde{U}\tilde{V}XY})$$

[12] See [31], [8] for the extension to multiple rounds.

$$+ (\alpha+2)D(\mathrm{P}_{\tilde{X}\tilde{Y}}\|\mathrm{P}_{XY}) + \alpha\big(D(\mathrm{P}_{\tilde{U}|\tilde{X}\tilde{Y}}\|\mathrm{P}_{\tilde{U}|\tilde{X}}|\mathrm{P}_{\tilde{X}\tilde{Y}}) + H(\tilde{F}|\tilde{Y},\tilde{U}) + H(\tilde{F}|\tilde{X},\tilde{U},\tilde{V})\big)\big]$$

$$= \min_{\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}} \big[I(\tilde{U},\tilde{V}\wedge\tilde{X}|\tilde{Y}) + I(\tilde{U},\tilde{V}\wedge\tilde{Y}|\tilde{X}) + (2\alpha+2)D(\mathrm{P}_{\tilde{X}\tilde{Y}}\|\mathrm{P}_{XY})$$

$$+ \alpha\big(2I(\tilde{U}\wedge\tilde{Y}|\tilde{X}) + I(\tilde{V}\wedge\tilde{X}|\tilde{Y},\tilde{U}) + H(\tilde{F}|\tilde{Y},\tilde{U}) + H(\tilde{F}|\tilde{X},\tilde{U},\tilde{V})\big)\big],$$

$\tilde{F} = f(\tilde{X},\tilde{Y})$, and $\mathrm{Q}_{\tilde{U}\tilde{V}XY} = \mathrm{P}_{\tilde{V}|\tilde{U}\tilde{Y}}\mathrm{P}_{\tilde{U}|\tilde{X}}\mathrm{P}_{XY}$ is the distribution induced from each $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}$ and respects the information constraints of the coding problem. The ranges $\mathcal{U}$ and $\mathcal{V}$ of $\tilde{U}$ and $\tilde{V}$ can be restricted to $|\mathcal{U}| \le |\mathcal{X}||\mathcal{Y}|$ and $|\mathcal{V}| \le |\mathcal{X}|^2|\mathcal{Y}|^2$. The two forms described above have different utilities as described in Remark 1. As in the previous section, we divide the proof of strong converse into two parts. The main technical component is the following result.

Theorem 6. *For every $n \in \mathbb{N}$ and $\alpha > 0$, we have*

$$R_f^\alpha(\mathrm{P}_{XY}^n) \ge nR_f^\alpha(\mathrm{P}_{XY}),$$

where in defining $R_f^\alpha(\mathrm{P}_{XY}^n)$ we use $\tilde{F}^n = (f(\tilde{X}_1,\tilde{Y}_1),...,f(\tilde{X}_n,\tilde{Y}_n))$.

As corollary, we get the strong converse theorem for function computation.

Corollary 7. *For every $0 < \varepsilon < 1$, we have $R_f(\varepsilon|\mathrm{P}_{XY}) = R_f(\mathrm{P}_{XY})$.*

The proof of corollary follows from Theorem 6 using similar steps as the proof of Corollary 5 where the changed measure $\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}$ is now obtained by conditioning on the set of inputs for which no error occurs, i.e., the set

$$\mathcal{D} = \{(x^n,y^n) : \psi_1(x^n,\varphi_2(y^n,\varphi_1(x^n))) = \psi_2(y^n,\varphi_1(x^n)) = (f(x_1,y_1),...,f(x_n,y_n))\}.$$

We close this section with a proof of Theorem 6.

Proof of Theorem 6: By setting

$$G_1(\mathrm{P}_{\tilde{X}^n,\tilde{Y}^n}) := \big[H(\tilde{X}^n|\tilde{Y}^n) + D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}_{X^nY^n})\big]$$

$$+ (2\alpha+1)\big[H(\tilde{Y}^n|\tilde{X}^n) + D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}_{X^nY^n})\big]$$

and

$$G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n\tilde{F}^n}) := -H(\tilde{X}^n|\tilde{Y}^n,\tilde{U},\tilde{V}) - H(\tilde{Y}^n|\tilde{X}^n,\tilde{U},\tilde{V}) - 2\alpha H(\tilde{Y}^n|\tilde{X}^n,\tilde{U})$$

$$+ \alpha I(\tilde{V}\wedge\tilde{X}^n|\tilde{Y}^n,\tilde{U}) + \alpha H(\tilde{F}^n|\tilde{Y}^n,\tilde{U}) + \alpha H(\tilde{F}^n|\tilde{X}^n,\tilde{U},\tilde{V})$$

for given $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}$ with $\tilde{F}_i = f(\tilde{X}_i, \tilde{Y}_i)$ for $1 \le i \le n$, we can write

$$R_f^\alpha(\mathrm{P}_{XY}^n) = \min_{\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}} \big[G_1(\mathrm{P}_{\tilde{X}^n,\tilde{Y}^n}) + G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n\tilde{F}^n}) \big].$$

Fix arbitrary $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}$. By Proposition 1, we get

$$G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}) \ge n G_1(\mathrm{P}_{\tilde{X}_J\tilde{Y}_J}), \tag{15}$$

where J is distributed uniformly over $\{1, ..., n\}$. For $G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n\tilde{F}^n})$, since removing condition increases entropy, we have

$$-H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}, \tilde{V}) - H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) \ge -nH(\tilde{X}_J|\tilde{Y}_J, \tilde{U}_J, J, \tilde{V}) - nH(\tilde{Y}_J|\tilde{X}_J, \tilde{U}_J, J, \tilde{V})$$

where $\tilde{U}_j = (\tilde{U}, \tilde{X}_j^-, \tilde{Y}_j^+)$. Furthermore, noting that

$$\begin{aligned}
&- 2H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}) + I(\tilde{V} \wedge \tilde{X}^n|\tilde{Y}^n, \tilde{U}) + H(\tilde{F}^n|\tilde{Y}^n, \tilde{U}) + H(\tilde{F}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) \\
&= 2[H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}) - H(\tilde{Y}^n|\tilde{X}^n, \tilde{U})] + [H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) - H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}, \tilde{V})] \\
&\quad - H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}) - H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) + H(\tilde{F}^n|\tilde{Y}^n, \tilde{U}) + H(\tilde{F}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) \\
&= 2[H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}) - H(\tilde{Y}^n|\tilde{X}^n, \tilde{U})] + [H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) - H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}, \tilde{V})] \\
&\quad - H(\tilde{X}^n, \tilde{F}^n|\tilde{Y}^n, \tilde{U}) - H(\tilde{Y}^n, \tilde{F}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) + H(\tilde{F}^n|\tilde{Y}^n, \tilde{U}) + H(\tilde{F}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) \\
&= 2[H(\tilde{X}^n|\tilde{U}) - H(\tilde{Y}^n|\tilde{U})] + [H(\tilde{Y}^n|\tilde{U}, \tilde{V}) - H(\tilde{X}^n|\tilde{U}, \tilde{V})] \\
&\quad - H(\tilde{X}^n|\tilde{Y}^n, \tilde{F}^n, \tilde{U}) - H(\tilde{Y}^n|\tilde{X}^n, \tilde{F}^n, \tilde{U}, \tilde{V}),
\end{aligned}$$

where we used the fact that $\tilde{F}^n$ is function of $(\tilde{X}^n, \tilde{Y}^n)$ to append $\tilde{F}^n$ in the second equality. Thus, by using (4) twice, we obtain

$$\begin{aligned}
&- 2H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}) + I(\tilde{V} \wedge \tilde{X}^n|\tilde{Y}^n, \tilde{U}) + H(\tilde{F}^n|\tilde{Y}^n, \tilde{U}) + H(\tilde{F}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) \\
&\ge 2n(H(\tilde{X}_J|\tilde{U}_J, J) - H(\tilde{Y}_J|\tilde{U}_J, J)) + n(H(\tilde{Y}_J|\tilde{U}_J, J, \tilde{V}) - H(\tilde{X}_J|\tilde{U}_J, J, \tilde{V})) \\
&\quad - nH(\tilde{X}_J|\tilde{Y}_J, \tilde{F}_J, \tilde{U}_J, J) - nH(\tilde{Y}_J|\tilde{X}_J, \tilde{F}_J, \tilde{U}_J, J, \tilde{V}) \\
&= -2nH(\tilde{Y}_J|\tilde{X}_J, \tilde{U}_J, J) + nI(\tilde{V} \wedge \tilde{X}_J|\tilde{Y}_J, \tilde{U}_J, J) + nH(\tilde{F}_J|\tilde{Y}_J, \tilde{U}_J, J) + nH(\tilde{F}_J|\tilde{X}_J, \tilde{U}_J, J, \tilde{V}),
\end{aligned}$$

where we used the fact $\tilde{F}_J = f(\tilde{X}_J, \tilde{Y}_J)$ to remove the unnecessary $\tilde{F}_J$ in the previous identity. Upon combining the bounds above, we obtain

$$G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n\tilde{F}^n}) \ge n G_2(\mathrm{P}_{(\tilde{U}_J,J)\tilde{V}\tilde{X}_J\tilde{Y}_J\tilde{F}_J}). \tag{16}$$

Since (15) and (16) hold for arbitrary $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}$, the proof is complete. □

V. Common Randomness Generation and Secret Key Agreement

We now move to the closely related problems of common randomness generation and secret key agreement. In these problems, an additional challenge arises due to the presence of a total variation distance constraint. We circumvent this difficulty by replacing the total variation distance constraint by a constraint on log-likelihood; the resulting set is used in our change of measure arguments. A similar approach will be used later for the wiretap channel strong converse where, too, the security constraint poses a similar challenge.

We begin with the common randomness problem and extend to the secret key agreement case using the connection between the two problems. Note that while the change of measure arguments presented here prove the strong converse for secret key agreement with limited communication, the strong converse for secret key agreement with unlimited communication is available in [43]. In fact, the conditional independence testing bound of [43] yields even the precise second-order term (cf. [24]); it is unclear if our change of measure approach can do the same.

A. Common Randomness Generation

Consider a source P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$. An (2-round) interactive common randomness generation protocol[13] π with input (X^n, Y^n) consists of mappings $\varphi_1 : \mathcal{X}^n \to \{0,1\}^{l_1}$ and $\varphi_2 : \mathcal{Y}^n \times \{0,1\}^{l_1} \to \{0,1\}^{l_2}$; the length $|\pi|$ of such a protocol π is $l_1 + l_2$. The random transcript of the protocol is denoted by $\Pi = (\Pi_1, \Pi_2)$, where $\Pi_1 = \varphi_1(X^n)$ and $\Pi_2 = \varphi_2(Y^n, \Pi_1)$.

Given a protocol, a pair of random variables (K_1, K_2) taking values in a finite set $\mathcal{K}$ constitute an (ε, δ)-CR recoverable from π if there exist $\psi_1 : \mathcal{X}^n \times \{0,1\}^{l_2} \to \mathcal{K}$ and $\psi_2 : \mathcal{Y}^n \times \{0,1\}^{l_1} \to \mathcal{K}$ such that $K_1 = \psi_1(X^n, \Pi_2)$, $K_2 = \psi_2(Y^n, \Pi_1)$, and

$$\mathbb{P}\left(K_1 \neq K_2\right) \leq \varepsilon, \tag{17}$$

$$d(\mathrm{P}_{K_1}, \mathrm{P}_{\mathtt{unif}}) \leq \delta, \tag{18}$$

where $\mathrm{P}_{\mathtt{unif}}$ is the uniform distribution on $\mathcal{K}$. The quantity $\log|\mathcal{K}|$ denotes the length of the CR.

[13] For ease of presentation, we restrict to 2-rounds. Our approach easily extends to higher (but fixed) number of rounds.

A rate pair $(R_{\mathtt{c}}, R_{\mathtt{r}})$ is (ε, δ)-achievable if, for all n sufficiently large, there exists a protocol π of length $|\pi| \leq nR_{\mathtt{c}}$ that recovers an (ε, δ)-CR of length $\log|\mathcal{K}| \geq nR_{\mathtt{r}}$. Let $\mathcal{R}_{\mathtt{CR}}(\varepsilon, \delta|\mathrm{P}_{XY})$ be the closure of the set of all (ε, δ)-achievable rate pairs. Define

$$\mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XY}) := \bigcap_{0<\varepsilon,\delta<1} \mathcal{R}_{\mathtt{CR}}(\varepsilon, \delta|\mathrm{P}_{XY}).$$

The following characterization of $\mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XY})$ was given in [3]:

$$\begin{aligned}\mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XY}) = \big\{(R_{\mathtt{c}}, R_{\mathtt{r}}) : \exists(U,V) \text{ s.t. } |\mathcal{U}| \leq |\mathcal{X}|+1, |\mathcal{V}| \leq |\mathcal{X}||\mathcal{Y}|+1, U \mathrel{-\!\!\circ\!\!-} X \mathrel{-\!\!\circ\!\!-} Y, V \mathrel{-\!\!\circ\!\!-} (Y,U) \mathrel{-\!\!\circ\!\!-} X \\ R_{\mathtt{c}} \geq I(U,V \wedge X|Y) + I(U,V \wedge Y|X), R_{\mathtt{r}} \leq I(U,V \wedge X,Y)\big\}\end{aligned}$$

The set $\mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XY})$ is closed and convex, and it can be expressed alternatively using tangent lines as follows:

$$\mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XY}) = \bigcap_{\mu \geq 0} \big\{(R_{\mathtt{r}}, R_{\mathtt{c}}) : R_{\mathtt{r}} - \mu R_{\mathtt{c}} \leq R_{\mathtt{CR}}^{\mu}(\mathrm{P}_{XY})\big\},$$

where

$$\begin{aligned}R_{\mathtt{CR}}^{\mu}(\mathrm{P}_{XY}) := \max\big\{&I(U,V \wedge X,Y) - \mu\big(I(U,V \wedge X|Y) + I(U,V \wedge Y|X)\big) : \\ &\exists(U,V) \text{ s.t. } |\mathcal{U}| \leq |\mathcal{X}|, |\mathcal{V}| \leq |\mathcal{X}||\mathcal{Y}|, U \mathrel{-\!\!\circ\!\!-} X \mathrel{-\!\!\circ\!\!-} Y, V \mathrel{-\!\!\circ\!\!-} (Y,U) \mathrel{-\!\!\circ\!\!-} X\big\}. \end{aligned} \tag{19}$$

As before, we circumvent the Markov chain conditions by using the following alternative form:

$$R_{\mathtt{CR}}^{\mu}(\mathrm{P}_{XY}) = \inf_{\alpha>0} R_{\mathtt{CR}}^{\mu,\alpha}(\mathrm{P}_{XY}), \tag{20}$$

where (see Remark 1)

$$\begin{aligned}R_{\mathtt{CR}}^{\mu,\alpha}(\mathrm{P}_{XY}) &:= \max_{\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}} \big[I(\tilde{U},\tilde{V} \wedge \tilde{X},\tilde{Y}) - \mu\big(I(\tilde{U},\tilde{V} \wedge \tilde{X}|\tilde{Y}) + I(\tilde{U},\tilde{V} \wedge \tilde{Y}|\tilde{X})\big) \\ &\qquad - \alpha D(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}} \| \mathrm{Q}_{\tilde{U}\tilde{V}XY}) - D(\mathrm{P}_{\tilde{U}\tilde{V}|\tilde{X}\tilde{Y}} \| \mathrm{P}_{\tilde{U}|\tilde{X}}\mathrm{P}_{\tilde{V}|\tilde{Y}\tilde{U}}|\mathrm{P}_{\tilde{X}\tilde{Y}}) - 2\mu D(\mathrm{P}_{\tilde{X}\tilde{Y}} \| \mathrm{P}_{XY})\big] \\ &= \max_{\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}} \big[I(\tilde{U},\tilde{V} \wedge \tilde{X},\tilde{Y}) - \mu\big(I(\tilde{U},\tilde{V} \wedge \tilde{X}|\tilde{Y}) + I(\tilde{U},\tilde{V} \wedge \tilde{Y}|\tilde{X})\big) \\ &\qquad - (\alpha+2\mu)D(\mathrm{P}_{\tilde{X}\tilde{Y}} \| \mathrm{P}_{XY}) - (\alpha+1)\big(I(\tilde{U} \wedge \tilde{Y}|\tilde{X}) + I(\tilde{V} \wedge \tilde{X}|\tilde{Y},\tilde{U})\big)\big],\end{aligned} \tag{21}$$

and $\mathrm{Q}_{\tilde{U}\tilde{V}XY} = \mathrm{P}_{\tilde{V}|\tilde{U}\tilde{Y}}\mathrm{P}_{\tilde{U}|\tilde{X}}\mathrm{P}_{XY}$ is the distribution induced from each $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}$. The ranges $\mathcal{U}$ and $\mathcal{V}$ of $\tilde{U}$ and $\tilde{V}$ can be restricted to $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}|$ and $|\mathcal{V}| \leq |\mathcal{X}|^2|\mathcal{Y}|^2$.

Theorem 8. *For every $n \in \mathbb{N}$, $\mu \geq 0$, and $\alpha > 0$, we have*

$$R_{\mathtt{CR}}^{\mu,\alpha}(\mathrm{P}_{XY}^{n}) \leq n R_{\mathtt{CR}}^{\mu,\alpha}(\mathrm{P}_{XY}).$$

Corollary 9. *For every $0 < \varepsilon, \delta < 1$ with $\varepsilon + \delta < 1$, we have $\mathcal{R}_{\mathtt{CR}}(\varepsilon, \delta | \mathrm{P}_{XY}) = \mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XY})$.*

Proof of Corollary 9: For a given protocol π and (ε, δ)-CR (K_1, K_2) satisfying (17) and (18), we first replace the uniformity constraint (18) with a constraint on log-likelihood. Specifically, for a given $\gamma > 0$, which will be specified later, let

$$\mathcal{T}_\gamma := \left\{ k \in \mathcal{K} : \log \frac{1}{\mathrm{P}_{K_1}(k)} \geq \log |\mathcal{K}| - \gamma \right\}. \tag{22}$$

Then, by (18) and the standard argument in the information-spectrum methods (cf. [21, Lemma 2.1.2]), we have

$$\begin{aligned} \delta &\geq d(\mathrm{P}_{K_1}, \mathrm{P}_{\mathtt{unif}}) \\ &\geq \mathrm{P}_{K_1}\left(\mathcal{T}_\gamma^c\right) - \mathrm{P}_{\mathtt{unif}}\left(\mathcal{T}_\gamma^c\right) \\ &\geq \mathrm{P}_{K_1}\left(\mathcal{T}_\gamma^c\right) - 2^{-\gamma}. \end{aligned} \tag{23}$$

We now define the set $\mathcal{D}$ over which our CR generation protocol behaves ideally. Let

$$\mathcal{D} := \left\{ (x^n, y^n) : \psi_1(x^n, \varphi_2(y^n, \varphi_1(x^n))) \in \mathcal{T}_\gamma,\ \psi_1(x^n, \varphi_2(y^n, \varphi_1(x^n))) = \psi_2(y^n, \varphi_1(x^n)) \right\}. \tag{24}$$

By (17) and (23), for $\gamma = \log \frac{2}{1-\varepsilon-\delta}$, we have

$$\begin{aligned} \mathrm{P}_{XY}^n(\mathcal{D}) &\geq 1 - \mathrm{P}_{K_1}\left(\mathcal{T}_\gamma^c\right) - \mathbb{P}\left(K_1 \neq K_2\right) \\ &\geq 1 - \varepsilon - \delta - 2^{-\gamma} \\ &= \frac{1-\varepsilon-\delta}{2}. \end{aligned} \tag{25}$$

Denote by $\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}$ the pmf

$$\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) := \frac{\mathrm{P}_{XY}^n(x^n, y^n)\mathbb{1}[(x^n, y^n) \in \mathcal{D}]}{\mathrm{P}_{XY}^n(\mathcal{D})}. \tag{26}$$

Then, (25) implies

$$\begin{aligned} D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n} \| \mathrm{P}_{XY}^n) &= \log \frac{1}{\mathrm{P}_{XY}^n(\mathcal{D})} \\ &\leq \log \frac{2}{1-\varepsilon-\delta}. \end{aligned} \tag{27}$$

Consider an execution of protocol π for input $(\tilde{X}^n, \tilde{Y}^n) \sim \mathrm{P}_{\tilde{X}^n\tilde{Y}^n}$. Set $\tilde{\Pi}_1 = \varphi_1(\tilde{X}^n)$, $\tilde{\Pi}_2 = \varphi_2(\tilde{Y}^n, \tilde{\Pi}_1)$, $\tilde{\Pi} = (\tilde{\Pi}_1, \tilde{\Pi}_2)$, $\tilde{K}_1 = \psi_1(\tilde{X}^n, \tilde{\Pi}_2)$, and $\tilde{K}_2 = \psi_2(\tilde{Y}^n, \tilde{\Pi}_1)$. Note that $\tilde{K}_1 = \tilde{K}_2$ with probability 1. Furthermore, since the support of $\mathrm{P}_{\tilde{K}_1}$ satisfies $\mathtt{supp}(\mathrm{P}_{\tilde{K}_1}) \subseteq \mathcal{T}_\gamma$, we have

$$\begin{aligned}\mathrm{P}_{\tilde{K}_1}(k) &= \frac{1}{\mathrm{P}^n_{XY}(\mathcal{D})} \sum_{\substack{(x^n,y^n)\in\mathcal{D}: \\ \psi_1(x^n,\varphi_2(y^n,\varphi_1(x^n)))=k}} \mathrm{P}^n_{XY}(x^n,y^n) \\ &\le \frac{\mathrm{P}_{K_1}(k)}{\mathrm{P}^n_{XY}(\mathcal{D})} \\ &\le \frac{2^\gamma}{\mathrm{P}^n_{XY}(\mathcal{D})|\mathcal{K}|},\end{aligned}$$

for every $k \in \mathtt{supp}(\mathrm{P}_{\tilde{K}_1})$. Thus, we get

$$\begin{aligned}H_{\min}(\mathrm{P}_{\tilde{K}_1}) &= \min_{k\in\mathtt{supp}(\mathrm{P}_{\tilde{K}_1})} \log \frac{1}{\mathrm{P}_{\tilde{K}_1}(k)} \\ &\ge \log|\mathcal{K}| - 2\log\frac{2}{1-\varepsilon-\delta},\end{aligned}$$

where we used (25) once again in the inequality.

By noting $H_{\min}(\mathrm{P}_{\tilde{K}_1}) \le H(\tilde{K}_1)$, we have

$$\begin{aligned}n(R_{\mathtt{r}} - \mu R_{\mathtt{c}}) - 2\log\frac{2}{1-\varepsilon-\delta} &\le \log|\mathcal{K}| - \mu|\pi| - 2\log\frac{2}{1-\varepsilon-\delta} \\ &\le H(\tilde{K}_1) - \mu H(\tilde{\Pi}) \\ &\le H(\tilde{\Pi}, \tilde{K}_1) - \mu\big(H(\tilde{\Pi}|\tilde{X}^n) + H(\tilde{\Pi}|\tilde{Y}^n)\big) \\ &= H(\tilde{\Pi}, \tilde{K}_2) - \mu\big(H(\tilde{\Pi}, \tilde{K}_2|\tilde{X}^n) + H(\tilde{\Pi}, \tilde{K}_2|\tilde{Y}^n)\big) \\ &\le I(\tilde{\Pi}, \tilde{K}_2 \wedge \tilde{X}^n, \tilde{Y}^n) - \mu\big(I(\tilde{\Pi}, \tilde{K}_2 \wedge \tilde{X}^n|\tilde{Y}^n) + I(\tilde{\Pi}, \tilde{K}_2 \wedge \tilde{Y}^n|\tilde{X}^n)\big) \\ &\le I(\tilde{\Pi}, \tilde{K}_2 \wedge \tilde{X}^n, \tilde{Y}^n) - \mu\big(I(\tilde{\Pi}, \tilde{K}_2 \wedge \tilde{X}^n|\tilde{Y}^n) + I(\tilde{\Pi}, \tilde{K}_2 \wedge \tilde{Y}^n|\tilde{X}^n)\big) \\ &\quad - (\alpha+1)\big(I(\tilde{\Pi}_1 \wedge \tilde{Y}^n|\tilde{X}^n) + I(\tilde{\Pi}_2, \tilde{K}_2 \wedge \tilde{X}^n|\tilde{Y}^n, \tilde{\Pi}_1)\big) \\ &\quad - (\alpha+2\mu)D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}^n_{XY}) + (\alpha+2\mu)\log\frac{2}{1-\varepsilon-\delta} \\ &\le R^{\mu,\alpha}_{\mathtt{CR}}(\mathrm{P}^n_{XY}) + (\alpha+2\mu)\log\frac{2}{1-\varepsilon-\delta},\end{aligned}$$

where we used a well-known property of interactive communication in the third inequality (eg. see [33, Eq. (3.2)]); the identity follows from the fact that $\tilde{K}_1$ and $\tilde{K}_2$ are recoverable perfectly from $(\tilde{X}^n, \tilde{\Pi})$ and $(\tilde{Y}^n, \tilde{\Pi})$, respectively, and $\tilde{K}_1 = \tilde{K}_2$ with probability 1; and we used the fact that costs $I(\tilde{\Pi}_1 \wedge \tilde{Y}^n|\tilde{X}^n)$ and $I(\tilde{\Pi}_2, \tilde{K}_2 \wedge \tilde{X}^n|\tilde{Y}^n, \tilde{\Pi}_1)$ are both 0 and (27) in the fifth inequality; in the last inequality, we regarded

$\tilde{\Pi}_1$ and $(\tilde{\Pi}_2, \tilde{K}_2)$ as $\tilde{U}$ and $\tilde{V}$, respectively. Finally, by applying Theorem 8, we have

$$R_{\mathtt{r}} - \mu R_{\mathtt{c}} \leq R_{\mathtt{CR}}^{\mu,\alpha}(\mathrm{P}_{XY}) + \frac{(\alpha + 2\mu + 2)}{n} \log \frac{2}{1 - \varepsilon - \delta},$$

which together with (20) imply the strong converse. □

Proof of Theorem 8: First note that we can expand

$$I(\tilde{U}, \tilde{V} \wedge \tilde{X}^n, \tilde{Y}^n) = I(\tilde{U} \wedge \tilde{X}^n) + I(\tilde{V} \wedge \tilde{Y}^n|\tilde{U}) + I(\tilde{U} \wedge \tilde{Y}^n|\tilde{X}^n) + I(\tilde{V} \wedge \tilde{X}^n|\tilde{Y}^n, \tilde{U}).$$

Then, by setting

$$G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}) := H(\tilde{X}^n) - \mu\big[H(\tilde{X}^n|\tilde{Y}^n) + D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}_{XY}^n)\big] - (\alpha + \mu)\big[H(\tilde{Y}^n|\tilde{X}^n) + D(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}\|\mathrm{P}_{XY}^n)\big]$$

and

$$\begin{aligned} G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}) := -H(\tilde{X}^n|\tilde{U}) + I(\tilde{V} \wedge \tilde{Y}^n|\tilde{U}) + \mu\big(H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}, \tilde{V}) + H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}, \tilde{V})\big) \\ + \alpha\big(H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}) - I(\tilde{V} \wedge \tilde{X}^n|\tilde{Y}^n, \tilde{U})\big) \end{aligned}$$

for given $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}$, we can write

$$R_{\mathtt{CR}}^{\mu,\alpha}(\mathrm{P}_{XY}^n) = \max_{\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}} \big[G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}) + G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n})\big].$$

Fix arbitrary $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}$. By noting $H(\tilde{X}^n) \leq nH(\tilde{X}_J)$ and by using Proposition 1, we get

$$G_1(\mathrm{P}_{\tilde{X}^n\tilde{Y}^n}) \leq nG_1(\mathrm{P}_{\tilde{X}_J\tilde{Y}_J}), \tag{28}$$

where J is distributed uniformly over $\{1, \ldots, n\}$. For $G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n})$, by using (4), we have

$$\begin{aligned} -H(\tilde{X}^n|\tilde{U}) + I(\tilde{V} \wedge \tilde{Y}^n|\tilde{U}) &= H(\tilde{Y}^n|\tilde{U}) - H(\tilde{X}^n|\tilde{U}) - H(\tilde{Y}^n|\tilde{U}, \tilde{V}) \\ &\leq n\big[H(\tilde{Y}_J|\tilde{U}_J, J) - H(\tilde{X}_J|\tilde{U}_J, J) - H(\tilde{Y}_J|\tilde{U}_J, J, \tilde{V})\big], \end{aligned}$$

where $\tilde{U}_j = (\tilde{U}, \tilde{X}_j^-, \tilde{Y}_j^+)$. Also,

$$H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}, \tilde{V}) + H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}, \tilde{V}) \leq n\big(H(\tilde{X}_J|\tilde{Y}_J, \tilde{U}_J, J, \tilde{V}) + H(\tilde{Y}_J|\tilde{X}_J, \tilde{U}_J, J, \tilde{V})\big).$$

Furthermore, by using (4) once more, we obtain

$$\begin{aligned} H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}) - I(\tilde{V} \wedge \tilde{X}^n|\tilde{Y}^n, \tilde{U}) &= H(\tilde{Y}^n|\tilde{U}) - H(\tilde{X}^n|\tilde{U}) + H(\tilde{X}^n|\tilde{Y}^n, \tilde{U}, \tilde{V}) \\ &\leq n\big(H(\tilde{Y}_J|\tilde{U}_J, J) - H(\tilde{X}_J|\tilde{U}_J, J) + H(\tilde{X}_J|\tilde{Y}_J, \tilde{U}_J, J, \tilde{V})\big) \end{aligned}$$

$$= n\big(H(\tilde{Y}_J|\tilde{X}_J, \tilde{U}_J, J) - I(\tilde{V} \wedge \tilde{X}_J|\tilde{Y}_J, \tilde{U}_J, J)\big).$$

Upon combining the bounds above, we obtain

$$G_2(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}) \le nG_2(\mathrm{P}_{\tilde{U}_J J\tilde{V}\tilde{X}_J\tilde{Y}_J}). \tag{29}$$

Since (28) and (29) hold for arbitrary $\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}^n\tilde{Y}^n}$, the proof is complete. □

Remark 3. When randomization is allowed, the achievable region is given by

$$\tilde{\mathcal{R}}_{\mathtt{CR}}(\mathrm{P}_{XY}) = \big\{(R_{\mathtt{c}}, R_{\mathtt{r}}) : \exists\, t \ge 0 \text{ s.t. } (R_{\mathtt{c}} - t, R_{\mathtt{r}} - t) \in \mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XY})\big\}.$$

We can extend the proof above to randomized protocols easily by appending two independent i.i.d. sources A^n and B^n (taking values in sufficiently large alphabets $\mathcal{A}$ and $\mathcal{B}$) to X^n and Y^n, respectively. By Corollary 9 we obtain $\mathcal{R}_{\mathtt{CR}}(\varepsilon, \delta|\mathrm{P}_{XAYB}) = \mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XAYB})$. Also, noting that (cf. [3])

$$\bigcup_{\mathrm{P}_{AB}} \mathcal{R}_{\mathtt{CR}}(\mathrm{P}_{XAYB}) = \tilde{\mathcal{R}}_{\mathtt{CR}}(\mathrm{P}_{XY}),$$

where the union is taken over all distributions such that A and B are independent, we have the strong converse even with randomized protocols. A similar approach has been pursued in [40, proof of Theorem III.2] to handle randomization.

B. Secret Key Agreement

Next, we consider the secret key agreement problem. The formulation and analysis is very similar to the common randomness generation problem; we only highlight the differences. Specifically, an (ε, δ)-SK (K_1, K_2) recoverable from a protocol π is an (ε, δ)-CR with the uniformity condition (18) replaced by the secrecy condition

$$d(\mathrm{P}_{K_1\Pi}, \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{\Pi}) \le \delta. \tag{30}$$

An (ε, δ)-achievable secret key rate pair $(R_{\mathtt{c}}, R_{\mathtt{s}})$ and the rate regions $\mathcal{R}_{\mathtt{SK}}(\varepsilon, \delta|\mathrm{P}_{XY})$ and $\mathcal{R}_{\mathtt{SK}}(\mathrm{P}_{XY})$ are defined exactly as before. The following characterization of $\mathcal{R}_{\mathtt{SK}}(\mathrm{P}_{XY})$ was given in [42]:

$$\begin{aligned}\mathcal{R}_{\mathtt{SK}}(\mathrm{P}_{XY}) = \big\{(R_{\mathtt{c}}, R_{\mathtt{s}}) : \exists (U,V) \text{ s.t. } &|\mathcal{U}| \le |\mathcal{X}| + 1, |\mathcal{V}| \le |\mathcal{X}||\mathcal{Y}| + 1, U -\!\circ\!- X -\!\circ\!- Y, V -\!\circ\!- (Y,U) -\!\circ\!- X\\ &R_{\mathtt{c}} \ge I(U,V \wedge X|Y) + I(U,V \wedge Y|X),\\ &R_{\mathtt{s}} \le I(U,V \wedge X,Y) - I(U,V \wedge X|Y) - I(U,V \wedge Y|X)\big\}.\end{aligned}$$

Define $R^{\mu}_{\mathtt{SK}}(\mathrm{P}_{XY})$ and $R^{\mu,\alpha}_{\mathtt{SK}}(\mathrm{P}_{XY})$ analogously to (19) and (21), respectively. Then, it can be easily verified that

$$R^{\mu}_{\mathtt{SK}}(\mathrm{P}_{XY}) = R^{\mu+1}_{\mathtt{CR}}(\mathrm{P}_{XY}),$$
$$R^{\mu,\alpha}_{\mathtt{SK}}(\mathrm{P}_{XY}) = R^{\mu+1,\alpha}_{\mathtt{CR}}(\mathrm{P}_{XY}).$$

Thus, Theorem 8 can be rewritten as follows.

Theorem 10. *For every $n \in \mathbb{N}$, $\mu \geq 0$, and $\alpha > 0$, we have*

$$R^{\mu,\alpha}_{\mathtt{SK}}(\mathrm{P}^n_{XY}) \leq nR^{\mu,\alpha}_{\mathtt{SK}}(\mathrm{P}_{XY}).$$

Furthermore, by using Theorem 10, we have the following strong converse theorem.

Corollary 11. *For every $0 < \varepsilon, \delta < 1$ with $\varepsilon + \delta < 1$, we have $\mathcal{R}_{\mathtt{SK}}(\varepsilon, \delta|\mathrm{P}_{XY}) = \mathcal{R}_{\mathtt{SK}}(\mathrm{P}_{XY})$.*

Proof of Corollary 11: The proof is mostly the same as the proof of Corollary 9; we only highlight the modifications required. Instead of the set defined by (22), we consider

$$\mathcal{T}_\gamma := \left\{(k, \tau) : \log \frac{1}{\mathrm{P}_{K_1|\Pi}(k|\tau)} \geq \log|\mathcal{K}| - \gamma\right\}.$$

We can verify that

$$\delta \geq \mathrm{P}_{K_1\Pi}\left(\mathcal{T}^c_\gamma\right) - 2^{-\gamma}.$$

In place of (24), define the set $\mathcal{D}$ as

$$\mathcal{D} := \big\{(x^n, y^n) : (\psi_1(x^n, \varphi_2(y^n, \varphi_1(x^n))), \varphi_1(x^n), \varphi_2(y^n, \varphi_1(x^n))) \in \mathcal{T}_\gamma,$$
$$\psi_1(x^n, \varphi_2(y^n, \varphi_1(x^n))) = \psi_2(y^n, \varphi_1(x^n))\big\}.$$

Then, for the changed measure (26), we recover the bound (27). Also,

$$H_{\min}(\mathrm{P}_{\tilde{K}_1\tilde{\Pi}}|\mathrm{P}_{\tilde{\Pi}}) := \min_{(k,\tau)\in\mathtt{supp}(\mathrm{P}_{\tilde{K}_1\tilde{\Pi}})} \log \frac{1}{\mathrm{P}_{\tilde{K}_1|\tilde{\Pi}}(k|\tau)}$$
$$\geq \log|\mathcal{K}| - 2\log\frac{2}{1-\varepsilon-\delta}.$$

Therefore, upon noting $H_{\min}(\mathrm{P}_{\tilde{K}_1\tilde{\Pi}}|\mathrm{P}_{\tilde{\Pi}}) \leq H(\tilde{K}_1|\tilde{\Pi})$, we obtain

$$n(R_{\mathtt{s}} - \mu R_{\mathtt{c}}) - 2\log\frac{2}{1-\varepsilon-\delta} \leq H(\tilde{K}_1|\tilde{\Pi}) - \mu H(\tilde{\Pi})$$

$$
\begin{aligned}
&= H(\tilde{\Pi}, \tilde{K}_1) - (\mu + 1)H(\tilde{\Pi}) \\
&\leq R^{\mu,\alpha}_{\mathtt{SK}}(\mathrm{P}^n_{XY}) + (\alpha + 2\mu + 2)\log\frac{2}{1-\varepsilon-\delta}.
\end{aligned}
$$

Finally, the strong converse follows from Theorem 10. □

VI. Wiretap Channel

A wiretap channel code enables reliable transmission of a message over a noisy channel while keeping it secure from an eavesdropper who can see another noisy version of transmissions. Formally, given a discrete memoryless channel (DMC) $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$, an $(N, n, \varepsilon, \delta)$-wiretap code for W consists of a (possibly randomized) encoder $\varphi : \{1, ..., N\} \to \mathcal{X}^n$ and a decoder $\psi : \mathcal{Y}^n \to \{1, ..., N\}$ such that when a message M distributed uniformly over $\{1, ..., N\}$ is transmitted over the channel as $X^n = \varphi(M)$, the estimate $\hat{M} = \psi(Y^n)$ has probability of error satisfying $\mathbb{P}(\hat{M} \neq M) \leq \varepsilon$ and leakage $d(\mathrm{P}_{MZ^n}, \mathrm{P}_M \times \mathrm{P}_{Z^n}) \leq \delta$.

A rate $R > 0$ is (ε, δ)-achievable if there exists an $(\lfloor 2^{nR} \rfloor, n, \varepsilon, \delta)$-wiretap code for all n sufficiently large. The (ε, δ)-wiretap capacity $C_{\mathtt{s}}(\varepsilon, \delta|W)$ is the supremum over all (ε, δ)-achievable rates. The wiretap capacity $C_{\mathtt{s}}(W)$ is the infimum of $C_{\mathtt{s}}(\varepsilon, \delta|W)$ over all $\varepsilon, \delta \in (0, 1)$. The following characterization of $C_{\mathtt{s}}(W)$ was derived in [10]:

$$
C_{\mathtt{s}}(W) = \max_{\substack{\mathrm{P}_{UXYZ}: \\ \mathrm{P}_{YZ|XU} = W}} \big[I(U \wedge Y) - I(U \wedge Z) \big],
$$

where the cardinality $|\mathcal{U}|$ of U can be restricted to be $|\mathcal{U}| \leq |\mathcal{X}|$. Using Proposition 3, the expression on the right above can be written alternatively as[14]

$$
C_{\mathtt{s}}(W) = \inf_{\alpha > 0} C^{\alpha}_{\mathtt{s}}(W), \tag{31}
$$

where

$$
C^{\alpha}_{\mathtt{s}}(W) = \max_{\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}}} \big[I(\tilde{U} \wedge \tilde{Y}) - I(\tilde{U} \wedge \tilde{Z}) - \alpha D(\mathrm{P}_{\tilde{Y}\tilde{Z}|\tilde{X}\tilde{U}} \| W | \mathrm{P}_{\tilde{X}\tilde{U}}) \big],
$$

where the cardinality $|\mathcal{U}|$ of U can be restricted to be $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}||\mathcal{Z}|$. The next theorem shows that the quantity $C^{\alpha}_{\mathtt{s}}(W)$ satisfies the required sub-additivity property.

Theorem 12. *Consider a DMC $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ such that $W(y, z|x) = W_1(y|x)W_2(z|x)$. For every $n \in \mathbb{N}$ and $\alpha > 0$,*

$$
C^{2\alpha}_{\mathtt{s}}(W^n) \leq nC^{\alpha}_{\mathtt{s}}(W).
$$

[14] The proof of (31) is very similar to that of other variational formulae such as (7) and is omitted.

As a corollary, we obtain the strong converse for wiretap channel.[15]

Corollary 13. *For every* $0 < \varepsilon, \delta < 1$ *with* $\varepsilon + \delta < 1$*, we have* $C_{\mathtt{s}}(\varepsilon, \delta|W) = C_{\mathtt{s}}(W)$.

Proof of Corollary 13: Consider an $(N, n, \varepsilon, \delta)$-wiretap code with a randomized encoder φ and (deterministic) decoder ψ . Note that without loss of generality we may assume $W(y, z|x) = W_1(y|x)W_2(z|x)$ since the error and secrecy criterion, respectively, depend only on the marginals (X^n, Y^n) and (X^n, Z^n). The first step in our proof is to convert the average probability of error and secrecy requirements to a worst-case version. Specifically, since

$$\begin{aligned}\varepsilon + \delta &\geq \mathbb{P}(\hat{M} \neq M) + d(\mathrm{P}_{MZ^n}, \mathrm{P}_M \times \mathrm{P}_{Z^n}) \\ &= \frac{1}{N}\sum_{m=1}^{N}\left[\mathbb{P}(\hat{M} \neq m|M = m) + d(\mathrm{P}_{Z^n|M=m}, \mathrm{P}_{Z^n})\right],\end{aligned}$$

there exists a subset $\mathcal{M}'$ of size $|\mathcal{M}'| \geq (1 - \varepsilon - \delta)N/(1 + \varepsilon + \delta)$ such that for every message $m \in \mathcal{M}'$,

$$\mathbb{P}(\hat{M} \neq m|M = m) + d(\mathrm{P}_{Z^n|M=m}, \mathrm{P}_{Z^n}) \leq \frac{1 + \varepsilon + \delta}{2}.$$

For $m \in \mathcal{M}'$, consider the sets

$$\mathcal{A}_m = \{y^n : \psi(y^n) = m\}$$

and, for $\gamma > 0$ specified later,

$$\mathcal{B}_m = \left\{z^n : \log\frac{\mathrm{P}_{Z^n|M}(z^n|m)}{\mathrm{P}_{Z^n}(z^n)} \leq \gamma\right\}.$$

The set $\mathcal{B}_m$ denotes, roughly, the set of observations that do not reveal much information to the wiretapper about the message m – the wiretapper cannot distinguish reliably if the observation was generated from $\mathrm{P}_{Z^n|M=m}$ or from P_{Z^n}. By the standard argument in the information-spectrum methods (cf. [21]), the set $\mathcal{B}_m$ satisfies

$$\mathrm{P}_{Z^n}(\mathcal{B}_m^c) \leq 2^{-\gamma}.$$

Furthermore, from the definition of the total variation distance, we further have

$$\mathrm{P}_{Z^n|M=m}(\mathcal{B}_m^c) \leq 2^{-\gamma} + d(\mathrm{P}_{Z^n|M=m}, \mathrm{P}_{Z^n})$$

[15] We remark that we consider strong converse for the wiretap channel only when information leakage is measured by $d(\mathrm{P}_{MZ^n}, \mathrm{P}_M \times \mathrm{P}_{Z^n})$, and not for other measures of secrecy such as those considered in [7].

for every $m \in \mathcal{M}'$. Therefore, upon choosing $2^{-\gamma} = (1-\varepsilon-\delta)/4$, we have

$$\begin{aligned}\mathbb{P}\left(Y^n \in \mathcal{A}_m, Z^n \in \mathcal{B}_m | M = m\right) &\geq 1 - \mathrm{P}_{Y^n|M=m}\left(\mathcal{A}_m^c\right) - \mathrm{P}_{Z^n|M=m}\left(\mathcal{B}_m^c\right) \\ &\geq \frac{1-\varepsilon-\delta}{4}\end{aligned}$$

for every $m \in \mathcal{M}'$. Denote $\eta = 1 - (1-\varepsilon-\delta)/4$ and by $\mathcal{C}_m$ the set of $x^n \in \mathtt{supp}(\mathrm{P}_{X^n|M=m})$ such that

$$\mathbb{P}\left(Y^n \in \mathcal{A}_m, Z^n \in \mathcal{B}_m | X^n = x^n\right) \geq 1 - \sqrt{\eta}, \tag{32}$$

which satisfies

$$\mathbb{P}\left(X^n \in \mathcal{C}_m | M = m\right) \geq 1 - \sqrt{\eta} \tag{33}$$

by the reverse Markov inequality. We now define our modified random variables for which the code is perfectly error-free and has a small leakage of information to the wiretapper; however, unlike the original random variables satisfying the Markov constraint $M -\!\!\circ\!\!- X^n -\!\!\circ\!\!- (Y^n, Z^n)$, the modified random variables do not satisfy the Markov constraint (see also Remark 4). Specifically, consider random variables $(\tilde{U}, \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n)$ such that $\tilde{U}$ is uniformly distributed on $\mathcal{M}'$, and

$$\begin{aligned}\mathrm{P}_{\tilde{X}^n|\tilde{U}}\left(x^n|m\right) &= \frac{\mathrm{P}_{X^n|M}\left(x^n|m\right)\mathbb{1}[x^n \in \mathcal{C}_m]}{\mathrm{P}_{X^n|M}\left(\mathcal{C}_m|m\right)}; \\ \mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\left(y^n, z^n|x^n, m\right) &= \frac{\mathrm{P}_{Y^nZ^n|X^n}\left(y^n, z^n|x^n\right)\mathbb{1}[y^n \in \mathcal{A}_m, z^n \in \mathcal{B}_m]}{\mathrm{P}_{Y^nZ^n|X^n}\left(\mathcal{A}_m \times \mathcal{B}_m|x^n\right)}, \quad \forall x^n \in \mathcal{C}_m.\end{aligned}$$

Using the conditional independence assumption $W(y,z|x) = W_1(y|x)W_2(z|x)$, we further get that

$$\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\left(y^n, z^n|x^n, m\right) = \frac{\mathrm{P}_{Y^n|X^n}\left(y^n|x^n\right)\mathbb{1}[y^n \in \mathcal{A}_m]}{\mathrm{P}_{Y^n|X^n}\left(\mathcal{A}_m|x^n\right)} \cdot \frac{\mathrm{P}_{Z^n|X^n}\left(z^n|x^n\right)\mathbb{1}[z^n \in \mathcal{B}_m]}{\mathrm{P}_{Z^n|X^n}\left(\mathcal{B}_m|x^n\right)},$$

whereby

$$\mathrm{P}_{\tilde{Z}^n|\tilde{X}^n\tilde{U}}\left(z^n|x^n, m\right) = \frac{\mathrm{P}_{Z^n|X^n}\left(z^n|x^n\right)\mathbb{1}[z^n \in \mathcal{B}_m]}{\mathrm{P}_{Z^n|X^n}\left(\mathcal{B}_m|x^n\right)}. \tag{34}$$

Since $\tilde{U} = \psi(\tilde{Y}^n)$ with probability 1, we get

$$\log N - \log\frac{2}{1-\varepsilon-\delta} \leq \log|\mathcal{M}'| \leq I(\tilde{U} \wedge \tilde{Y}^n). \tag{35}$$

To bound the leakage $I(\tilde{U} \wedge \tilde{Z}^n)$, note that

$$I(\tilde{U} \wedge \tilde{Z}^n) + D(\mathrm{P}_{\tilde{Z}^n}\|\mathrm{P}_{Z^n}) = \mathbb{E}\left[\log\frac{\mathrm{P}_{\tilde{Z}^n|\tilde{U}}(\tilde{Z}^n|\tilde{U})}{\mathrm{P}_{Z^n}(\tilde{Z}^n)}\right]$$

$$
\begin{aligned}
&\leq \max_{m\in\mathcal{M}', z^n\in\mathcal{B}_m} \log \frac{\mathrm{P}_{\tilde{Z}^n|\tilde{U}}\left(z^n|m\right)}{\mathrm{P}_{Z^n}\left(z^n\right)} \\
&\leq \log \frac{4}{1-\varepsilon-\delta} + \max_{m\in\mathcal{M}', z^n\in\mathcal{B}_m} \log \frac{\mathrm{P}_{\tilde{Z}^n|\tilde{U}}\left(z^n|m\right)}{\mathrm{P}_{Z^n|M}\left(z^n|m\right)},
\end{aligned}
$$

where the previous inequality uses the definition of $\mathcal{B}_m$ and $\gamma = \log \frac{4}{1-\varepsilon-\delta}$. For the second term on the right-side above, for every $z^n \in \mathcal{B}_m$ it holds that

$$
\begin{aligned}
\frac{\mathrm{P}_{\tilde{Z}^n|\tilde{U}}\left(z^n|m\right)}{\mathrm{P}_{Z^n|M}\left(z^n|m\right)} &= \frac{\sum_{x^n\in\mathcal{C}_m} \mathrm{P}_{\tilde{X}^n|\tilde{U}}\left(x^n|m\right)\mathrm{P}_{\tilde{Z}^n|\tilde{X}^n\tilde{U}}\left(z^n|x^n,m\right)}{\sum_{x'^n\in\mathcal{X}^n} \mathrm{P}_{X^n|M}\left(x'^n|m\right)\mathrm{P}_{Z^n|X^n}\left(z^n|x'^n\right)} \\
&\leq \frac{\sum_{x^n\in\mathcal{C}_m} \mathrm{P}_{\tilde{X}^n|\tilde{U}}\left(x^n|m\right)\mathrm{P}_{\tilde{Z}^n|\tilde{X}^n\tilde{U}}\left(z^n|x^n,m\right)}{\sum_{x'^n\in\mathcal{C}_m} \mathrm{P}_{X^n|M}\left(x'^n|m\right)\mathrm{P}_{Z^n|X^n}\left(z^n|x'^n\right)} \\
&= \frac{\sum_{x^n\in\mathcal{C}_m} \mathrm{P}_{X^n|M}\left(x^n|m\right)\mathrm{P}_{Z^n|X^n}\left(z^n|x^n\right)/\{\mathrm{P}_{X^n|M}\left(\mathcal{C}_m|m\right)\mathrm{P}_{Z^n|X^n}\left(\mathcal{B}_m|x^n\right)\}}{\sum_{x'^n\in\mathcal{C}_m} \mathrm{P}_{X^n|M}\left(x'^n|m\right)\mathrm{P}_{Z^n|X^n}\left(z^n|x'^n\right)} \\
&\leq \frac{1}{(1-\sqrt{\eta})^2},
\end{aligned}
$$

where the second equality uses (34) and the final inequality is by (32) and (33). Using the bounds above, we get

$$
I(\tilde{U} \wedge \tilde{Z}^n) \leq \log \frac{4}{1-\varepsilon-\delta} + 2\log \frac{1}{1-\sqrt{\eta}}.
$$

Combining this bound with (35), for every $\alpha > 0$ and with

$$
\Delta(\varepsilon,\delta) = 2\log\frac{1}{1-\varepsilon-\delta} + 2\log\frac{1}{1-\sqrt{1-(1-\varepsilon-\delta)/4}} + 3,
$$

we get

$$
\begin{aligned}
\log N &\leq I(\tilde{U}\wedge\tilde{Y}^n) - I(\tilde{U}\wedge\tilde{Z}^n) + \Delta(\varepsilon,\delta) \\
&\leq C_{\mathsf{s}}^{2\alpha}(W^n) + 2\alpha D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) + \Delta(\varepsilon,\delta) \\
&\leq nC_{\mathsf{s}}^{\alpha}(W) + 2\alpha D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) + \Delta(\varepsilon,\delta), \qquad (36)
\end{aligned}
$$

where the final bound uses Theorem 12. It only remains to bound the divergence term on the right-side above. To that end, note

$$
\begin{aligned}
D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) &= \sum_{x^n,m} \mathrm{P}_{\tilde{X}^n\tilde{U}}\left(x^n,m\right)\log\frac{1}{\mathrm{P}_{Y^nZ^n|X^n}\left(\mathcal{A}_m\times\mathcal{B}_m|x^n\right)} \\
&\leq \log\frac{1}{1-\sqrt{\eta}},
\end{aligned}
$$

where we have used the fact that support of $\mathrm{P}_{\tilde{X}^n|\tilde{U}=m}$ is $\mathcal{C}_m$ and (32). This bound along with (36) yields

$$\log N \leq nC_\alpha(W) + 2\log\frac{1}{1-\varepsilon-\delta} + (2\alpha+2)\log\frac{1}{1-\sqrt{1-(1-\varepsilon-\delta)/4}} + 3,$$

which yields the strong converse by (31). □

Remark 4. Unlike the standard choice of auxiliary random variable in the wiretap channel, the random variables $(\tilde{U}, \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n)$ in the above proof do not satisfy the Markov relation $\tilde{U} -\!\!\circ\!\!- \tilde{X}^n -\!\!\circ\!\!- (\tilde{Y}^n, \tilde{Z}^n)$. Instead, we have added the cost $D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}})$.

Proof of Theorem 12: For any distribution $\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}$, note first that (see [12, Lemma 17.12])

$$I(\tilde{U}\wedge\tilde{Y}^n) - I(\tilde{U}\wedge\tilde{Z}^n) = n[I(\tilde{U}\wedge\tilde{Y}_J|V_J, J) - I(\tilde{U}\wedge\tilde{Z}_J|V_J, J)], \tag{37}$$

where J is distributed uniformly over $\{1, ..., n\}$ and $V_j = (\tilde{Y}_j^-, \tilde{Z}_j^+)$. Next, consider

$$D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) = D(\mathrm{P}_{\tilde{Y}^n|\tilde{X}^n\tilde{U}}\|W_1^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) + D(\mathrm{P}_{\tilde{Z}^n|\tilde{Y}^n\tilde{X}^n\tilde{U}}\|W_2^n|\mathrm{P}_{\tilde{Y}^n\tilde{X}^n\tilde{U}}). \tag{38}$$

The first term on the right is bounded below by

$$\begin{aligned}
&D(\mathrm{P}_{\tilde{Y}^n|\tilde{X}^n\tilde{U}}\|W_1^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) - D(\mathrm{P}_{\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W_2^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) \\
&= \mathbb{E}\left[\log\frac{W_2^n(\tilde{Z}^n|\tilde{X}^n)}{W_1^n(\tilde{Y}^n|\tilde{X}^n)}\right] + H(\tilde{Z}^n|\tilde{X}^n, \tilde{U}) - H(\tilde{Y}^n|\tilde{X}^n, \tilde{U}) \\
&= \sum_{j=1}^n \left[\mathbb{E}\left[\log\frac{W_2(\tilde{Z}_j|\tilde{X}_j)}{W_1(\tilde{Y}_j|\tilde{X}_j)}\right] + H(\tilde{Z}_j|\tilde{X}^n, \tilde{U}, V_j) - H(\tilde{Y}_j|\tilde{X}^n, \tilde{U}, V_j)\right] \\
&= \sum_{j=1}^n \left[D(\mathrm{P}_{\tilde{Y}_j|\tilde{X}^n\tilde{U}V_j}\|W_1|\mathrm{P}_{\tilde{X}^n\tilde{U}V_j}) - D(\mathrm{P}_{\tilde{Z}_j|\tilde{X}^n\tilde{U}V_j}\|W_2|\mathrm{P}_{\tilde{X}^n\tilde{U}V_j})\right],
\end{aligned}$$

where the second equality follows from (4). For the second term on the right-side of (38), we have

$$\begin{aligned}
D(\mathrm{P}_{\tilde{Z}^n|\tilde{Y}^n\tilde{X}^n\tilde{U}}\|W_2^n|\mathrm{P}_{\tilde{Y}^n\tilde{X}^n\tilde{U}}) &= \sum_{j=1}^n D(\mathrm{P}_{\tilde{Z}_j|\tilde{Z}_j^+\tilde{Y}^n\tilde{X}^n\tilde{U}}\|W_2|\mathrm{P}_{\tilde{Z}_j^+\tilde{Y}^n\tilde{X}^n\tilde{U}}) \\
&\geq \sum_{j=1}^n D(\mathrm{P}_{\tilde{Z}_j|\tilde{X}^n\tilde{U}V_j}\|W_2|\mathrm{P}_{\tilde{X}^n\tilde{U}V_j}),
\end{aligned}$$

where the inequality uses the convexity of $D(\mathrm{P}\|\mathrm{Q})$ in (P, Q). Using these bounds with (38), it follows that

$$D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) \geq \sum_{j=1}^n D(\mathrm{P}_{\tilde{Y}_j|\tilde{X}^n\tilde{U}V_j}\|W_1|\mathrm{P}_{\tilde{X}^n\tilde{U}V_j})$$

$$\begin{aligned} &\geq \sum_{j=1}^{n} D(\mathrm{P}_{\tilde{Y}_j|\tilde{X}_j\tilde{U}V_j}\|W_1|\mathrm{P}_{\tilde{X}_j\tilde{U}V_j}) \\ &= nD(\mathrm{P}_{\tilde{Y}_J|\tilde{X}_J\tilde{U}V_JJ}\|W_1|\mathrm{P}_{\tilde{X}_J\tilde{U}V_JJ}). \end{aligned} \tag{39}$$

Also,

$$\begin{aligned} D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) &\geq D(\mathrm{P}_{\tilde{Z}^n|\tilde{Y}^n\tilde{X}^n\tilde{U}}\|W_2^n|\mathrm{P}_{\tilde{Y}^n\tilde{X}^n\tilde{U}}) \\ &= \sum_{j=1}^{n} D(\mathrm{P}_{\tilde{Z}_j|\tilde{Z}_j^+\tilde{Y}^n\tilde{X}^n\tilde{U}}\|W_2|\mathrm{P}_{\tilde{Z}_j^+\tilde{Y}^n\tilde{X}^n\tilde{U}}) \\ &\geq \sum_{j=1}^{n} D(\mathrm{P}_{\tilde{Z}_j|\tilde{Y}_j\tilde{X}_j\tilde{U}V_j}\|W_2|\mathrm{P}_{\tilde{Y}_j\tilde{X}_j\tilde{U}V_j}) \\ &= nD(\mathrm{P}_{\tilde{Z}_J|\tilde{Y}_J\tilde{X}_J\tilde{U}V_JJ}\|W_2|\mathrm{P}_{\tilde{Y}_J\tilde{X}_J\tilde{U}V_JJ}). \end{aligned} \tag{40}$$

The bounds (39) and (40) yield

$$2D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) \geq nD(\mathrm{P}_{\tilde{Y}_J\tilde{Z}_J|\tilde{X}_J\tilde{U}V_JJ}\|W|\mathrm{P}_{\tilde{X}_J\tilde{U}V_JJ}).$$

Consequently, we have

$$\begin{aligned} &I(\tilde{U}\wedge\tilde{Y}^n) - I(\tilde{U}\wedge\tilde{Z}^n) - 2\alpha D(\mathrm{P}_{\tilde{Y}^n\tilde{Z}^n|\tilde{X}^n\tilde{U}}\|W^n|\mathrm{P}_{\tilde{X}^n\tilde{U}}) \\ &\leq n\big[I(\tilde{U}\wedge\tilde{Y}_J|V_J,J) - I(\tilde{U}\wedge\tilde{Z}_J|V_J,J) - \alpha D(\mathrm{P}_{\tilde{Y}_J\tilde{Z}_J|\tilde{X}_J\tilde{U}V_JJ}\|W|\mathrm{P}_{\tilde{X}_J\tilde{U}V_JJ})\big] \\ &\leq nC_{\mathtt{s}}^{\alpha}(W), \end{aligned}$$

where, in the last inequality, we removed (V_J, J) by taking the maximum over realizations of (V_J, J). Since $\mathrm{P}_{\tilde{U}\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}$ is arbitrary, the proof is completed. □

VII. Discussion

Our proofs of strong converse have followed a common recipe where an important step is to establish the super-additivity and sub-additivity, respectively, of the lower and upper bounds involving the changed measure. To facilitate this, we have used appropriately crafted variational formulae for these bounds which allowed us to establish the desired additivity properties. These results, Theorem 4, Theorem 6, Theorem 8, Theorem 10, and Theorem 12, along with Proposition 1 may be of independent interest.

We restricted our treatment to the case of random variables taking finitely many values. But this assumption was used only to establish the variational formulae (7), (14), (20), and (31), and our results will hold whenever these formulae can be established. In particular, a technical difficulty in generalizing these formulae is to replace the use of uniform continuity of the information quantities in our proofs with suitable conditions. We only need this in the neighborhood of product distributions; we have not pursued generalization in this direction in the current paper, but techniques we develop can be applied to more general distributions. Regarding continuous channels, the strong converse theorems were proved in [14], [15], [16] for the Gaussian multiple access channel, the Gaussian broadcast channel, and some class of Gaussian networks.

An intriguing direction of research is if the change of measure argument can be used to derive second-order converse bounds, namely extending the results of [25], [38] to the multiterminal setting. For centralized coding problems, such as the Gray-Wyner network, an application of the argument in [19] (namely, the change of measure argument without introducing penalty terms) to each type class leads to the exact second-order converse bound (cf. [46], [52]). A difficulty for distributed coding problems is the evaluation of the variational formulae; to derive second-order bounds, we need to take the limit of block length n and the multiplier α simultaneously. Recently, following-up on an early version of this paper, an evaluation method for the variational formulae was developed in [36] which used the bound in (10) to derive a second-order converse bound for the Wyner-Ziv problem.

The strong converse claim considered in this paper is that the capacity remains unchanged even if a constant error $0 < \varepsilon < 1$ is allowed. A stronger notion of strong converse, termed the exponential strong converse or Arimoto converse, requires that the error converges to 1 exponentially rapidly when the rate exceeds the capacity [6]. In fact, our proofs give exponential strong converses as well. For instance, in the lossy source coding with side-information, by setting $\varepsilon = 1 - 2^{-\xi n}$ in the final bound (10) of the proof of Corollary 5, we can show that any code with excess distortion probability less than $1 - 2^{-\xi n}$ must satisfy

$$R + \mu D \geq R_{\mathtt{WZ}}^{\mu,\alpha}(\mathrm{P}_{XY}) - (\alpha+1)\xi. \tag{41}$$

Suppose that $R + \mu D \leq R_{\mathtt{WZ}}^{\mu}(\mathrm{P}_{XY}) - 2\nu$ for some $\nu > 0$. The variational formula (7) implies that there exists sufficiently large α such that $R_{\mathtt{WZ}}^{\mu}(\mathrm{P}_{XY}) \leq R_{\mathtt{WZ}}^{\mu,\alpha}(\mathrm{P}_{XY}) + \nu$. Thus, if we take ξ so that $\xi < \frac{\nu}{\alpha+1}$, then (41) is violated, which implies that the excess distortion probability must be larger than $1 - 2^{-\xi n}$. However, the above argument does not give an explicit lower bound for the exponent of the convergence speed. Such an explicit bound has been derived recently by Oohama for certain multiterminal problems (cf. [35], [34]).

Another interesting problem, which we have not considered, is that of the multiple access channel (MAC). The strong converse for MAC was established in [13], [1] using a technical tool called the "wringing lemma." While we can recast the proof of [13], [1] in our change of measure language, but it does not offer any extra insight. In particular, we cannot circumvent the wringing lemma and simplify the proof of [13], [1]; indeed, it is of interest to simplify this opaque and technical proof.

Finally, it is of interest to examine the applicability of our strong converse proof recipe to problems studied in other fields. One such instance was recently demonstrated in [44] where this recipe was used to provide an alternative proof for the multiprover nonsignaling parallel repetition theorem, an important result in theoretical computer science and physics.

Even though we have illustrated the utility of our recipe only for several representative problems, we believe that this recipe provides strong converse theorems for any problems as long as single-letter characterizations of the optimal rates under weak converse are known. An interesting future direction will be an application of this change-of-measure argument to problems such that single-letter characterizations of weak converse are unknown. A partial attempt for this problem was made in [20] for centralized coding problems.[16] A research in such a direction will establish a folklore in information theory: Strong converse holds for any stationary memoryless system.

Appendix

A. *Proof of variational formaula* (7)

The proof is almost the same as the proof of Proposition 2. Clearly, the left-side is greater than or equal to the right-side. To prove the other direction, for each $\alpha > 0$, let $\mathrm{P}^{\alpha}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}}$ be the minimizer for the inner minimum on the right-side, and let $\mathrm{Q}^{\alpha}_{\tilde{U}XY\tilde{Z}} = \mathrm{P}^{\alpha}_{\tilde{Z}|\tilde{U}\tilde{Y}}\mathrm{P}^{\alpha}_{\tilde{U}|\tilde{X}}\mathrm{P}_{XY}$ be the induced distribution. Since $G(\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}}) = I(\tilde{U} \wedge \tilde{X}|\tilde{Y}) + \mathbb{E}[d(\tilde{X}, \tilde{Z})]$ is nonnegative and bounded above by $a = \log|\mathcal{X}| + D_{\max}$, it must hold that $D(\mathrm{P}^{\alpha}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}} \| \mathrm{Q}^{\alpha}_{\tilde{U}XY\tilde{Z}}) \le (a/\alpha)$.

Furthermore, since $G(\mathrm{P}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}})$ is uniformly continuous, there exists a function $\Delta(t)$ satisfying $\Delta(t) \to 0$ as $t \to 0$ such that

$$\begin{aligned} R^{\mu,\alpha}_{\mathtt{WZ}}(\mathrm{P}_{XY}) &\ge G(\mathrm{P}^{\alpha}_{\tilde{U}\tilde{X}\tilde{Y}\tilde{Z}}) \\ &\ge G(\mathrm{Q}^{\alpha}_{\tilde{U}XY\tilde{Z}}) - \Delta(a/\alpha) \\ &\ge R^{\mu}_{\mathtt{WZ}}(\mathrm{P}_{XY}) - \Delta(a/\alpha). \end{aligned}$$

Thus, we obtain the desired inequality by taking $\alpha \to \infty$, which completes the proof. □

[16] Instances of strong converses for problems with unknown single-letter characterization of the optimal rate are available; see [2], [28]. Both these results apply the blowing-up lemma in a non-trivial manner.

B. Proof of variational formula (14)

The proof mimics the one above, but has been included for completeness. As before, it is easy to see that the left-side is greater than or equal to the right-side. For the other direction, for each $\alpha > 0$, let $\mathrm{P}^{\alpha}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}$ be the minimizer for the inner minimum on the right-side, and let $\mathrm{Q}^{\alpha}_{\tilde{U}\tilde{V}XY} = \mathrm{P}^{\alpha}_{\tilde{V}|\tilde{U}\tilde{Y}}\mathrm{P}^{\alpha}_{\tilde{U}|\tilde{X}}\mathrm{P}_{XY}$ be the induced distribution. Since $G(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}) = I(\tilde{U},\tilde{V} \wedge \tilde{X}|\tilde{Y}) + I(\tilde{U},\tilde{V} \wedge \tilde{Y}|\tilde{X})$ is nonnegative and bounded above by $a = \log|\mathcal{X}||\mathcal{Y}|$, it must hold that

$$D(\mathrm{P}^{\alpha}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}} \| \mathrm{Q}^{\alpha}_{\tilde{U}\tilde{V}XY}) \le \frac{a}{\alpha} \tag{42}$$

and[17]

$$H_{\mathrm{P}^{\alpha}}(\tilde{F}|\tilde{Y},\tilde{U}) \le \frac{a}{\alpha}, \qquad H_{\mathrm{P}^{\alpha}}(\tilde{F}|\tilde{X},\tilde{U},\tilde{V}) \le \frac{a}{\alpha}. \tag{43}$$

Using the compactness of the finite dimensional probability simplex, there exists a subsequence $\{\mathrm{Q}^{\alpha_i}_{\tilde{U}\tilde{V}XY}\}_{i=1}^{\infty}$ of $\{\mathrm{Q}^{\alpha}_{\tilde{U}\tilde{V}XY}\}_{\alpha\in\mathbb{N}}$ that converges to $\mathrm{Q}^{*}_{\tilde{U}\tilde{V}XY}$. By uniform continuity of the entropy, (42) and (43) imply that the limit point $\mathrm{Q}^{*}_{\tilde{U}\tilde{V}XY}$ satisfies $H_{\mathrm{Q}^*}(F|Y,\tilde{U}) = H_{\mathrm{Q}^*}(F|X,\tilde{U},\tilde{V}) = 0$. Furthermore, since $G(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}})$ is also uniform continuous, there exists a function $\Delta(t)$ satisfying $\Delta(t) \to 0$ as $t \to 0$ such that

$$\begin{aligned} R_f^{\alpha_i}(\mathrm{P}_{XY}) &\ge G(\mathrm{P}^{\alpha_i}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}) \\ &\ge G(\mathrm{Q}^{\alpha_i}_{\tilde{U}\tilde{V}XY}) - \Delta(a/\alpha_i). \end{aligned}$$

Thus, by taking the limit $i \to \infty$, we have

$$\begin{aligned} \sup_{\alpha>0} R_f^{\alpha}(\mathrm{P}_{XY}) &\ge G(\mathrm{Q}^{*}_{\tilde{U}\tilde{V}XY}) \\ &\ge R_f(\mathrm{P}_{XY}), \end{aligned}$$

which completes the proof. □

C. Proof of variational formula (20)

The proof is a minor variant of those of (7) and (14). We only need to observe that the function $G(\mathrm{P}_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}) = I(\tilde{U},\tilde{V}\wedge\tilde{X},\tilde{Y}) - \mu\big(I(\tilde{U},\tilde{V}\wedge\tilde{X}|\tilde{Y}) + I(\tilde{U},\tilde{V}\wedge\tilde{Y}|\tilde{X})\big)$ is bounded above by $a = \log|\mathcal{X}||\mathcal{Y}|$ and below by $b = -\mu\log|\mathcal{X}||\mathcal{Y}|$. With this observation, the same arguments go through. □

[17] We have put the subscripts in (43) to emphasize the underlying measure.

REFERENCES

[1] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," *J. Combinatorics, Information and System Sciences*, vol. 7, no. 3, pp. 216–230, 1982.

[2] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, July 1986.

[3] ——, "Common randomness in information theory and cryptography–part ii: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, January 1998.

[4] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 34, pp. 157–177, 1976.

[5] R. Ahlswede and J. Körner, "Source coding with side information and a converse for the degraded broadcast channel," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, November 1975.

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

[7] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, December 2013.

[8] M. Braverman and A. Rao, "Information equals amortized communication," *Annual Symposium on Foundations of Computer Science*, pp. 748–757, 2011.

[9] M. Braverman and O. Weinstein, "An interactive information odometer with applications," *STOC*, pp. 341–350, 2015.

[10] I. Csiszár and J. Körner, "Broadcast channel with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[11] ——, "Feedback does not affect the reliability function of a DMC at rates above capacity," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 92–93, January 1982.

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

[13] G. Dueck, "The strong converse of the coding theorem for the multiple-access channel," *Journal of Combinatorics, Information & System Sciences*, vol. 6, no. 3, pp. 187–196, 1981.

[14] S. L. Fong and V. Y. F. Tan, "A proof of the strong converse theorem for Gaussian multiple access channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4376–4394, August 2016.

[15] ——, "A proof of the strong converse theorem for Gaussian broadcast channels via the Gaussian Poincaré inequality," *IEEE Trans. Inf. Theory*, vol. 63, no. 12, pp. 7737–7746, Dec. 2017.

[16] ——, "Strong converse theorems for multimessage networks with tight cut-set bound," *Problems of Information Transmission*, vol. 55, no. 1, pp. 67–100, April 2019.

[17] E. Graves and T. Wong, "Wiretap channel capacity: Secrecy criteria, strong converse, and phase change," *Proc. IEEE International Symposium on Information Theory*, 2017, arXiv:1701.07347.

[18] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Labs. Technical Journal*, vol. 53, no. 9, pp. 1681–1721, November 1974.

[19] W. Gu and M. Effros, "A strong converse for a collection of network source coding problem," *Proc. IEEE International Symposium on Information Theory*, pp. 2316–2320, 2009.

[20] ——, "A strong converse in source coding for super-source networks," *Proc. IEEE International Symposium on Information Theory*, pp. 395–399, 2011.

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

[22] E. A. Haroutunian, "Bounds on the error probability exponent for semicontinuous memoryless channels," *Prob. Pered. Inform.*, vol. 4, no. 4, pp. 37–48, 1968, in Russian.

[23] M. Hayashi, H. Tyagi, and S. Watanabe, "Strong converse for a degraded wiretap channel via active hypothesis testing," *Proc. Conference on Communication, Control, and Computing (Allerton)*, 2014.

[24] ——, "Secret key agreement: General capacity and second-order asymptotics," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3796–3810, July 2016.

[25] 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.

[26] S. T. Jose and A. A. Kulkarni, "Linear programming-based converse for finite blocklength lossy joint source-channel coding," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7066–7094, November 2017.

[27] ——, "Improved finite blocklength converses for Slepian-Wolf coding via linear programming," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2423–2441, April 2019.

[28] O. Kosut and J. Kliewer, "Strong converses are just edge removal properties," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3315–3339, June 2019.

[29] J. Liu, "Dispersion bound for the Wyner-Ahlswede-Körner network via reverse hypercontractivity on types," *Proc. IEEE International Symposium on Information Theory*, pp. 1854–1858, 2018.

[30] J. Liu, R. Handel, and S. Verdú, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," *Proc. IEEE International Symposium on Information Theory*, pp. 943–947, 2017.

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

[32] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, no. 3, pp. 445–446, May 1986.

[33] P. Narayan and H. Tyagi, *Multiterminal Secrecy by Public Discussion.* Hanover, MA, USA: Now Publishers Inc., 2016.

[34] Y. Oohama, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, p. 352, May 2018.

[35] ——, "Exponential strong converse for one helper source coding problem," *Entropy*, vol. 21, no. 6, p. 567, June 2019.

[36] ——, "An inequality useful for proofs of strong converse theorems in network information theory," in *IEEE International Symposium on Information Theory*, 2019, pp. 2883–2887.

[37] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, March 2001.

[38] 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.

[39] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with nonvanishing error probability," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 5–21, Jan. 2014.

[40] M. Sudan, H. Tyagi, and S. Watanabe, "Communication for generating correlation: A survey," *arXiv:1904.09563*, 2019.

[41] V. Y. F. Tan and M. R. Bloch, "Information spectrum approach to strong converse theorems for degraded wiretap channels," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1891–1904, 2015.

[42] H. Tyagi, "Common information and secret key capacity," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5627–5640, Sep. 2013.

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

[44] ——, “A new proof of nonsignaling multiprover parallel repetition theorem,” *Proc. IEEE International Symposium on Information Theory*, pp. 967–971, 2019.

[45] S. Watanabe, “A converse bound on Wyner-Ahlswede-Körner network via Gray-Wyner network,” in *Proceedings of the 2017 IEEE Information Theory Workshop*, Nov. 2017.

[46] ——, “Second-order region for Gray-Wyner network,” *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 1006–1018, February 2017.

[47] Y.-P. Wei and S. Ulukus, “Partial strong converse for the non-degraded wiretap channel,” in *Proceedings of the 54th Annual Allerton Conference on Communications, Control and Computing*, 2016, arXiv:1610.04215.

[48] J. Wolfowitz, *Coding theorems of information theory*. New York:Springer-Verlag, 1961.

[49] A. D. Wyner, “On source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, May 1975.

[50] ——, “The wiretap channel,” *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1367, October 1975.

[51] A. D. Wyner and J. Ziv, “The rate distortion function for source coding with side information,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[52] L. Zhou, V. Y. F. Tan, and M. Motani, “Discrete lossy Gray-Wyner revisited: Second-order asymptotics, large and moderate deviations,” *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1766–1791, March 2017.