Simulation of a Channel With Another Channel

Farzin Haddadpour, Mohammad Hossein Yassaee, Salman Beigi, Amin Gohari, and Mohammad Reza Aref

***Abstract*— In this paper, we study the problem of simulating a discrete memoryless channel (DMC) from another DMC under an average-case and an exact model. We present several achievability and infeasibility results, with tight characterizations in special cases. In particular, for the exact model, we fully characterize when a binary symmetric channel can be simulated from a binary erasure channel when there is no shared randomness. We also provide infeasibility and achievability results for the simulation of a binary channel from another binary channel in the case of no shared randomness. To do this, we use the properties of Rényi capacity of a given order. We also introduce a notion of "channel diameter" which is shown to be additive and satisfy a data processing inequality.**

***Index Terms*— Channel simulation, coordination, point to point channel, broadcast, BIBO, OSRB.**

I. Introduction

Characterizing when a stochastic resource, such as a channel, can simulate another stochastic resource is of theoretical and practical interest [26], [27]. In particular, this general problem relates to the question of how much randomness one can one distill from an imperfect stochastic resource, or how much randomness is required to synthesis a given stochastic resource (e.g. see [20], [21]). In this work, we are primarily interested in channels as stochastic resources, and whether one can use a channel to simulate another channel. As another example, assume that we have designed an error-correction code for an intended channel. But, it turns out that the actual channel differs from the intended channel. Then, one may wish to augment the wrong channel so that the same error-correction code can be utilized for even the wrong channel.

As a concrete example, consider a scenario in which memoryless copies of a BEC with erasure probability ϵ from Alice to Bob is available. Alice and Bob aim to use this resource to simulate memoryless copies of a BSC channel with crossover probability p. We require that the number of consumed BECs to be equal to the number of generated BSCs. This is clearly possible when $p \in [\frac{\epsilon}{2}, \frac{1}{2}]$, since Bob can degrade the output of the BEC channel by mapping the erasure symbol to 0 or 1 with equal probabilities (see Fig. 1). On the other hand, channel simulation is impossible when $p \in \left[0, h^{-1}(\epsilon)\right)$, since the capacity of the consumed erasure channel should be greater than or equal to the capacity of the simulated BSC channel. Then, the question is whether channel simulation is possible when $p \in \left[h^{-1}(\epsilon), \frac{\epsilon}{2}\right)$. Moreover, what would be the answer, if Alice and Bob are additionally provided with shared randomness at a limited rate? The answer to this question depends whether we want to *approximately or exactly* simulate a given channel (the notion of approximate simulation is made precise later). For instance, in the presence of infinite shared random randomness, any two channels of equal capacity can approximately simulate one another [1], [2]. As a result, in the presence of infinite randomness, approximate BEC to BSC simulation is possible if and only if $p \in \left[h^{-1}(\epsilon), \frac{1}{2}\right]$. On the other hand, we show in this work that exact simulation is possible if and only if $p \in \left[\frac{\epsilon}{2}, \frac{1}{2}\right]$ in presence of no shared randomness. Thus, the answers to the above questions depend on the amount of shared randomness and the notion of channel simulation we are considering.

The channel simulation problem can be described as follows: Alice has an x^n sequence, and n memoryless copies of the channel $p(\tilde{y}|\tilde{x})$ as a resource. She creates $\tilde{x}^n$ as a (stochastic) function of x^n, and sends it to Bob over n copies of the channel $p(\tilde{y}|\tilde{x})$. Bob receives $\tilde{y}^n$ and passes it through

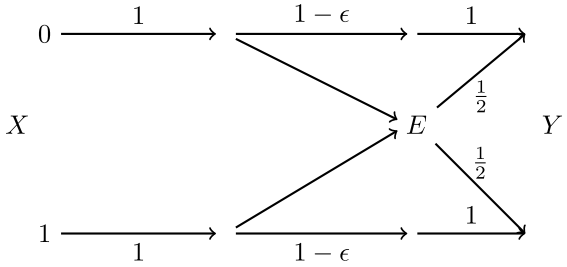

Fig. 1. Simulation of a BSC($\epsilon/2$) from a BEC(ϵ) channel.

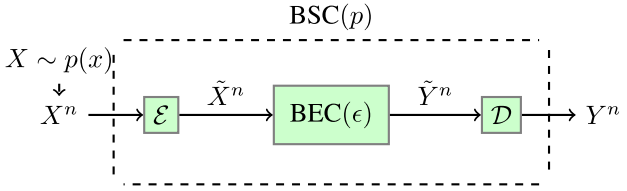

Fig. 2. Channel simulation in absence of common randomness.

Manuscript received October 14, 2015; revised September 29, 2016; accepted November 26, 2016. Date of publication December 5, 2016; date of current version April 19, 2017. This work was supported in part by the Institute for Research in Fundamental Sciences under Grant 94050034, and in part by the Iran National Science Foundation under Grant 95824657 and Grant 9232575. This work was presented at the 2013 IEEE Information Theory Workshop.

F. Haddadpour, M. H. Yassaee, A. Gohari, and M. R. Aref are with the Department of Electrical Engineering, Sharif University of Technology, Tehran 11155-4363, Iran (e-mail: farzin.haddadpour@gmail.com; yassaee@gmail.com; aminzadeh@sharif.edu; aref@sharif.edu).

S. Beigi is with the School of Mathematics, Institute for Research in Fundamental Sciences, Tehran 19395-5746, Iran (e-mail: salman.beigi@gmail.com).

Communicated by D. Tuninetti, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2635660

another stochastic map to generate y^n. Their goal is that the induced channel, which maps x^n to y^n, to be exactly equal, or ϵ-close to n copies of a target memoryless channel $p(y|x)$, for some sufficiently small ϵ.

To formally define the simulation problem, one has to specify a model for the input sequences, x^n. Furthermore, one should also specify the simulation error ϵ. Here, we consider two models of *worst-case* and *average-case* for the input sequence, and two models of *asymptotically reliable* ($\epsilon \to 0$) and *exact* ($\epsilon = 0$) for the error. Briefly speaking, in the worst-case model we demand reliable simulation *for all* possible input sequences x^n. In the average-case model, on the other hand, we assume that the input X^n is i.i.d. according to some distribution $p(x)$. We use the total variation distance between the simulated channel and the desired channel to measure the performance of the simulation protocol. In an asymptotically reliable simulation, we want the total variation distance to vanish asymptotically, whereas in the exact simulation we want it to be exactly zero. All in all, three models of channel simulation emerges:

- average-case with asymptotically vanishing distortion,
- worst-case with asymptotically vanishing distortion,
- exact model.

In the exact model, the zero distortion constraint implies exact simulation for all possible input sequences x^n. The average-case model would require approximate channel simulation for most typical sequences x^n. Finally, the worst-case model with asymptotically vanishing distortion requires approximate channel simulation for all sequences x^n.

Another aspect of the channel simulation problem is the existence or lack thereof of shared randomness. Thus, any of the models can be conceived when limited share randomness exists between the transmitter and the receiver. We would like to emphasize that one can also conceive other notions of channel simulation, which we do not consider in this work, *e.g.,* average distortion ($\bar{d}$-distance) measure as in [18]–[20], agreement probability as in [22], the empirical coordination of [23] or the exact simulation of [5]; see also [6], [7].

While the problem of simulating an arbitrary channel from another one has been formally defined in the literature, *e.g.* in [8], but only the following special cases are studied in the literature:

Simulation of a Noiseless Link: Suppose that we would like to simulate a noiseless link from a given channel $p(\tilde{y}|\tilde{x})$. In both the average-case and worst-case asymptotic models, a noiseless link can be simulated with $p(\tilde{y}|\tilde{x})$ only if its rate is less than the Shannon capacity of $p(\tilde{y}|\tilde{x})$. Indeed, channel simulation in this special case, corresponds to the problems of computing the average probability of error capacity, and maximum probability of error capacity of $p(\tilde{y}|\tilde{x})$ which are known to be equal. Similarly, for exact simulation, the rate of the noiseless link should be less than the zero-error capacity of $p(\tilde{y}|\tilde{x})$. Shared randomness does not affect the zero-error and average error capacity of a channel, hence is useless for simulation of a noiseless link.

Simulation With a Noiseless Link: Conversely, suppose that we would like to simulate a channel $p(y|x)$ from a noiseless link as well as limited shared randomness. The reverse Shannon theorem shows the ability of a noiseless link to simulate a noisy channel of equal capacity in the presence of infinite shared randomness for the average-case asymptotic model [1]. In other words, a noiseless link whose rate is equal to the capacity of the channel $p(y|x)$, is sufficient for simulating the channel $p(y|x)$. Therefore, all channels with the same capacity are equivalent in the presence of infinite shared randomness for the average-case asymptotic model. However, the assumption of infinite shared randomness is crucial here. This is demonstrated by the result of Cuff in [2].

In the exact model, we saw above that simulation of a noiseless link from a given channel reduces to the zero-error capacity of a channel, and the amount of shared randomness is not relevant. Unfortunately computing the zero-error capacity of a channel is an open problem. Fortunately, the reverse problem has a clean solution when infinite shared randomness is available, given in [8]. While Cubitt *et al.* [8] do not explicitly express their solution in terms of the Rényi mutual information, one can observe that their answer is the Rényi capacity of order infinity of the channel (see [6] for a discussion).

A. Main Results

In this paper, we only study the two models of exact and asymptotic average-case. We also study the problem of simulating a broadcast channel with another broadcast channel. The main contributions of this paper are as follows.

Asymptotic average-case model: The channel simulation problem in the asymptotic average-case model is essentially a joint source-channel coding problem as we compress and code X^n into the channel inputs $\tilde{X}^n$. From a mathematical perspective, the problem is complicated due to the following fact: consider the joint pmf on random variables $X^n, \tilde{X}^n, \tilde{Y}^n, Y^n$ induced by a simulation code of length n. In the case of no-shared randomness, we have the Markov chain $X^n \to \tilde{X}^n \to \tilde{Y}^n \to Y^n$. However, $X^n, \tilde{X}^n, \tilde{Y}^n, Y^n$ do not necessarily have a joint i.i.d. distribution (even in the average-case model we only require X^n, Y^n to be almost i.i.d.).

For this model, we provide an inner bound for the point-to-point channel simulation problem using the OSRB technique [9]. This inner bound is based on a "hybrid coding" scheme to perform joint source-channel coding. We also provide an outer bound. Then, we compare these bounds for simulation of a BSC channel from a BEC channel in the absence of shared common randomness. We show that in this example, the degradation strategy is sub-optimal for any non-uniform input distribution.

Exact model: Our main result here are two infeasibility results in the absence of shared common randomness. Our first result is based on the observation that if channel $p(y|x)$ can be exactly simulated using $p(\tilde{y}|\tilde{x})$, then the channel $p(\tilde{y}|\tilde{x})$ must have a better *error exponent* than $p(y|x)$ for all communication rates. Error exponents are known to have a characterization in terms of Rényi capacity. Our first infeasibility result is also in terms of Rényi capacity. Our second infeasibility result is in terms a quantity that we introduce and

name the "channel diameter." Just like the Rényi capacity of a channel, this quantity is shown to be additive and satisfy a data processing inequality. We use these two infeasibility results to show that the degradation strategy is optimal for simulation of a BSC channel from a BEC channel in the absence of shared common randomness. We also study when we can simulate a binary input binary output channel (BIBO) from another BIBO channel.

Broadcast channel simulation: We consider an extension of the point-to-point channel simulation problem under the asymptotic average-case model. We apply the OSRB technique to the broadcast channel simulation problem and present an inner bound for it. Simulation of a broadcast network was stated as an open problem in [10, p. 100].

The rest of the paper is organized as follows: In Section II we set up the notation. In Section II-B, two models for channel simulation are introduced. In Sections III-A and III-B we provide our results for a point-to-point channel. Section III-C contains our results for the broadcast channel. In Section IV our result for exact channel simulation is presented. The proofs are included in Section V.

II. Notation and Preliminaries

Throughout this paper, we restrict ourselves to discrete random variables, taking values in finite sets. We generally use $q(\cdot)$ to denote the pmf's induced by a code, and $p(\cdot)$ to denote the desired i.i.d. pmf's. We also use ω as a random variable to denote shared randomness. To distinguish the uniform distribution we use the notation p^{U}. The total variation distance between two pmf's p and q on the same alphabet $\mathcal{X}$, is denoted by $\|p-q\|_1$. We say that $p \overset{\epsilon}{\approx} q$ if $\|p-q\|_1 \leq \epsilon$. A sequence of random variables $X_1, X_2, \cdots, X_j$ is denoted by $X_{[1:j]}$ where $[1:j] = \{1, 2, \cdots, j\}$, and $\bar{\alpha} = 1-\alpha$. We use $H(X)$ to denote the entropy of a random variable X and $h(p)$ to denote the binary entropy of a Bernoulli random variable with parameter p. All the logarithms in this paper are in base 2.

We frequently use the concept of random pmf's, which we denote by capital letters (e.g., P_X). To be more precise, let $\Delta^{\mathcal{X}}$ be the probability simplex over the set $\mathcal{X}$. A random pmf P_X is indeed a probability distribution over $\Delta^{\mathcal{X}}$. In other words, if we use $\mathbf{\Omega}$ to denote some sample space, we may have a random variable $P_X(x;\omega)$ such that for any $\omega \in \mathbf{\Omega}$ and $x \in \mathcal{X}$ we have $P_X(x;\omega) \geq 0$, and $\sum_x P_X(x;\omega) = 1$ for all ω. Then, $P_X(.;\omega)$ is a vector of random variables which resembles a random distribution. We denote such a random pmf by P_X.

We can definite random $P_{X,Y}$ on a product set $\mathcal{X} \times \mathcal{Y}$ similarly. We note that we can continue to use the law of total probability with random pmf's (e.g., to write $P_X(x) = \sum_y P_{X,Y}(x,y)$ meaning that $P_X(x;\omega) = \sum_y P_{X,Y}(x,y;\omega)$ for all ω) and conditional probability pmf's (e.g., to write $P_{Y|X}(y|x) = \frac{P_{XY}(x,y)}{P_X(x)}$ meaning that $P_{Y|X;\omega}(y|x) = \frac{P_{XY}(x,y;\omega)}{P_X(x;\omega)}$ for all ω).

A. Rényi Divergence

Definition 1: Rényi divergence of order α, or the α-Rényi divergence, for $\alpha \in (0,1) \cup (1,\infty)$, between two pmf's is defined as

$$D_\alpha(p\|q) := \frac{1}{\alpha-1}\log\big(\sum_x p_x^\alpha q_x^{1-\alpha}\big).$$

We also define

$$H_\alpha(p) := -D_\alpha(p\|I) = \frac{1}{1-\alpha}\log\big(\sum_x p_x^\alpha\big).$$

Note that $\lim_{\alpha\to 1} D_\alpha(p\|q) = D(p\|q)$ as well as $\lim_{\alpha\to 1} H_\alpha(p) = H(p)$ which are KL divergence and Shannon entropy, respectively.

It is known that Rényi divergence satisfies the data processing inequality similar to KL divergence (see e.g., [17]).

There are several definitions for mutual information of order α. Here we use Sibson's choice [11] as follows:

$$\begin{aligned} I_\alpha(X;Y) &:= \min_{q_Y} D_\alpha(p_{XY}\|p_X\cdot q_Y) \\ &= \frac{\alpha}{\alpha-1}\log\left(\sum_y\left[\sum_x p(x)p(y|x)^\alpha\right]^{1/\alpha}\right). \end{aligned}$$

The α-capacity of a channel $p(y|x)$ is then defined as (see [12] for a review):

$$\begin{aligned} C_\alpha\big(p(y|x)\big) &= \max_{p(x)} I_\alpha(X;Y) \\ &= \max_{p(x)} \frac{\alpha}{\alpha-1}\log\left(\sum_y\left[\sum_x p(x)p(y|x)^\alpha\right]^{1/\alpha}\right). \end{aligned} \tag{1}$$

In particular, the capacity of order infinity (∞-capacity) is equal to

$$C_\infty(p(y|x)) = \log\left(\sum_y \max_x\ p(y|x)\right). \tag{2}$$

For $\alpha = 1$, we let $C_1(p(y|x)) = C(p(y|x))$ to be the Shannon's capacity of the channel.

Interestingly, just like the Shannon capacity, the α-capacity is also additive for product channels.

Theorem 1 [13]: For $\alpha > 0$ and product of identical channels $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$ we have

$$C_\alpha\big(p(y^n|x^n)\big) = nC_\alpha\big(p(y|x)\big).$$

Using the data processing property for Rényi divergence, one can easily show the data processing property for mutual information of order α:

Theorem 2 [14]: If $X - Y - Z - W$ and $\alpha > 0$ we have

$$I_\alpha(Y;Z) \leq I_\alpha(X;W).$$

Lemma 1: $C_\alpha\big(p(y|x)\big)$ is quasi-convex in $p(y|x)$ for any $\alpha > 0$.

Proof: The mapping $p(y|x) \mapsto \zeta_\alpha(I_\alpha(X;Y))$ for any fixed $p(x)$ is convex, where

$$\zeta_\alpha = \frac{1}{\alpha-1}\exp(t - \frac{t}{\alpha})$$

is a monotonically increasing function [12, Th. 4]. Since the inverse of ζ_α is also a monotonically increasing function,

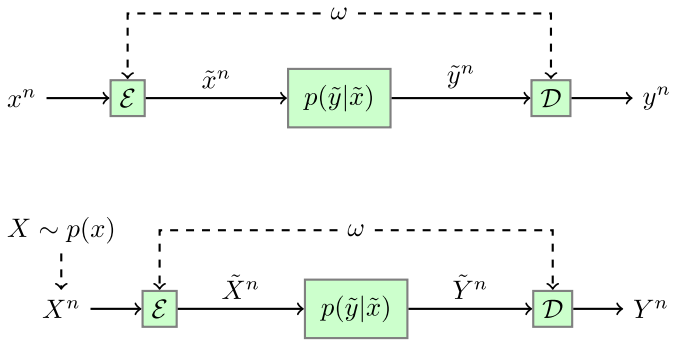

Fig. 3. Simulation of $p(y|x)$ using $p(\tilde{y}|\tilde{x})$ using shared randomness. (Top figure) the general channel simulation problem for a point-to-point channel (Bottom figure) In the average-case model, the input to the channel, i.e., X is assumed to be i.i.d.

and the composition of any convex function with an increasing function is quasi-convex, we conclude that the mapping $p(y|x) \mapsto I_\alpha(X;Y)$ is quasi-convex for any fixed $p(x)$. This implies that the mapping $p(y|x) \mapsto C_\alpha(p(y|x))$ is a maximum of quasi-convex functions, and hence quasi-convex itself. ■

1) Optimal Error Exponent: Rényi mutual information finds an operational meaning in connection to the optimal error exponent of a channel. Given a channel $p(y|x)$ and a fixed transmission rate R, let $P_e^{(n)}$ denote the error probability of the best code with blocklength n. Then, the optimum error exponent $E_{p(y|x)}(R)$, at some fixed transmission rate R, is the supremum of $-\frac{1}{n}\ln P_e^{(n)}$ as n tends to infinity. The optimal error exponent $E(R)$ is also known as the channel reliability function. The channel reliability function is the same under average and maximal notions of error probability; this can be proved using the standard argument of pruning a code with low average probability of error to construct a code with low maximal probability of error. It is known that $E_{p(y|x)}(R)$ is related to capacity of order α:

Theorem 3 [24], [25]: For any rate R less than the channel capacity C, we have that

$$\begin{aligned}&\max_{\alpha\in[\frac{1}{2},1]}(\alpha^{-1}-1)\left(C_\alpha(p(y|x))-R\right)\\&\quad\le E_{p(y|x)}(R)\\&\quad\le \max_{\alpha\in(0,1]}(\alpha^{-1}-1)\left(C_\alpha(p(y|x))-R\right).\end{aligned}$$

B. Two Models for Channel Simulation

Consider the problem of simulating memoryless copies of the channel $p(y|x)$ given memoryless copies of $p(\tilde{y}|\tilde{x})$ as depicted in Fig. 3. Alice's input is denoted by X^n. The shared randomness between Alice and Bob is denoted by the random variable ω, which is independent of X^n and uniformly distributed over $[1:2^{nR}]$. A simulation code consists of

- A (stochastic) encoder with conditional pmf $q^{\text{enc}}(\tilde{x}^n|x^n,\omega)$.
- A (stochastic) decoder with conditional pmf $q^{\text{dec}}(y^n|\tilde{y}^n,\omega)$.

Thus, the joint distribution induced by the code is as follows:

$$\begin{aligned}&q(\tilde{x}^n,\tilde{y}^n,y^n|\omega,x^n)\\&\quad=q^{\text{enc}}(\tilde{x}^n|x^n,\omega)\Big(\prod_{i=1}^n p(\tilde{y}_i|\tilde{x}_i)\Big)q^{\text{dec}}(y^n|\tilde{y}^n,\omega).\end{aligned}\tag{3}$$

An (n,R) *exact* simulation code requires that for every x^n and y^n

$$q(y^n|x^n)=\prod_{i=1}^n p(y_i|x_i),\tag{4}$$

where $q(y^n|x^n)$ is the pmf induced by the code.

On the other hand, an *average-case* simulation code assumes that Alice observes i.i.d. copies of a source X (taking values in the finite set $\mathcal{X}$ and having pmf $p(x)$). Then an (n,R,ϵ) average-case simulation code would require that

$$\Big\|q(x^n,y^n)-\prod_{i=1}^n p(x_i)p(y_i|x_i)\Big\|_1\le\epsilon.$$

Definition 2: The channel $p(y|x)$ is said to be in the admissible region of the channel-rate pair $\big(p(\tilde{y}|\tilde{x}),R\big)$ for the exact simulation model if one can find an (n,R) exact simulation code for some n.

The input distribution-channel pair $\big(p(x),p(y|x)\big)$ is said to be in the admissible region of the channel-rate pair $\big(p(\tilde{y}|\tilde{x}),R\big)$ for the asymptotic average-case model if one can find a sequence of (n,R,ϵ_n) average-case simulation codes such that $\lim_{n\to\infty}\epsilon_n=0$.

Clearly, exact channel simulation implies average case simulation for any arbitrary $p(x)$.

III. Average-Case Model

In this section we state our results on the channel simulation problem in the average-case model. For the proofs of these results refer to Section V.

A. Point-to-Point Channel: Inner Bound

Theorem 4 (Inner Bound): A pair $(p(x),p(y|x))$ is in the admissible region of the channel-rate pair $\big(p(\tilde{y}|\tilde{x}),R\big)$ if one can find channels $p(u,\tilde{x}|x)$ and $p(y|\tilde{y},u)$ such that

$$p(u,\tilde{x},\tilde{y},x,y)=p(x)p(u,\tilde{x}|x)p(\tilde{y}|\tilde{x})p(y|\tilde{y},u),\tag{5}$$

with the given marginal $p(x,y)=p(x)p(y|x)$, and that

$$\begin{aligned}R+I(U;\tilde{Y})&>I(U;XY),\\I(U;\tilde{Y})&>I(U;X).\end{aligned}\tag{6}$$

The main idea of the above theorem is to use an encoding scheme similar to hybrid codes, by utilizing an auxiliary random variable to allow a coupling between the channels, as opposed to simply reducing the resource channel to a noiseless link according to its channel capacity (see [4] for another uses of hybrid coding).

Let us examine the following (non-optimal) general strategy for channel simulation and compare it to the above theorem. We can always convert the channel $p(\tilde{y}|\tilde{x})$ to a noiseless link

of rate $\tilde{C} = \max_{p(\tilde{x})} I(\tilde{X};\tilde{Y})$ and then use the achievability part of the result of [2] (described in the introduction) to simulate the channel $p(y|x)$ from the noiseless link. This strategy is a special case of the above theorem. Let $p(\tilde{x})$ denote the pmf that maximizes $I(\tilde{X};\tilde{Y})$. Also let U in the above theorem to be of the form $U = (\tilde{X}, U')$ where (U', X, Y) is independent of $\tilde{X}$ and satisfies $X - U' - Y$. Then (5) holds and (6) reduces to

$$\begin{aligned} R + \tilde{C} &> I(U';XY), \\ \tilde{C} &> I(U';X). \end{aligned} \tag{7}$$

This region has appeared in [2].

In the special case of $\tilde{Y} = f(\tilde{X})$ where $\tilde{Y}$ is a function of $\tilde{X}$, our inner bound matches (7), which is expected due to the converse given in [2]. To see this, take an arbitrary $p(u, \tilde{x}, \tilde{y}, x, y)$ satisfying (5). In this case we have $X - U\tilde{Y} - Y$, i.e., we have $X - U' - Y$ for $U' = (U, \tilde{Y})$. As a result,

$$\begin{aligned} I(U';XY) &= I(U;XY) + I(\tilde{Y};XY|U) \\ &\le I(U;XY) + H(\tilde{Y}|U) \\ &= I(U;XY) - I(U;\tilde{Y}) + H(\tilde{Y}) \\ &< R + I(U;\tilde{Y}) - I(U;\tilde{Y}) + H(\tilde{Y}). \end{aligned}$$

Thus, $R + H(\tilde{Y}) = R + I(\tilde{X};\tilde{Y}) > I(U';XY)$, and hence $R + \tilde{C} > I(U';XY)$. To verify the second inequality, observe that

$$\begin{aligned} I(U';X) &= I(U;X) + I(\tilde{Y};X|U) \\ &\le I(U;X) + H(\tilde{Y}|U) \\ &= I(U;X) - I(U;\tilde{Y}) + H(\tilde{Y}) \\ &< I(U;\tilde{Y}) - I(U;\tilde{Y}) + H(\tilde{Y}) \\ &= H(\tilde{Y}). \end{aligned}$$

This means that $\tilde{C} \ge I(\tilde{X};\tilde{Y}) = H(\tilde{Y}) > I(U';X)$.

While the cardinality of U in the above theorem can be bounded from above by $|\mathcal{X}||\mathcal{Y}||\tilde{\mathcal{X}}||\tilde{\mathcal{Y}}|$, it is not easy to explicitly compute the inner bound for a given arbitrary pair of channels. In particular, it is not easy to compute the achievable region for the BEC-BSC example that we discussed in the introduction. Nonetheless, we show that the above inner bound gives a strictly better strategy than the degradation scheme, for any non-uniform input pmf in the BEC to BSC simulation problem.

Theorem 5: Degradation for simulating a BSC(p) from a BEC(ϵ) in the average-case model with non-uniform input pmf is suboptimal. In other words, given any $\epsilon \in (0,1)$ and any non-uniform binary input pmf $p(x)$, there exists $p < \epsilon/2$ such that $(p(x), \text{BSC}(p))$ is in the admissible region of $(\text{BEC}(\epsilon), 0)$.

B. Point-to-Point Channel: Outer Bound

Theorem 6 (Outer Bound): If the input distribution-channel pair $(p(x), p(y|x))$ is in the admissible region of the channel-rate pair $\big(p(\tilde{y}|\tilde{x}), R\big)$, then for any non-negative reals β, γ, and θ we have

$$\begin{aligned} I(X;Y) + \min_{U:X-U-Y} \big[\beta I(U;XY) + \gamma I(U;X) + \theta I(U;Y)\big] \\ \le \max_{p(\tilde{x})} \Bigg[I(\tilde{X};\tilde{Y}) + \min_{\tilde{U}:\tilde{X}-\tilde{U}-\tilde{Y}} \big[\beta I(\tilde{U};\tilde{X}\tilde{Y}) + \gamma I(\tilde{U};\tilde{X}) \\ + \theta I(\tilde{U};\tilde{Y})\big] \Bigg] + (\beta + \gamma + 2\theta)R. \end{aligned} \tag{8}$$

Further, to compute the above minimums one can put the cardinality bounds of $|\mathcal{U}| \le |\mathcal{X}| \cdot |\mathcal{Y}|$ and $|\tilde{\mathcal{U}}| \le |\tilde{\mathcal{X}}| \cdot |\tilde{\mathcal{Y}}|$.

According to [1] and [2], when infinite shared randomness is available, $\tilde{C} = C(p(\tilde{y}|\tilde{x})) \ge I(X;Y)$ is a sufficient and necessary condition for channel simulation. By setting $\beta = \gamma = \theta = 0$ in the above theorem, we recover the necessity of this condition.

Corollary 1: $(p(x), p(y|x))$ is in the admissible region of $\big(p(\tilde{y}|\tilde{x}), \infty\big)$ if and only if $\tilde{C} = C(p(\tilde{y}|\tilde{x})) \ge I(X;Y)$.

A channel $p(y|x)$ is called symmetric if for every permutation π_X on $\mathcal{X}$, there is a permutation π_Y on $\mathcal{Y}$ such that $p(\pi_Y(y)|\pi_X(x)) = p(y|x)$. For such channels, the outer bound of the above theorem can be simplified as the maximum on the right hand side occurs at uniform distribution $p^{\mathsf{u}}(\tilde{x})$.

Theorem 7 (Outer Bound for Symmetric Channels): Suppose that the channel $p(\tilde{y}|\tilde{x})$ is symmetric. If $(p(x), p(y|x))$ is in the admissible region of $\big(p(\tilde{y}|\tilde{x}), 0\big)$, then $I_{p^{\mathsf{u}}(\tilde{x})}(\tilde{X};\tilde{Y}) \ge I(X;Y)$ and $\mathcal{S}(p(\tilde{y}|\tilde{x}), p^{\mathsf{u}}(\tilde{x})) \subseteq \mathcal{S}(p(y|x), p(x))$ where

$$\begin{aligned} \mathcal{S}(p(y|x), p(x)) \triangleq \bigcup_{X-U-Y} \big\{(a_1, a_2, a_3) : a_1 \ge I(U;XY), \\ a_2 \ge I(U;X), a_3 \ge I(U;Y)\big\}. \end{aligned}$$

Remark 1: The minimum value of a_1 such that $(a_1, \infty, \infty) \in \mathcal{S}(p(y|x), p(x))$ is Wyner's common information between X and Y, denoted by $C(X,Y)$.

As an application of the above theorem, we show that for any $0 < p < 1$ and $\epsilon < 1/2$, it is impossible to simulate a BEC(ϵ) with uniform input distribution from a BSC(p) when there is no shared randomness. In other words, we show that $(p^{\mathsf{u}}, \text{BEC}(\epsilon))$ is not admissible for $(\text{BSC}(p), 0)$. To prove this, observe that the simple mutual information bound is inadequate as it gives $1 - \epsilon < 1 - h(p)$ which does not exclude the possibility of simulation for the entire range of $0 < p < 1$. Nevertheless, if simulation is possible, from the above outer bound for symmetric channels, we conclude that the Wyner common information of BEC(ϵ) with uniform input distribution must be less than or equal to the maximum of the Wyner common information of BSC(p) over all input distributions. For $\epsilon < 1/2$, the Wyner common information of a BEC(ϵ) with uniform input is one [2, Sec. IV. A]. On the other hand, for any input distribution $p(x)$ the Wyner common information $C(X,Y)$ is strictly less than one, since when $p \in (0,1)$, one can write the BSC(p) channel as the cascade of two *non-trivial* BSC channels $X \to U \to Y$. Then $I(U;XY)$ will be strictly less than $H(U)$ which is less than one. This concludes our claim. Since the average-case model is less restrictive than the exact model, we can conclude that

simulating a BEC(ϵ) from a BSC(p) is impossible under the exact model as well when $0 < p < 1$ and $\epsilon < 1/2$.

In the following theorem we characterize the set $\mathcal{S}(p(y|x), p(x))$ for BSC and BEC channels with uniform input distributions.

Theorem 8: (i) $\mathcal{S}(\text{BEC}(\epsilon), p^{\mathsf{u}})$ is the set of triples (a_1, a_2, a_2) such that for some $a \in [1-\epsilon, 1]$ we have

$$\begin{aligned} a_1 &\geq -\epsilon \log(\epsilon) + a - a\log(a) \\ &\quad +(a-1+\epsilon)\log(a-1+\epsilon), \\ a_2 &\geq a, \\ a_3 &\geq 1-\epsilon-\epsilon\log(\epsilon) - a\log(a) \\ &\quad +(a-1+\epsilon)\log(a-1+\epsilon). \end{aligned}$$

(ii) Assuming Conjecture 1 (stated in Section V), the set $\mathcal{S}(\text{BSC}(p), p^{\mathsf{u}})$ consists of triples (a_1, a_2, a_3) such that for some $0 \leq \alpha, \beta \leq 1$ with $p = \alpha\bar{\beta} + \bar{\alpha}\beta$ we have

$$\begin{aligned} a_1 &\geq 1 + h(p) - h(\alpha) - h(\beta), \\ a_2 &\geq 1 - h(\alpha), \\ a_3 &\geq 1 - h(\beta). \end{aligned}$$

Example 1: Let us use the above theorem as well as Theorem 7 to study the problem of simulating a BSC channel from a BEC channel. More precisely, we want to characterize the range of $p \in [0, 1/2]$ such that $(p^{\mathsf{u}}, \text{BSC}(p))$ is in the admissible region of $(\text{BEC}(\epsilon), 0)$. The degradation scheme (of Fig. 1) for this problem gives the achievability of $p = \epsilon/2$. On the other hand, if simulation is possible, the trivial mutual information bound implies that $1-\epsilon \geq 1-h(p)$, which does not match the achievability bound. Nevertheless, the above results give a stronger converse.

If simulation is possible, then for any $a \geq 1-\epsilon$ there exist α and β such that $p = \alpha\bar{\beta} + \bar{\alpha}\beta$ and

$$\begin{aligned} &-\epsilon\log(\epsilon) + a - a\log(a) + (a-1+\epsilon)\log(a-1+\epsilon) \\ &\quad \geq 1 + h(\alpha\bar{\beta} + \bar{\alpha}\beta) - h(\alpha) - h(\beta), \qquad (9) \\ &a \geq 1 - h(\alpha), \qquad (10) \\ &1 - \epsilon - \epsilon\log(\epsilon) - a\log(a) + (a-1+\epsilon)\log(a-1+\epsilon) \\ &\quad \geq 1 - h(\beta). \qquad (11) \end{aligned}$$

In other words, simulation is possible only if $p \geq p^*$ with

$$p^* = \max_{a \geq 1-\epsilon} \min_{\alpha,\beta} (\alpha\bar{\beta} + \bar{\alpha}\beta),$$

where the minimum is over $\alpha, \beta \in [0, 1]$ satisfying (9)-(11). Observe that without loss of generality we can restrict to $\alpha, \beta \in [0, 1/2]$.

We show that the above converse is stronger than $1-\epsilon \geq 1-h(p)$ for all values of $\epsilon > \epsilon^* \approx 0.7729$, where $\epsilon^* \in (0, 1]$ is the unique positive solution of $\epsilon^* = h(\epsilon^*)$. To prove this, let us choose $a = 1-\epsilon$ and consider the following lower bound on p^*:

$$p^* \geq \min_{\alpha,\beta} (\alpha\bar{\beta} + \bar{\alpha}\beta), \qquad (12)$$

over pairs (α, β) satisfying (9)-(10) with $a = 1-\epsilon$. We have plotted p^* in terms of ϵ in Fig. 4, and compared it with $p = h^{-1}(\epsilon) \in [0, 1/2]$ that one gets from the trivial upper bound. However, to show the improvement analytically, note that equations (10) and (11) imply that

$$h(\alpha) \geq \epsilon, \qquad h(\beta) \geq \epsilon - h(\epsilon).$$

Since $\epsilon > \epsilon^*$, we have $\epsilon - h(\epsilon) > 0$ and hence $h(\beta) > 0$. Thus, $\beta \in [\beta^*, 1/2]$ where $\beta^* > 0$ is the solution of $h(\beta^*) = \epsilon - h(\epsilon)$. Therefore, $\alpha\bar{\beta} + \bar{\alpha}\beta \geq \alpha\bar{\beta}^* + \bar{\alpha}\beta^*$. We next have $h(\alpha) \geq \epsilon$, and thus, $\alpha \in [\alpha^*, 1/2]$ where $\alpha^* \in [0, 1/2]$ is such that $h(\alpha^*) = \epsilon$. Hence, after solving the minimization over all (α, β), one ends up with

$$p^* \geq \min_{\alpha,\beta} (\alpha\bar{\beta} + \bar{\alpha}\beta) = \alpha^*\bar{\beta}^* + \bar{\alpha}^*\beta^*.$$

Since for any $\alpha, \beta \in (0, 1/2]$ we have $h(\alpha\bar{\beta} + \bar{\alpha}\beta) > h(\alpha)$, we find that $h(p^*) > \epsilon$, which results in a strictly better bound than $h(p) \geq \epsilon$ that one gets from the trivial upper bound.

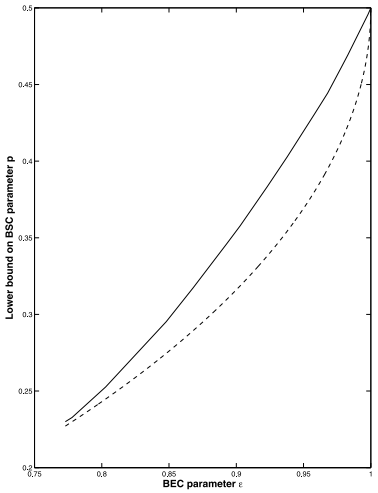

Fig. 4. The value of p^* given in (12) (upper curve), and $p = h^{-1}(\epsilon) \in [0, 1/2]$ (lower curve) in terms of the BEC parameter ϵ, for $\epsilon > \epsilon^* \approx 0.7729$.

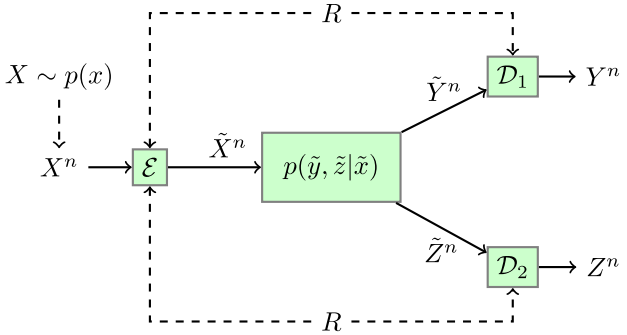

Fig. 5. Channel simulation over the Broadcast channel $p(\tilde{y}, \tilde{z}|\tilde{x})$.

C. Broadcast Channel Simulation

Consider the problem of simulating memoryless copies of the channel $p(y, z|x)$ given memoryless copies of $p(\tilde{y}, \tilde{z}|\tilde{x})$ as depicted in Fig. 5. In this setting, the input terminal observes i.i.d. copies of a source X (taking values in finite sets $\mathcal{X}$ and having joint pmf $p(x)$). The three terminals are provided with a shared randomness at rate R, denoted by random variable

$$I(WU;\tilde{Y}) > I(UW;X), \tag{14}$$
$$R + I(WU;\tilde{Y}) > I(UW;XYZ), \tag{15}$$
$$I(WV;\tilde{Z}) > I(WV;X), \tag{16}$$
$$R + I(WV;\tilde{Z}) > I(WV;XYZ), \tag{17}$$
$$I(WU;\tilde{Y}) + I(WV;\tilde{Z}) > I(UW;X) + I(WV;X) + I(U;V|WX), \tag{18}$$
$$2R + I(UW;\tilde{Y}) + I(VW;\tilde{Z}) > I(UW;XYZ) + I(VW;XYZ) + I(U;V|WXYZ), \tag{19}$$
$$\min\{I(W;\tilde{Y}), I(W;\tilde{Z})\} + I(U;\tilde{Y}|W) + I(V;\tilde{Z}|W) > I(W;X) + I(U;X|W) + I(V;X|W) + I(U;V|WX), \tag{20}$$
$$R + \min\{I(W;\tilde{Y}), I(W;\tilde{Z})\} + I(U;\tilde{Y}|W) + I(V;\tilde{Z}|W) > I(W;XYZ) + I(U;XYZ|W) + I(V;XYZ|W) + I(U;V|WXYZ), \tag{21}$$
$$R + I(W;YZ|X) + I(UW;\tilde{Y}) + I(VW;\tilde{Z}) > I(UW;XYZ) + I(VW;XYZ) + I(U;V|WXYZ). \tag{22}$$

ω, which is uniformly distributed over $[1 : 2^{nR}]$ and is independent of X^n. An (n, R) code consists of

- An (stochastic) encoder with conditional pmf's $q^{\text{enc}}(\tilde{x}^n|x^n, \omega)$,
- Two (stochastic) decoders with conditional pmf's $q^{\text{dec}_1}(y^n|\tilde{y}^n, \omega)$ and $q^{\text{dec}_1}(z^n|\tilde{z}^n, \omega)$.

Thus, the joint distribution induced by the code is as follows:

$$q(\hat{x}^n, \hat{y}^n, \hat{z}^n, y^n, z^n|\omega, x^n) = q^{\text{enc}}(\tilde{x}^n|x^n, \omega) \times \Big(\prod_{i=1}^{n} p(\tilde{y}_i, \tilde{z}_i|\tilde{x}_i)\Big) q^{\text{dec}_1}(y^n|\tilde{y}^n, \omega) q^{\text{dec}_2}(z^n|\tilde{z}^n, \omega). \tag{13}$$

Definition 3: An input distribution-channel pair $\big(p(x), p(y, z|x)\big)$ is said to be in the admissible region of the channel-rate triple $\big(p(\tilde{y}, \tilde{z}|\tilde{x}), R\big)$ if one can find a sequence of (n, R) simulation codes whose induced joint distributions have marginal distributions $q(x^n, y^n, z^n)$ that satisfy

$$\lim_{n\to\infty} \Big\| q(x^n, y^n, z^n) - \prod_{i=1}^{n} p(x_i, y_i, z_i) \Big\|_1 = 0.$$

We can now state our achievability bound for the broadcast channel simulation problem.

Theorem 9 (Inner Bound): $\big(p(x), p(y, z|x)\big)$ is in the admissible region of the channel-rate pair $\big(p(\tilde{y}, \tilde{z}|\tilde{x}), R\big)$ if there exist $p(u, v, w, \tilde{x}|x)$, $p(y|\tilde{y}, u, w)$ and $p(z|\tilde{z}, v, w)$ such that $(X, U, V, W, \tilde{X}, \tilde{Y}, \tilde{Z}, Y, Z)$ is distributed according to

$$p(x, u, v, w, \tilde{x}, \tilde{y}, \tilde{z}, y, z) = p(x)p(u, v, w, \tilde{x}|x)p(\tilde{y}, \tilde{z}|\tilde{x}) \times p(y|\tilde{y}, u, w)p(z|\tilde{z}, v, w),$$

that has the given marginal $p(x, y, z)$, and satisfies equations (14)-(22), as shown at the top of this page.

IV. Exact Channel Simulation

Recall that we say a channel $p(y|x)$ can be simulated exactly with a channel $p(\tilde{y}|\tilde{x})$ if there are n, encoding map $q^{\text{enc}}(\tilde{x}^n|x^n, \omega)$, and decoding map $q^{\text{dec}}(y^n|\tilde{y}^n, \omega)$ such that the induced distribution given by (3) satisfies

$$q(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i). \tag{23}$$

In this section we state our results about the channel simulation problem in the exact model.

Theorem 10: A channel $p(y|x)$ can be simulated from $p(\tilde{y}|\tilde{x})$ in the exact model with infinite shared randomness only if for any $\alpha > 0$ we have

$$C_\alpha(p(y|x)) \le C_\alpha(p(\tilde{y}|\tilde{x})),$$
$$\mathsf{Diam}_\alpha(p(y|x)) \le \mathsf{Diam}_\alpha(p(\tilde{y}|\tilde{x})),$$

where C_α denotes the α-capacity defined in (1), and

$$\mathsf{Diam}_\alpha(p(y|x)) \triangleq \max_{p(x),q(x)} D_\alpha(p(y)\|q(y)).$$

Remark 2: To the best of our knowledge, we define the notion of channel diameter $\mathsf{Diam}_\alpha(p(y|x))$ for the first time. Theorem 10 implies that the exact simulation of $p(y|x)$ with a noiseless link of rate R with infinite shared randomness is possible only if $C_\infty(p(y|x)) \le R$. In fact, as shown in [8] (see [6] for a discussion), $C_\infty(p(y|x)) \le R$ is a sufficient and necessary condition for exact simulation of $p(y|x)$ with a noiseless link of rate R.

Remark 3: As it becomes clear from the proof of Theorem 10, any function $\Phi(p(y|x))$ that satisfies additivity and data processing properties, namely,

$$\Phi\Big(\prod_{i=1}^{n} p(y_i|x_i)\Big) = n\Phi\big(p(y|x)\big)$$

and

$$\Phi(p(y|x)) \ge \Phi(p(z|x)), \quad \text{if} \quad p(z|x) = \sum_y p(y|x)p(z|y)$$

can be used instead of C_α and Diam_α above to write an infeasibility result when there is no shared randomness. Furthermore, if $\Phi(p(y|x))$ is quasi-convex in $p(y|x)$, one can

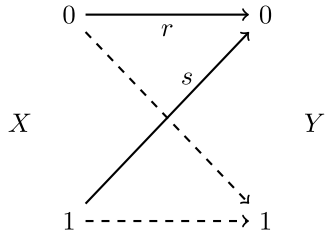

Fig. 6. A BIBO channel $p(y|x)$ with parameters (r, s).

claim the infeasibility result even when there is infinite shared randomness. The above particular choices of $\Phi(p(y|x))$ as C_α and Diam_α are shown later to be particularly useful in finding *tight bounds* for a few examples we consider. But it is possible to find other choices for $\Phi(p(y|x))$: for instance, $E_{p(y|x)}(R)$, the reliability function of $p(y|x)$ at a given rate R satisfies the additivity and data processing inequalities by definition, and is quasi-convex since shared randomness does not increase the optimal error exponent.

Using the above theorem, we show that the degradation strategy of Fig. 1 is optimal in the exact model for simulating a BSC channel from a BEC channel (for any amount of shared randomness).

Theorem 11: For any amount of shared randomness, the BSC(p) channel with parameter $p \in [0, 1/2]$ can be exactly simulated from the BEC(ϵ) if and only if $p \geq \epsilon/2$. In particular, shared randomness is not helpful in this exact simulation problem.

A channel is called binary-input binary-output (BIBO) if both the input and the output of the channel are a single bit. A BIBO channel can be characterized with two parameters $r = p_{Y|X}(0|0)$ and $s = p_{Y|X}(0|1)$ as depicted in Fig. 6. The following theorem studies BIBO channels that can be exactly simulated from another BIBO channel.

Theorem 12: Depending on whether we have shared randomness or not, we have

- (Infinite shared randomness): A channel with parameters (r, s) can be simulated exactly from a channel with parameters (r', s') if and only if (r, s) is in the convex polygon with six vertices $(0, 0)$, $(1, 1)$, (r', s'), (s', r'), $(\bar{s}', \bar{r}')$, and $(\bar{r}', \bar{s}')$ where $\bar{s}' = 1 - s'$ and $\bar{r}' = 1 - r'$ (see Fig. 7).
- (No shared randomness): A BIBO channel with parameters (r, s) can be simulated from another BIBO channel with parameters (r', s') in the exact model if (r, s) is in the union of two parallelogram with vertices $(0, 0)$, (r', s'), $(\bar{r}', \bar{s}')$, $(1, 1)$ and $(0, 0)$, (s', r'), $(\bar{s}', \bar{r}')$, $(1, 1)$ (see Fig. 7). Conversely, if a channel with parameters (r, s) can be simulated exactly from a channel with parameters (r', s'), then (r, s) has to be in the convex polygon with six vertices $(0, 0)$, $(1, 1)$, (r', s'), (s', r'), $(\bar{s}', \bar{r}')$, and $(\bar{r}', \bar{s}')$ depicted in Fig. 7. In other words, the outer bound is the convex hull of the given achievable inner bound.

V. Proofs

This section is devoted to the proofs of the results stated in the previous two sections.

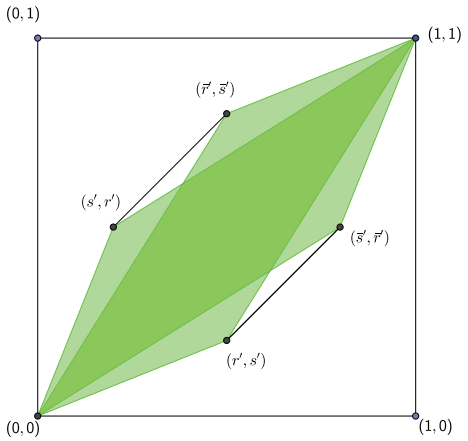

Fig. 7. The inner bound (union of two parallelogram) and outer bound (convex polygon) for exact channel simulation of BIBO channel from another BIBO channel when there is no shared randomness. The convex polygon bound is the simulation region when there is infinite shared randomness.

A. Point to Point Channel: Inner Bound

Proof of Theorem 4: We apply the OSRB technique of [9] to prove the theorem. Our proof consists of three parts. In the first part we introduce two protocols, A and B, each of which induces a pmf on a certain set of random variables. Protocol A has the desired i.i.d. property on X^n and Y^n, but leads to no concrete coding algorithm. However, Protocol B is suitable for construction of a code, with one exception: Protocol B is assisted with an extra common randomness that does not really exist in the model. In the second part of the proof we find conditions on R implying that two certain induced distributions are almost identical. In the third part of the proof, we eliminate the extra common randomness given to Protocol B without significantly disturbing the pmf induced on the desired random variables (X^n, Y^n). This makes Protocol B useful for code construction.

Part (1): We define two protocols each of which induces a joint pmf on random variables of the corresponding protocol.

Protocol A [Not useful for coding]: Let $(U^n, X^n, \tilde{X}^n, \tilde{Y}^n, Y^n)$ be n i.i.d. copies of the joint pmf $p(u, x, \tilde{x}, \tilde{y}, y)$. Consider the following construction:

- To each $u^n \in \mathcal{U}^n$ assign two random bin indices $g \in [1 : 2^{n\tilde{R}}]$ and $\omega \in [1 : 2^{nR}]$.
- Consider a Slepian-Wolf decoder for estimating $\hat{u}^n$ from $(\omega, g, \tilde{y}^n)$. Here we are considering $\tilde{y}^n$ as side information and ω, g as the random bins of the source u^n that we want to decode.

The rate constraints on R, $\tilde{R}$ for the success of this decoder will be imposed later, although this decoder can be conceived even when there is no guarantee of successful decoding. We denoted the random pmf induced by the random binning and the Slepian-Wolf decoder by Q and Q^{SW}, respectively. We then obtain the following joint distribution:

$$\begin{aligned} &Q(x^n, u^n, \hat{u}^n, \tilde{x}^n, \tilde{y}^n, y^n, g, \omega) \\ &\quad = p(x^n)p(u^n, \tilde{x}^n|x^n)p(\tilde{y}^n|\tilde{x}^n) \\ &\qquad \times Q(g, \omega|u^n)Q^{\mathsf{SW}}(\hat{u}^n|g, \omega, \tilde{y}^n)p(y^n|\tilde{y}^n, u^n) \\ &\quad = p(x^n)Q(g, \omega|x^n)Q(u^n, \tilde{x}^n|x^n, g, \omega) \\ &\qquad \times p(\tilde{y}^n|\tilde{x}^n)Q^{\mathsf{SW}}(\hat{u}^n|g, \omega, \tilde{y}^n)p(y^n|\tilde{y}^n, u^n). \end{aligned} \tag{24}$$

Note that we have used capital Q for random pmf's in the above equation.

Protocol B [Useful for coding after removing an extra common randomness]. In this protocol we assume that Alice and Bob have access to the common randomness ω and an extra common randomness G, where G is mutually independent of X^n and ω. We assume that G is distributed uniformly over the set $[1 : 2^{n\tilde{R}}]$. Now we use the following protocol:

- First, Alice having (g, ω, x^n) generates u^n according to the pmf $Q(u^n, \tilde{x}^n|g, \omega, x^n)$ of Protocol A, and sends $\tilde{x}^n$ over the channel to Bob. Then Bob receives $\tilde{y}^n$. Having $(g, \omega, \tilde{y}^n)$, Bob uses the Slepian-Wolf decoder of protocol A to generate $\hat{u}^n$ as a estimation of u^n.
- Having $(\hat{u}^n, \tilde{y}^n)$, Bob generates Y^n according to $p(y^n|\tilde{y}^n, \hat{u}^n) = \prod_{i=1}^n p_{Y|\tilde{Y}U}(y_i|\tilde{y}_i, \hat{u}_i)$.

The random pmf induced by the protocol, denoted by $\hat{Q}$, is equal to

$$\begin{aligned}\hat{Q}(x^n, u^n, \hat{u}^n, \tilde{x}^n, \tilde{y}^n, y^n, g, \omega) &= p^{\mathsf{u}}(\omega)p^{\mathsf{u}}(g)p(x^n)Q(u^n, \tilde{x}^n|g, \omega, x^n)p(\tilde{y}^n|\tilde{x}^n) \\ &\quad \times Q^{\mathsf{sw}}(\hat{u}^n|\omega, g, \tilde{y}^n)p(y^n|\tilde{y}^n, \hat{u}^n), \end{aligned} \tag{25}$$

where p^{u} denotes the uniform distribution.

Part (2): In this part we put sufficient conditions under which the induced distributions Q and $\hat{Q}$ given by (24) and (25) are approximately the same. The first step is to observe that g and ω are the bin indices of u^n. Substituting $T = 2$, $X_1 \leftarrow U$, $X_2 \leftarrow U$, $Z \leftarrow X$, $b_1 \leftarrow g$ and $b_2 \leftarrow \omega$ in [9, Th. 1], implies that if

$$R + \tilde{R} < H(U|X), \tag{26}$$

then there exists $\epsilon_0^{(n)} \to 0$ such that

$$\mathbb{E}\|Q(g, \omega|x^n) - p^{\mathsf{u}}(\omega)p^{\mathsf{u}}(g)\|_1 \le \epsilon_0^{(n)}.$$

Observe that $p^{\mathsf{u}}(\omega)p^{\mathsf{u}}(g) = \hat{Q}(g, \omega)$. This implies that the joint pmf of all random variables, excluding y^n, of the two protocols are close in total variation distance, i.e,

$$\mathbb{E}\|\hat{Q}(x^n, u^n, \hat{u}^n, \tilde{x}^n, \tilde{y}^n, g, \omega) - Q(x^n, u^n, \hat{u}^n, \tilde{x}^n, \tilde{y}^n, g, \omega)\|_1 \le \epsilon_0^{(n)}. \tag{27}$$

To ensure the above equation with y^n included, we begin by investigating conditions that make the Slepian-Wolf decoder of Protocol A succeed with high probability. By the Slepian-Wolf theorem as long as

$$R + \tilde{R} > H(U|\tilde{Y}) \tag{28}$$

holds, we have

$$Q(x^n, u^n, \hat{u}^n, \tilde{x}^n, \tilde{y}^n, g, \omega) \overset{\epsilon_1^{(n)}}{\approx} Q(x^n, u^n, \tilde{x}^n, \tilde{y}^n, g, \omega)\mathbf{1}\{\hat{u}^n = u^n\}. \tag{29}$$

for some vanishing sequence $\epsilon_1^{(n)}$. Then using equations (27) and (29) we can apply [9, Lemma 3] to write

$$\hat{Q}(x^n, u^n, \hat{u}^n, \tilde{x}^n, \tilde{y}^n, g, \omega) \overset{\epsilon_0^{(n)}+\epsilon_1^{(n)}}{\approx} Q(x^n, u^n, \tilde{x}^n, \tilde{y}^n, g, \omega)\mathbf{1}\{\hat{u}^n = u^n\}. \tag{30}$$

Moreover, the third part of [9, Lemma 3] implies that

$$\begin{aligned}\hat{Q}(x^n, u^n, \hat{u}^n, \tilde{x}^n, \tilde{y}^n, g, \omega)p(y^n|\hat{u}^n, \tilde{y}^n) &\overset{\epsilon_0^{(n)}+\epsilon_1^{(n)}}{\approx} Q(x^n, u^n, \tilde{x}^n, \tilde{y}^n, g, \omega)\mathbf{1}\{\hat{u}^n = u^n\}p(y^n|\hat{u}^n, \tilde{y}^n) \\ &= Q(x^n, u^n, \tilde{x}^n, \tilde{y}^n, g, \omega)\mathbf{1}\{\hat{u}^n = u^n\}p(y^n|u^n, \tilde{y}^n) \\ &= Q(x^n, u^n, \tilde{x}^n, \tilde{y}^n, g, \omega)p(y^n|u^n, \tilde{y}^n)\mathbf{1}\{\hat{u}^n = u^n\} \\ &= Q(x^n, u^n, \tilde{x}^n, \tilde{y}^n, y^n, g, \omega)\mathbf{1}\{\hat{u}^n = u^n\}.\end{aligned}$$

Then, by part 1 (second item) of [9, Lemma 3] we have $\hat{Q}(g, x^n, y^n) \overset{\epsilon_0^{(n)}+\epsilon_1^{(n)}}{\approx} Q(g, x^n, y^n)$. Note, in particular, that the marginal pmf on (X^n, Y^n) of Q is equal to $p(x^n, y^n)$ implying that $\hat{Q}(x^n, y^n)$ is within $\epsilon_0^{(n)} + \epsilon_1^{(n)}$ distance of $p(x^n, y^n)$.

To summarize, assuming (26) and (28), and having access to common randomness ω, G, Alice and Bob can simulate the channel $p(y|x)$ using the channel $p(\tilde{y}|\tilde{x})$ according to Protocol B. As discussed above, with high probability $\hat{u}^n$ generated by Bob would be equal to u^n, and the final pmf induced on (X^n, Y^n) would be close to the desired pmf $p(x^n, y^n)$.

Part (3): In the above protocol we assumed that Alice and Bob have access to an extra randomness G which is not present in the model. To eliminate this extra common randomness, we will fix a particular instance g of G and show that the same protocol works even if we fix $G = g$. To prove this note that by letting $G = g$, the induced pmf $\hat{Q}(x^n, y^n)$ changes to the conditional pmf $\hat{Q}(x^n, y^n|g)$. But if G is almost independent of (X^n, Y^n), the conditional pmf $\hat{Q}(x^n, y^n|g)$ would be close to the desired distribution as well. To obtain the independence, we again use [9, Th. 1]. Substituting $T = 1$, $X_1 \leftarrow U$ and $Z \leftarrow XY$ in [9, Th. 1], we find that if

$$\tilde{R} < H(U|XY), \tag{31}$$

then $Q(x^n, y^n, g) \overset{\epsilon_2^{(n)}}{\approx} p^{\mathsf{u}}(g)p(x^n, y^n)$, for some vanishing $\epsilon_2^{(n)}$. Thus, by triangular inequality for total variation distance, we have $\hat{Q}(x^n, y^n, g) \overset{\epsilon^{(n)}}{\approx} p^{\mathsf{u}}(g)p(x^n, y^n)$, where $\epsilon^{(n)} = \sum_{i=0}^2 \epsilon_i^{(n)}$. From the definition of total variation distance for random pmf's, the average of the total variation distance between $\hat{q}(x^n, y^n, g)$ and $p^{\mathsf{u}}(g)p(x^n, y^n)$ over all random binnings is small. Thus, there exists a fixed binning with the corresponding pmf $\hat{q}$ such that $\hat{q}(x^n, y^n, g) \overset{\epsilon^{(n)}}{\approx} p^{\mathsf{u}}(g)p(x^n, y^n)$. Next, [9, Lemma 9] guarantees the existence of an instance g such that $\hat{p}(x^n, y^n|g) \overset{2\epsilon^{(n)}}{\approx} p(x^n, y^n)$. Then the extra shared randomness G can be eliminated by fixing it to be $G = g$.

Finally, observe that the conditions of (6) are seen to be equivalent to (26), (28) and (31) after eliminating $\tilde{R}$ using Fourier-Motzkin elimination. ■

Proof of Theorem 5: We would like to prove that to simulate a BSC channel with a non-uniform input pmf from a BEC channel, we can do better than $p = \frac{\epsilon}{2}$ (obtained by a degradation scheme). To be more precise, let Bern(q) be the Bernoulli distribution with parameter q. We show that (Bern(q), BSC$(\frac{t\epsilon}{2})$) is in the admissible region of (BEC(ϵ), 0)

for any $t \in (0, 1]$ such that

$$t\left[\epsilon + (1-\epsilon)(1-h(q)) + h\left(\frac{\epsilon}{2}\right)\right] - \epsilon - h\left(\frac{t\epsilon}{2}\right) \geq 0. \quad (32)$$

Indeed if $q \neq 1/2$, this inequality strictly holds for $t = 1$. So for any $q \neq 1/2$, one can find $t < 1$ so that this inequality is still valid. This would demonstrate the sub-optimality of a degradation scheme for non-uniform input distributions.

We use Theorem 4 to prove the above claim. For this we need to specify the joint pmf of random variables $X, Y, U, \tilde{X}, \tilde{Y}$ as follows:

- Let X to be distributed according to Bern(q).
- Assume that $\tilde{X}$ is uniform over $\{0, 1\}$ and independent of X.
- Let W be Bern(t) and independent of $(X, \tilde{X})$.
- Define U as follows: let $U = (W, K)$ where $K = (X, \tilde{X})$ if $W = 0$, and $K = X + \tilde{X}(\mod 2)$ if $W = 1$.
- To specify $p(y|\tilde{y}, u)$ we proceed as follows:
  - If $W = 0$, we let $Y = X$; note that in this case X is a part of U.
  - If $W = 1$, we look at $\tilde{Y}$; if it is the erasure flag, we choose Y uniformly at random. Otherwise, $\tilde{Y} = \tilde{X}$, so we may let $Y = \tilde{X} + K(\mod 2) = X$.

This procedure induces the following distribution on (X, Y): X is chosen according to Bern(q); with probability $(1-t) + t(1-\epsilon)$ we have $Y = X$, and Y with probability $t\epsilon$ is chosen uniformly at random (and independent of X). This is equivalent with the BSC(p) channel with Bern(q) input distribution where

$$p = \frac{t\epsilon}{2}.$$

We now need to verify (6) for $R = 0$, i.e., $I(U; \tilde{Y}) > I(U; XY) \geq I(U; X)$. We have

$$I(U; \tilde{Y}) = (1-\epsilon)I(U; \tilde{X}) = (1-\epsilon)\big[1 - t + t(1-h(q))\big].$$

Next,

$$I(U; XY) = I(WK; XY) = I(W; XY) + I(K; XY|W).$$

Moreover,

$$\begin{aligned} I(W; XY) &= H(XY) - H(XY|W) \\ &= H(XY) - (1-t)H(X) - tH(XY|W=1) \\ &= h(q) + h(p) - (1-t)h(q) - t \cdot h(q) - t \cdot h(\frac{\epsilon}{2}) \\ &= h(p) - t \cdot h(\frac{\epsilon}{2}) \\ &= h(\frac{t\epsilon}{2}) - t \cdot h(\frac{\epsilon}{2}), \end{aligned}$$

and

$$\begin{aligned} I(K; XY|W) &= (1-t)I(X\tilde{X}; XX) + tI(X + \tilde{X}; XY) \\ &= 1 - t, \end{aligned}$$

where we use the fact that $\tilde{X}$ is independent of (X, Y) if $W = 1$. Therefore,

$$\begin{aligned} &I(U; \tilde{Y}) - I(U; XY) \\ &= (1-\epsilon)\big[1 - t + t(1-h(q))\big] - h(\frac{t\epsilon}{2}) + t \cdot h(\frac{\epsilon}{2}) - 1 + t \\ &= t\big[\epsilon + (1-\epsilon)(1-h(q)) + h(\frac{\epsilon}{2})\big] - \epsilon - h(\frac{t\epsilon}{2}), \end{aligned}$$

which is positive by assumption (32). ∎

B. Point to Point Channel: Outer Bound

Proof of Theorem 6: Take an encoding map $q^{\text{enc}}(\tilde{x}^n|x^n, \omega)$ and a decoding map $q^{\text{dec}}(y^n|\tilde{y}^n, \omega)$ with the induced distribution $q(x^n, y^n, \tilde{x}^n, \tilde{y}^n, \omega)$ as described in (3) such that

$$\left\| q(x^n, y^n) - \prod_{i=1}^{n} p(x_i)\, p(y_i|x_i) \right\|_1 \leq \epsilon. \quad (33)$$

We have the Markov chains

$$X^n \to \omega\tilde{X}^n \to \omega\tilde{Y}^n \to Y^n,$$

and

$$\omega \to \tilde{X}^n \to \tilde{Y}^n.$$

Moreover, ω is independent of X^n. Therefore,

$$I(X^n; Y^n|\omega) \leq I(\tilde{X}^n; \tilde{Y}^n|\omega).$$

On the other hand,

$$I(X^n; Y^n|\omega) = I(X^n; Y^n\omega) \geq I(X^n; Y^n),$$

and

$$I(\tilde{X}^n; \tilde{Y}^n|\omega) \leq I(\omega\tilde{X}^n; \tilde{Y}^n) = I(\tilde{X}^n; \tilde{Y}^n).$$

As a result, $I(X^n; Y^n) \leq I(\tilde{X}^n; \tilde{Y}^n)$.

To proceed let

$$f_1 = \frac{1}{n}\sum_{i=1}^{n} I(X_{[1:i-1]}Y_{[1:i-1]}; X_iY_i).$$

Since by (33) the induced distribution on (X^n, Y^n) by the code is almost i.i.d., f_1 should be small. Indeed, Lemma 2 below shows that f_1 vanishes as ϵ goes to zero. We can then write

$$\begin{aligned} \sum_{i=1}^{n} I(X_i; Y_i) &= \sum_{i=1}^{n} H(X_i) + H(Y_i) - H(X_i, Y_i) \\ &\leq \sum_{i=1}^{n} H(X_i) + H(Y_i) - H(X_i, Y_i|X_{[1:i-1]}, Y_{[1:i-1]}) \\ &= \sum_{i=1}^{n} H(X_i|X_{[1:i-1]}) + H(Y_i|Y_{[1:i-1]}) \\ &\quad - H(X_i, Y_i|X_{[1:i-1]}, Y_{[1:i-1]}) + I(X_i; X_{[1:i-1]}) \\ &\quad + I(Y_i; Y_{[1:i-1]}) \\ &= H(X^n) + H(Y^n) - H(X^n, Y^n) + \sum_{i=1}^{n} I(X_i; X_{[1:i-1]}) \\ &\quad + I(Y_i; Y_{[1:i-1]}) \\ &\leq I(X^n; Y^n) + 2nf_1 \\ &\leq I(\tilde{X}^n; \tilde{Y}^n) + 2nf_1 \\ &\leq \sum_{i=1}^{n} I(\tilde{X}_i; \tilde{Y}_i) + 2nf_1, \end{aligned} \quad (34)$$

where in the last step we used the familiar expansion of mutual information for memoryless channels used in the converse proof of a point-to-point channel.

For $\tilde{X}_i, \tilde{Y}_i$, let $\tilde{U}_i$ be the random variable such that $\tilde{X}_i \to \tilde{U}_i \to \tilde{Y}_i$ forms a Markov chain and $\beta I(\tilde{U}_i; \tilde{X}_i\tilde{Y}_i) + \gamma I(\tilde{U}_i; \tilde{X}_i) + \theta I(\tilde{U}_i; \tilde{Y}_i)$ reaches its minimum. Assume that $\tilde{U}_i$ is constructed to be conditionally independent of other variables given $\tilde{X}_i, \tilde{Y}_i$. Then, we obtain a random variable $\tilde{U}^n = (\tilde{U}_1, \tilde{U}_2, \cdots, \tilde{U}_n)$ such that

$$\tilde{X}^n \to \tilde{U}^n \to \tilde{Y}^n,$$

forms a Markov chain, and that $q(\tilde{u}^n|\tilde{x}^n)$ and $q(\tilde{y}^n|\tilde{u}^n)$ are product channels. The joint pmf of all random variables factorizes as

$$\begin{aligned} q(x^n, \omega, \tilde{x}^n, \tilde{u}^n, \tilde{y}^n, y^n) &= p(x^n)p(\omega)q^{\text{enc}}(\tilde{x}^n|x^n, \omega) \\ &\quad \times q(\tilde{u}^n|\tilde{x}^n)q(\tilde{y}^n|\tilde{u}^n)q^{\text{dec}}(y^n|\tilde{y}^n, \omega) \\ &= q(x^n, \omega, \tilde{x}^n, \tilde{u}^n)q(\tilde{y}^n|\tilde{u}^n) \\ &\quad \times q^{\text{dec}}(y^n|\tilde{y}^n, \omega) \\ &= q(x^n, \omega, \tilde{x}^n, \tilde{u}^n)q(y^n\tilde{y}^n|\tilde{u}^n, \omega). \end{aligned} \tag{35}$$

This shows that $X^n\tilde{X}^n \to \tilde{U}^n\omega \to \tilde{Y}^nY^n$ forms a Markov chain. In particular, we conclude that

$$X^n \to \tilde{U}^n\omega \to Y^n,$$

forms a Markov chain.

Let $U_i = (\tilde{U}^n, \omega)$. Then $X^n \to U_i \to Y^n$, and hence $X_i \to U_i \to Y_i$ form Markov chains. Next, we have

$$\begin{aligned} \sum_{i=1}^n I(U_i; X_i, Y_i) &\le \sum_{i=1}^n I(U_iX_{[1:i-1]}Y_{[1:i-1]}; X_i, Y_i) \\ &= \sum_{i=1}^n I(X_{[1:i-1]}Y_{[1:i-1]}; X_i, Y_i) \qquad (36) \\ &\quad + \sum_{i=1}^n I(U_i; X_i, Y_i|X_{[1:i-1]}, Y_{[1:i-1]}) \\ &= nf_1 + \sum_{i=1}^n I(\tilde{U}^n, \omega; X_i, Y_i|X_{[1:i-1]}, Y_{[1:i-1]}) \\ &= nf_1 + I(\tilde{U}^n\omega; X^n, Y^n) \\ &\le nf_1 + H(\omega) + I(\tilde{U}^n; X^n, Y^n) \\ &\le nf_1 + H(\omega) + I(\tilde{U}^n; \tilde{X}^n, \tilde{Y}^n) \\ &\le nf_1 + nR + H(\tilde{U}^n) - H(\tilde{U}^n|\tilde{X}^n, \tilde{Y}^n) \\ &\le nf_1 + nR \qquad (37) \\ &\quad + \sum_{i=1}^n H(\tilde{U}_i) - H(\tilde{U}_i|\tilde{X}^n, \tilde{Y}^n, \tilde{U}_{[1:i-1]}) \\ &= nf_1 + nR + \sum_{i=1}^n H(\tilde{U}_i) - H(\tilde{U}_i|\tilde{X}_i, \tilde{Y}_i) \qquad (38) \\ &= nf_1 + nR + \sum_{i=1}^n I(\tilde{U}_i; \tilde{X}_i, \tilde{Y}_i), \qquad (39) \end{aligned}$$

where equation (38) follows from the fact that $p(\tilde{u}^n|\tilde{x}^n, \tilde{y}^n) = \prod_{i=1}^n p(\tilde{u}_i|\tilde{x}_i, \tilde{y}_i)$. This fact follows from the factorization

$$p(\tilde{x}^n, \tilde{u}^n, \tilde{y}^n) = p(\tilde{x}^n)\prod_{i=1}^n p(\tilde{u}_i, \tilde{y}_i|\tilde{x}_i).$$

Further, we have

$$\begin{aligned} \sum_{i=1}^n I(U_i; Y_i) &\le \sum_{i=1}^n I(U_i, Y_{[1:i-1]}; Y_i) \\ &= \sum_{i=1}^n I(Y_{[1:i-1]}; Y_i) + \sum_{i=1}^n I(U_i; Y_i|Y_{[1:i-1]}) \\ &\le nf_1 + \sum_{i=1}^n I(\tilde{U}^n, \omega; Y_i|Y_{[1:i-1]}) \\ &= nf_1 + I(\tilde{U}^n\omega; Y^n) \\ &\le nf_1 + 2H(\omega) + I(\tilde{U}^n; \tilde{Y}^n) \qquad (40) \\ &\le nf_1 + 2nR + H(\tilde{Y}^n) - H(\tilde{Y}^n|\tilde{U}^n) \\ &\le nf_1 + 2nR + \sum_{i=1}^n H(\tilde{Y}_i) - H(\tilde{Y}_i|\tilde{U}^n, \tilde{Y}_{[1:i-1]}) \\ &= nf_1 + 2nR + \sum_{i=1}^n H(\tilde{Y}_i) - H(\tilde{Y}_i|\tilde{U}_i) \qquad (41) \\ &= nf_1 + 2nR + \sum_{i=1}^n I(\tilde{U}_i; \tilde{Y}_i). \qquad (42) \end{aligned}$$

where equation (40) holds because

$$\begin{aligned} I(\tilde{U}^n\omega; Y^n) &\le H(\omega) + I(\tilde{U}^n; Y^n|\omega) \\ &\le H(\omega) + I(\tilde{U}^n; \tilde{Y}^nY^n|\omega) \\ &= H(\omega) + I(\tilde{U}^n; \tilde{Y}^n|\omega) \\ &\le 2H(\omega) + I(\tilde{U}^n; \tilde{Y}^n) \end{aligned}$$

and equation (41) is due to the fact that $p(\tilde{y}^n|\tilde{u}^n) = \prod_{i=1}^n p(\tilde{y}_i|\tilde{u}_i)$. Following similar steps, using the fact that $p(\tilde{u}^n|\tilde{x}^n) = \prod_{i=1}^n p(\tilde{u}_i|\tilde{x}_i)$ and

$$\begin{aligned} I(\tilde{U}^n\omega; X^n) &\le H(\omega) + I(\tilde{U}^n; X^n) \\ &\le H(\omega) + I(\tilde{U}^n; \tilde{X}^nX^n) \\ &= H(\omega) + I(\tilde{U}^n; \tilde{X}^n) \end{aligned}$$

which holds because of equation (35), one can show that

$$\sum_{i=1}^n I(U_i; X_i) \le nf_1 + nR + \sum_{i=1}^n I(\tilde{U}_i; \tilde{X}_i). \tag{43}$$

Let T be a random variable distributed uniformly over $[1 : n]$ and independent of previously defined random variables. Then, inequalities (34)-(43) can be equivalently written as

$$\begin{aligned} I(X_T; Y_T|T) &\le I(\tilde{X}_T; \tilde{Y}_T|T) + 2f_1, \\ I(U_T; X_TY_T|T) &\le I(\tilde{U}_T; \tilde{X}_T\tilde{Y}_T|T) + R + f_1, \\ I(U_T; X_T|T) &\le I(\tilde{U}_T; \tilde{X}_T|T) + R + f_1, \\ I(U_T; Y_T|T) &\le I(\tilde{U}_T; \tilde{Y}_T|T) + 2R + f_1. \end{aligned}$$

Let $f_2 = I(T; X_T, Y_T)$. Observe that by Lemma 2 at the end of the proof, f_2 vanishes as ϵ converges to zero. Then the above set of equations imply that

$$\begin{aligned} I(X_T; Y_T) &\le I(\tilde{X}_T; \tilde{Y}_T|T) + 2f_1 + f_2, \\ I(U_T; X_T, Y_T) &\le I(\tilde{U}_T; \tilde{X}_T\tilde{Y}_T|T) + R + f_1 + f_2, \\ I(U_T; X_T) &\le I(\tilde{U}_T; \tilde{X}_T|T) + R + f_1 + f_2, \\ I(U_T; Y_T) &\le I(\tilde{U}_T; \tilde{Y}_T|T) + 2R + f_1 + f_2. \end{aligned}$$

$$I(X;Y)+\min_{U:X-U-Y}\big[\beta I(U;XY)+\gamma\, I(U;X)+\theta I(U;Y)\big] \tag{44}$$

$$\le I(\tilde{X}_T;\tilde{Y}_T|T)+\Big[\beta I(\tilde{U}_T;\tilde{X}_T\tilde{Y}_T|T)+\gamma\, I(\tilde{U}_T;\tilde{X}_T|T)+\theta I(\tilde{U}_T;\tilde{Y}_T|T)\Big] +(\beta+\gamma+2\theta)R+f_1+(f_1+f_2)(1+\beta+\gamma+\theta) \tag{45}$$

$$\le \max_t\Big(I(\tilde{X}_T;\tilde{Y}_T|T=t)+\Big[\beta I(\tilde{U}_T;\tilde{X}_T\tilde{Y}_T|T=t)+\gamma\, I(\tilde{U}_T;\tilde{X}_T|T=t)+\theta I(\tilde{U}_T;\tilde{Y}_T|T=t)\Big]\Big) +(\beta+\gamma+2\theta)R+f_1+(f_1+f_2)(1+\beta+\gamma+\theta) \tag{46}$$

$$= \max_t\Big(I(\tilde{X}_t;\tilde{Y}_t)+\Big[\beta I(\tilde{U}_t;\tilde{X}_t\tilde{Y}_t)+\gamma\, I(\tilde{U}_t;\tilde{X}_t)+\theta I(\tilde{U}_t;\tilde{Y}_t)\Big]\Big) +(\beta+\gamma+2\theta)R+f_1+(f_1+f_2)(1+\beta+\gamma+\theta) \tag{47}$$

$$= \max_t\Bigg(I(\tilde{X}_t;\tilde{Y}_t)+\min_{\tilde{U}:\tilde{X}_t-\tilde{U}-\tilde{Y}_t}\Big[\beta I(\tilde{U};\tilde{X}_t\tilde{Y}_t)+\gamma\, I(\tilde{U};\tilde{X}_t)+\theta I(\tilde{U};\tilde{Y}_t)\Big]\Bigg) +(\beta+\gamma+2\theta)R+f_1+(f_1+f_2)(1+\beta+\gamma+\theta) \tag{48}$$

$$\le \max_{p(\tilde{x})}\Big[I(\tilde{X};\tilde{Y})+\min_{\tilde{U}:\tilde{X}-\tilde{U}-\tilde{Y}}\big[\beta I(\tilde{U};\tilde{X}\tilde{Y})+\gamma\, I(\tilde{U};\tilde{X})+\theta I(\tilde{U};\tilde{Y})\big]\Big] +(\beta+\gamma+2\theta)R+f_1+(f_1+f_2)(1+\beta+\gamma+\theta). \tag{49}$$

Let $X=X_T$ and $Y=Y_T$, and note that $X\to U_T\to Y$ forms a Markov chain. Then by the above inequalities for non-negative reals β, γ, and θ we can proceed with equations (44) to (49), as shown at the top of this page. Recall that both f_1 and f_2 converge to zero as ϵ goes to zero. Furthermore, by (33) the joint pmf of (X_T,Y_T) converges to the desired pmf $p(x)p(y|x)$ as ϵ converges to zero. Therefore, to complete the proof, it remains to show that the expression

$$I(X;Y)+\min_{U:X-U-Y}\big[\beta I(U;XY)+\gamma\, I(U;X)+\theta I(U;Y)\big]$$

is a continuous function of the joint distribution on (X,Y). Equivalently, we need to show that the function

$$g(\epsilon)=\min_{\substack{q_{uxy}\\ X-U-Y\\ q(xy)\overset{\epsilon}{\approx}p(xy)}}\big[\beta I(U;XY)+\gamma\, I(U;X)+\theta I(U;Y)\big],$$

for a fixed $p(xy)$, satisfies $\lim_{\epsilon\to 0}g(\epsilon)=g(0)$.

First, observe that $g(\epsilon)$ is a decreasing function of ϵ; in particular, $g(0)\ge g(\epsilon)$ for all $\epsilon>0$. Thus, the limit $\lim_{\epsilon\longrightarrow 0}g(\epsilon)$ exists and is at most $g(0)$. Second, for every $\epsilon>0$, the minimization over U, can be restricted to random variables U with cardinality bound $|\mathcal{U}|\le|\mathcal{X}\times\mathcal{Y}|$. Let $p_\epsilon(x,y,u)$ be an optimal point in the minimization. Since $p_\epsilon(x,y,u)$ belongs to the compact set of the probability simplex on a finite alphabet set, the set of optimal points has a limit point $p^*(xyu)$. We then have

$$g(0)\ge\lim_{\epsilon\to 0}g(\epsilon)=\beta I_{p^*}(U;XY)+\gamma\, I_{p^*}(U;X)+\theta I_{p^*}(U;Y).$$

Moreover, we have $p^*(xy)=p(xy)$, and by the continuity of mutual information, $X-U-Y$ holds for the limit distribution $p^*(xyu)$ as well. Now by definition we have

$$\begin{aligned} g(0)&=\min_{U:X-U-Y}\big[\beta I(U;XY)+\gamma\, I(U;X)+\theta I(U;Y)\big]\\ &\le \beta I_{p^*}(U;XY)+\gamma\, I_{p^*}(U;X)+\theta I_{p^*}(U;Y)\\ &=\lim_{\epsilon\to 0}g(\epsilon)\\ &\le g(0).\end{aligned}$$

This completes the proof. ∎

In the above proof we used the following lemma from [15].

Lemma 2 (Entropy and Timing Information of Nearly i.i.d. Sequences [15]): For any discrete random variables W^n whose pmf satisfies

$$\Big\|p(w^n)-\prod_{t=1}^n\hat{p}_t(w_t)\Big\|_1<\epsilon<\frac{1}{4},$$

for some $\hat{p}_1(w),\dots,\hat{p}_n(w)$, we have

$$\sum_{t=1}^n I(W_t;W^{t-1})\le 4n\epsilon(\log|\mathcal{W}|+\log\frac{1}{\epsilon}).$$

Moreover, for any random variable $T\in\{1,\cdots,n\}$ independent of W^n,

$$I(W_T;T)\le 4n\epsilon(\log|\mathcal{W}|+\log\frac{1}{\epsilon}).$$

1) Equivalent Characterization for Symmetric Channels: Proof of Theorem 7: We claim that, for a symmetric channel, the maximum on the right hand side of (8) is achieved at the uniform distribution $p^{\mathsf{u}}(\tilde{x})$. We prove this for binary input channels, and the proof for general channels is done in a similar way. More specifically, we show that if $p=p(\tilde{X}=0)$ and we let $g(p)$ to be equal to

$$I(\tilde{X};\tilde{Y})+\min_{\tilde{U}:\tilde{X}-\tilde{U}-\tilde{Y}}\big[\beta I(\tilde{U};\tilde{X}\tilde{Y})+\gamma\, I(\tilde{U};\tilde{X})+\theta I(\tilde{U};\tilde{Y})\big],$$

then $g(p)$ is maximized at $p=\frac{1}{2}$. To show this, we first claim that $g(p)=g(1-p)$. Take some $p(\tilde{x}\tilde{y}\tilde{u})$ such that $p=p(\tilde{X}=0)$ and $\tilde{X}-\tilde{U}-\tilde{Y}$. Let

$$\tilde{X}'=1-\tilde{X},\ \tilde{Y}'=\pi_Y(Y),\ \tilde{U}'=\tilde{U},$$

where π_Y is the permutation corresponding to permutation $\pi_X(0)=1,\pi_X(1)=0$ such that

$$p_{Y|X}(\pi_Y(y)|\pi_X(x))=p_{Y|X}(y|x).$$

Clearly, all the mutual information terms remain the same for $\tilde{X}',\tilde{U}',\tilde{Y}'$, and by the symmetry of $p(\tilde{y}|\tilde{x})$, the two channels $\tilde{X}\to\tilde{Y}$ and $\tilde{X}'\to\tilde{Y}'$ are the same. On the other hand,

$p(\tilde{X}_1 = 0) = 1 - p$. This means that, for every choice of $\tilde{U}$ in the minimization of $g(p)$ there is a choice of $\tilde{U}$ in minimization of $g(1-p)$ that leads to the same answer. As a result, $g(1-p) = g(p)$.

Let $p(\tilde{x}\tilde{y}\tilde{u})$ be the distribution with $\tilde{X} - \tilde{U} - \tilde{Y}$, that achieves the minimum in $g(1/2)$, i.e.,

$$g(1/2) = I(\tilde{X};\tilde{Y}) + \beta I(\tilde{U};\tilde{X}\tilde{Y}) + \gamma\, I(\tilde{U};\tilde{X}) + \theta I(\tilde{U};\tilde{Y}).$$

Now, fix the channel $p(\tilde{y}\tilde{u}|\tilde{x})$, and instead of the uniform distribution on $\tilde{X}$ put the distribution Bern(p) on $\tilde{X}$. Denote the resulting distribution by $q_p(\tilde{x}\tilde{y}\tilde{u})$. Then by definition we have

$$\begin{aligned} g(p) \le\ & I_{q_p}(\tilde{X};\tilde{Y}) + \beta I_{q_p}(\tilde{U};\tilde{X}\tilde{Y}) + \gamma\, I_{q_p}(\tilde{U};\tilde{X}) \\ & + \theta I_{q_p}(\tilde{U};\tilde{Y}). \end{aligned} \tag{50}$$

We similarly have that

$$\begin{aligned} g(1-p) \le\ & I_{q_{(1-p)}}(\tilde{X};\tilde{Y}) + \beta I_{q_{(1-p)}}(\tilde{U};\tilde{X}\tilde{Y}) \\ & + \gamma\, I_{q_{(1-p)}}(\tilde{U};\tilde{X}) + \theta I_{q_{(1-p)}}(\tilde{U};\tilde{Y}). \end{aligned} \tag{51}$$

Observe that for a fixed $p(\tilde{u}, \tilde{y}|\tilde{x})$ the expression

$$I(\tilde{X};\tilde{Y}) + \beta I(\tilde{U};\tilde{X}\tilde{Y}) + \gamma\, I(\tilde{U};\tilde{X}) + \theta I(\tilde{U};\tilde{Y}),$$

is a concave function of $p(\tilde{x})$; this is because mutual information is concave in input distribution for a fixed channel implying that the first, third and fourth term are concave; the third term is equal to $I(\tilde{U};\tilde{X}\tilde{Y}) = I(\tilde{U};\tilde{X}) + I(\tilde{U};\tilde{Y}|\tilde{X})$ which is a concave term plus a linear term. Therefore, by (50) and (51) and this concavity we obtain

$$g(p) = \frac{1}{2}(g(p) + g(1-p)) \le g(\frac{1}{2}).$$

This proves our claim.

Now by Theorem 6 and the above claim, for any non-negative real numbers β, γ, and θ we have

$$\begin{aligned} & I(X;Y) + \min_{U:X-U-Y} \big[\beta I(U;XY) + \gamma\, I(U;X) + \theta I(U;Y)\big] \\ & \le I(\tilde{X};\tilde{Y}) + \min_{\tilde{U}:\tilde{X}-\tilde{U}-\tilde{Y}} \big[\beta I(\tilde{U};\tilde{X}\tilde{Y}) + \gamma\, I(\tilde{U};\tilde{X}) \\ & \quad + \theta I(\tilde{U};\tilde{Y})\big], \end{aligned} \tag{52}$$

in which $p(\tilde{x})$ is fixed to be the uniform distribution. Since this inequality holds for all β, γ and θ, we find that $I(\tilde{X};\tilde{Y}) \ge I(X;Y)$ and further

$$\begin{aligned} & \min_{U:X-U-Y} \big[\beta I(U;XY) + \gamma\, I(U;X) + \theta I(U;Y)\big] \\ & \quad \le \min_{\tilde{U}:\tilde{X}-\tilde{U}-\tilde{Y}} \big[\beta I(\tilde{U};\tilde{X}\tilde{Y}) + \gamma\, I(\tilde{U};\tilde{X}) + \theta I(\tilde{U};\tilde{Y})\big]. \end{aligned}$$

Then by the definition of $\mathcal{S}(p(y|x), p(x))$, the supporting hyperplane theorem would imply statement of the theorem, i.e.,

$$\mathcal{S}(p(\tilde{y}|\tilde{x}), p(\tilde{x})) \subseteq \mathcal{S}(p(y|x), p(x)),$$

if we show that $\mathcal{S}(p(y|x), p(x))$ is a convex set.

Here we prove that $\mathcal{S}(p(y|x), p(x))$ is convex. Corresponding to any two points in $\mathcal{S}(p(y|x), p(x))$, one can find two random variables U_1, U_2 such that $X - U_1 - Y$ and $X - U_2 - Y$ form Markov chains. Let T be a uniform random variable on $\{1, 2\}$ and independent of X, Y, U_1, and U_2. Let $U = (T, U_T)$. We clearly have $I(X;Y|U) = 0$ and further $I(U;XY) = \frac{1}{2}(I(U_1;XY) + I(U_2;XY))$ etc. Therefore, we can use U to show that the average of the two original points belongs to $\mathcal{S}(p(y|x), p(x))$. ■

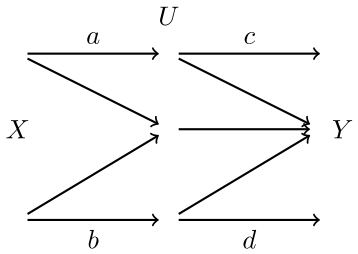

Fig. 8. The form of $X \to U \to Y$ in the definition of $\mathcal{S}(\text{BEC}(\epsilon), p^{\mathsf{U}})$.

C. Point-to-Point Channel: BEC vs BSC

Proof of Theorem 8:

(i) In this part, we would like to compute $\mathcal{S}(\text{BEC}(\epsilon), p^{\mathsf{U}})$ with uniform input distribution. Take some $p(u|xy)$ such that $X - U - Y$, and assume without loss of generality that $p(u) > 0$ for all $u \in \mathcal{U}$. Define U' as a function of U as follows:

$$U' = \begin{cases} 0 & \text{if } U = u \text{ and } p(X=0|U=u) = 1, \\ 1 & \text{if } U = u \text{ and } p(X=1|U=u) = 1, \\ e & \text{if } U = u \text{ and } p(X=1|U=u) > 0, \\ & P(X=0|U=u) > 0. \end{cases}$$

Then we claim that $X - U' - Y$ forms a Markov chain. Observe that $I(X;Y|U'=0) = I(X;Y|U'=1) = 0$, since X is deterministic if $U' = 0$ or $U' = 1$. Moreover, $U' = e$ implies that $Y = e$ is deterministic and hence $I(X;Y|U'=e) = 0$; this is because if for instance $p(Y=0|U'=e) > 0$, then

$$p(Y=0|X=1) \ge p(U'=e|X=1)p(Y=0|U'=e) > 0,$$

which is a contradiction. Therefore, $X - U' - Y$ forms a Markov chain.

Since U' is a function of U we have

$$\begin{aligned} I(U';XY) &\le I(U;XY), \\ I(U';X) &\le I(U;X), \\ I(U';Y) &\le I(U;Y). \end{aligned}$$

Therefore, in the definition of $\mathcal{S}(\text{BEC}(\epsilon), p^{\mathsf{U}})$ without loss of generality we may assume that U has the form of U' defined above. This form is depicted in Fig. 8. Here $a, b, c, d \in [1-\epsilon, 1]$ are arbitrary numbers with $\epsilon = 1 - ac = 1 - bd$. With the latter equations U is indeed determined by the pair (a, b) since c and d are computed in terms of a, b and ϵ.

We claim that for the uniform input distribution $p(x) = p^{\mathsf{U}}(x)$ it suffices to consider the symmetric case with $a = b$ and $c = d$. Observe that $p(x, y, u)$ is linear in terms

of a and b, e.g.,

$$\begin{aligned} p(X=0,Y=0,U=0) &= \frac{ac}{2} = \frac{1-\epsilon}{2}, \\ P(X=0,Y=e,U=0) &= \frac{a(1-c)}{2} = \frac{a-ac}{2} \\ &= \frac{a-1+\epsilon}{2}, \\ P(X=0,Y=e,U=e) &= \frac{1-a}{2}. \end{aligned}$$

On the other hand $I(U;XY) = H(XY) - H(XY|U)$ is a convex function when we linearly change the joint pmf of $p(x,u,y)$ while fixing $p(xy)$. Therefore, the value of $I(U;XY)$ at $(\frac{a+b}{2}, \frac{a+b}{2})$ is less than or equal to the average of its values at (a,b) and (b,a). Moreover, by symmetry, the value of $I(U;XY)$ is the same at (a,b) and (b,a). We conclude that $I(U;XY)$ at $(\frac{a+b}{2}, \frac{a+b}{2})$ is not greater than this value at (a,b). The same argument works for $I(U;X)$ and $I(U;Y)$ as well. Therefore, the three terms $I(U;XY), I(U;X)$ and $I(U;Y)$ are simultaneously minimized when $a=b$, and then $c=d$.

Using the Markov chain condition $X-U-Y$ we have

$$\begin{aligned} I(U;XY) &= H(XY) - H(XY|U) \\ &= H(XY) - H(X|U) - H(Y|U) \\ &= H(XY) - H(X) - H(Y) + H(X) + H(Y) \\ &\quad -H(X|U) - H(Y|U) \\ &= -I(X;Y) + I(X;U) + I(Y;U) \end{aligned}$$

Then, for the BEC channel with parameter ϵ we have

$$I(U;XY) = -1 + \epsilon + I(X;U) + I(Y;U).$$

Moreover, for $a=b \geq 1-\epsilon$ we have $H(X|U) = 1-a$, and $H(Y|U) = ah(\frac{1-\epsilon}{a})$. Then

$$\begin{aligned} I(U;XY) &= h(\epsilon) + a - ah(\frac{1-\epsilon}{a}) \\ I(U;X) &= a \\ I(U;Y) &= 1 - \epsilon + h(\epsilon) - ah(\frac{1-\epsilon}{a}) \end{aligned}$$

The result then follows by a straightforward computation.

(ii) We adapt the approach of Wyner to weighted sum calculations to prove our result. Take the channel BSC(p) with uniform input distribution. Take some arbitrary auxiliary U such that $X-U-Y$. We define two random variables as functions of U by

$$A = p(X=0|U), \qquad B = p(Y=0|U).$$

Then we have

$$H(X|U) = \mathbb{E}[h(A)], \qquad H(Y|U) = \mathbb{E}[h(B)].$$

Furthermore,

$$p(X=0) = \mathbb{E}[A], \qquad p(Y=0) = \mathbb{E}[B].$$

Also

$$p(X=0,Y=0|U) = p(X=0|U)p(Y=0|U) = AB,$$

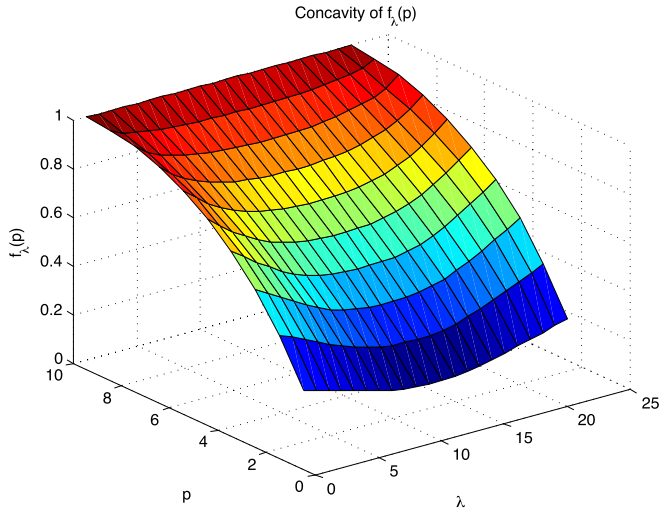

Fig. 9. Concavity of $f_\lambda(p)$ with respect to $0.05 \leq p \leq 0.5$ and for all $0 \leq \lambda \leq 1$.

and hence $p(X=0,Y=0) = \mathbb{E}[AB]$. Therefore, we have

$$I(U;XY) = 1 + h(p) - \mathbb{E}[h(A)] - \mathbb{E}[h(B)], \quad (53)$$
$$I(U;X) = 1 - \mathbb{E}[h(A)], \quad (54)$$
$$I(U;Y) = 1 - \mathbb{E}[h(B)]. \quad (55)$$

Here A, B are real-valued random variables satisfying:

$$\begin{aligned} A, B &\in [0,1], \\ \mathbb{E}[A] &= \mathbb{E}[B] = \frac{1}{2}, \\ \mathbb{E}[AB] &= \frac{1-p}{2}. \end{aligned}$$

By the above equations to compute $\mathcal{S}(\text{BSC}(\epsilon), p^{\mathsf{U}})$ we need to characterize the set

$$\bigcup_{A,B} \big\{(b_1,b_2) : b_1 \leq \mathbb{E}[h(A)], \quad b_2 \leq \mathbb{E}[h(B)]\big\},$$

where we take union over all real-valued random variables A, B satisfying the above constraints. Equivalently, for any $\lambda \in [0,1]$ we need to compute

$$\max \lambda \mathbb{E}[h(A)] + \bar{\lambda}\mathbb{E}[h(B)], \quad (56)$$

over all A, B. We show that here the maximum occurs at two binary random variables A and B that correspond to $X \to U$ and $U \to Y$ being BSC channels.

Let $X \to U$ be a BSC with parameter α, and let $U \to Y$ be another BSC with parameter β. We need the induced channel $X \to Y$ be a BSC with parameter $p \in [0, 1/2]$. This is equivalent to

$$p = \alpha * \beta = \alpha\bar{\beta} + \bar{\alpha}\beta. \quad (57)$$

For this special U we get

$$\max_{\substack{\alpha,\beta \\ \alpha*\beta=p}} \lambda H(X|U) + \bar{\lambda} H(Y|U) = \max_{\substack{\alpha,\beta \\ \alpha*\beta=p}} \lambda h(\alpha) + \bar{\lambda} h(\beta).$$

We then make the following conjecture:

conjecture 1: Let

$$f_\lambda(p) = \max_{\substack{\alpha,\beta \\ \alpha*\beta=p}} \lambda h(\alpha) + \bar{\lambda} h(\beta).$$

Then $f_\lambda(p)$ is a concave function of p for all λ, as plotted in Fig. 9.

Using this conjecture, we show that the answer to the maximization (56) is also $f_\lambda(p)$ defined above. From the definitions it is clear that $f_\lambda(p)$ is a lower bound on (56). To prove inequality in the other direction, take A, B with the above conditions. Assume that $(A, B) = (\alpha_i, \beta_j)$ happens with probability q_{ij}. We have

$$\begin{aligned}\lambda\mathbb{E}[h(A)] + \bar{\lambda}\mathbb{E}[h(B)] &= \sum_{i,j} q_{ij}[\lambda h(\alpha_i) + \bar{\lambda} h(\beta_j)] \\ &\le \sum_{i,j} q_{ij} f_\lambda(\alpha_i\bar{\beta}_j + \bar{\alpha}_i\beta_j) \\ &\le f_\lambda\Big(\sum_{i,j} q_{ij}(\alpha_i\bar{\beta}_j + \bar{\alpha}_i\beta_j)\Big) \\ &= f_\lambda(\mathbb{E}[A] + \mathbb{E}[B] - 2\mathbb{E}[AB]) \\ &= f_\lambda(p).\end{aligned}$$

Therefore, to compute (56) we may restrict to auxiliary U where $X \to U$ and $U \to Y$ are BSC channels with parameters α and β respectively, with $p = \alpha * \beta$. In this case, using equations (53)-(55) we have

$$\begin{aligned}I(U; XY) &= 1 + h(p) - h(\alpha) - h(\beta), \\ I(U; X) &= 1 - h(\alpha), \\ I(U; Y) &= 1 - h(\beta).\end{aligned}$$

These give the desired result. ■

D. Broadcast Channel

Proof of Theorem 9: The structure of the proof of this theorem is similar to that of Theorem 4 and has three parts.

Part (1): We define two protocols each of which induces a joint distribution on random variables that will be used in the proof.

Protocol A [Not useful for coding]: Let $(W^n, U^n, V^n, X^n, \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n, Y^n, Z^n)$ be n i.i.d. repetitions of the joint pmf $p(w, u, v, x, \tilde{x}, \tilde{y}, \tilde{z}, y, z)$. Consider the following construction:

- To each sequence $w^n \in \mathcal{W}^n$ assign two random bin indices $g_0 \in [1 : 2^{n\tilde{R}_0}]$ and $\omega \in [1 : 2^{nR}]$.
- To each pair of sequences (w^n, u^n), assign a random bin index $g_1 \in [1 : 2^{n\tilde{R}_1}]$.
- To each pair of sequences (w^n, v^n), assign a random bin index $g_2 \in [1 : 2^{n\tilde{R}_2}]$.
- Consider two Slepian-Wolf decoders to estimate $(\hat{w}_1^n, \hat{u}^n)$ and $(\hat{w}_2^n, \hat{v}^n)$ from $(\omega, g_1, \tilde{y}^n)$ and $(\omega, g_2, \tilde{z}^n)$, respectively. Here we are considering $\tilde{y}^n$ and $\tilde{z}^n$ as side information, and (ω, g_1) and (ω, g_2) as random bins of the sources (w^n, u^n) and (w^n, v^n) that we want to decode. Note that $\hat{w}_1^n$ and $\hat{w}_2^n$ are reconstructions of w^n by two different Slepian-Wolf decoders.

The constraints on the rates R, $\tilde{R}_0$, $\tilde{R}_1$ and $\tilde{R}_2$ for the success of the decoders will be imposed later. The random pmf induced by the random binning, denoted by Q, can be expressed as follows:

$$\begin{aligned}&Q(x^n, y^n, z^n, w^n, u^n, v^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\quad= p(x^n)p(w^n, u^n, v^n, \tilde{x}^n|x^n)Q(g_0, \omega|w^n)Q(g_1|w^n, u^n) \\ &\qquad\times Q(g_2|w^n, v^n)p(\tilde{y}^n, \tilde{z}^n|\tilde{x}^n)Q^{\mathsf{SW}}(\hat{w}_1^n, \hat{u}^n|g_0, g_1, \tilde{y}^n, \omega) \\ &\qquad\times Q^{\mathsf{SW}}(\hat{w}_2^n, \hat{v}^n|g_0, g_2, \tilde{z}^n, \omega)p(y^n|w^n, u^n, \tilde{y}^n) \\ &\qquad\times p(z^n|w^n, v^n, \tilde{z}^n) \\ &\quad= p(x^n)Q(w^n, u^n, v^n, \tilde{x}^n, g_{[0:2]}, \omega|x^n) \\ &\qquad\times p(\tilde{y}^n, \tilde{z}^n|\tilde{x}^n)Q^{\mathsf{SW}}(\hat{w}_1^n, \hat{u}^n|g_0, g_1, \tilde{y}^n, \omega) \\ &\qquad\times Q^{\mathsf{SW}}(\hat{w}_2^n, \hat{v}^n|g_0, g_2, \tilde{z}^n, \omega)p(y^n|w^n, u^n, \tilde{y}^n) \\ &\qquad\times p(z^n|w^n, v^n, \tilde{z}^n) \\ &\quad= p(x^n)Q(g_{[0:2]}, \omega|x^n)Q(w^n, u^n, v^n, \tilde{x}^n|x^n, g_{[0:2]}, \omega) \\ &\qquad\times p(\tilde{y}^n, \tilde{z}^n|\tilde{x}^n)Q^{\mathsf{SW}}(\hat{w}_1^n, \hat{u}^n|g_0, g_1, \tilde{y}^n, \omega) \\ &\qquad\times Q^{\mathsf{SW}}(\hat{w}_2^n, \hat{v}^n|g_0, g_2, \tilde{z}^n, \omega) \\ &\qquad\times p(y^n|w^n, u^n, \tilde{y}^n)p(z^n|w^n, v^n, \tilde{z}^n).\end{aligned}$$

Protocol B [Useful for coding after removing extra common randomnesses]. In this protocol we assume that the sender and receivers have access to the extra common randomness (G_0, G_1, G_2) where G_0, G_1, G_2 are mutually independent of X^n and ω. It is further assumed that G_0, G_1 and G_2 are distributed uniformly over the sets $[1 : 2^{n\tilde{R}_0}]$, $[1 : 2^{n\tilde{R}_1}]$ and $[1 : 2^{n\tilde{R}_2}]$, respectively. Now we use the following protocol:

- First, the sender having $(g_{[0:2]}, \omega, x^n)$ generates $(w^n, u^n, v^n, \tilde{x}^n)$ according to pmf $Q(w^n, u^n, v^n, \tilde{x}^n|x^n, g_{[0:2]}, \omega)$ of Protocol A, and sends $\tilde{x}^n$ over the memoryless broadcast channel $p(\tilde{y}^n, \tilde{z}^n|\tilde{x}^n)$. The first receiver gets $\tilde{y}^n$ and the second receiver gets $\tilde{z}^n$ from the channel. Having $(g_0, g_1, \omega, \tilde{y}^n)$, the first receiver uses the Slepian-Wolf decoder $Q^{\mathsf{SW}}(\hat{w}_1^n, \hat{u}^n|\omega, g_0, g_1, \tilde{y}^n)$ to estimate (w^n, u^n). Similarly, the second receiver uses the Slepian-Wolf decoder $Q^{\mathsf{SW}}(\hat{w}_2^n, \hat{v}^n|\omega, g_0, g_2, \tilde{z}^n)$ to obtain an estimate of (w^n, v^n). Here $\hat{w}_1^n$ and $\hat{w}_2^n$ are first and second receiver's estimate of w^n respectively.
- Having $(\tilde{y}^n, \hat{u}^n, \hat{w}_1^n)$, the first receiver generates y^n using $p(y^n|\tilde{y}^n, \hat{u}^n, \hat{w}_1^n) = \prod_{i=1}^n p(y_i|\tilde{y}_i, \hat{u}_i, \hat{w}_{1i})$. Similarly the second receiver generates z^n according to $p(z^n|\tilde{z}^n, \hat{v}^n, \hat{w}_2^n) = \prod_{i=1}^n p(z_i|\tilde{z}_i, \hat{v}_i, \hat{w}_{2i})$.

The random pmf induced by the second protocol, denoted by $\hat{P}$, is equal to

$$\begin{aligned}&\hat{Q}(x^n, y^n, z^n, w^n, u^n, v^n, \hat{w}_{[1:2]}^n, \hat{u}^n, \hat{v}^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\quad= p^{\mathsf{U}}(\omega)p^{\mathsf{U}}(g_{[0:2]})p(x^n) \\ &\qquad\times Q(w^n, u^n, v^n, \tilde{x}^n|x^n, g_{[0:2]}, \omega) \\ &\qquad\times p(\tilde{y}^n, \tilde{z}^n|\tilde{x}^n)Q^{\mathsf{SW}}(\hat{w}_1^n, \hat{u}^n|\omega, g_{[0:1]}, \tilde{y}^n) \\ &\qquad\times Q^{\mathsf{SW}}(\hat{w}_2^n, \hat{v}^n|\omega, g_0, g_2, \tilde{z}^n)p(y^n|\tilde{y}^n, \hat{u}^n, \hat{w}_1^n) \\ &\qquad\times p(z^n|\tilde{z}^n, \hat{v}^n, \hat{w}_2^n).\end{aligned}$$

Part (2): In this part we mention sufficient conditions under which the pmf's Q and $\hat{Q}$ induced by the above protocols are approximately equal. The first step is to observe that g_0, ω, g_1 and g_2 are bin indices of w^n, w^n, $w^n u^n$ and $w^n v^n$, respectively.

Substituting $T = 4$, $X_1 \leftarrow W, X_2 \leftarrow W, X_3 \leftarrow WU$, $X_4 \leftarrow WV$ and $Z \leftarrow \emptyset$ in [9, Th. 1], we find that if

$$\begin{aligned} R + \tilde{R}_0 &< H(W|X), \\ R + \tilde{R}_0 + \tilde{R}_1 &< H(WU|X), \\ R + \tilde{R}_0 + \tilde{R}_2 &< H(WV|X), \\ R + \tilde{R}_0 + \tilde{R}_1 + \tilde{R}_2 &< H(WUV|X), \end{aligned} \tag{58}$$

then there exists $\epsilon_0^{(n)} \to 0$ such that $Q(g_{[0:2]}, \omega|x^n) \overset{\epsilon_0^{(n)}}{\approx} p^{\mathsf{U}}(\omega)p^{\mathsf{U}}(g_{[0:2]}) = \hat{Q}(g_{[0:2]}, \omega)$. This implies that

$$\hat{Q}(x^n, w^n, u^n, v^n, \hat{w}_1^n, \hat{u}^n, \hat{w}_2^n, \hat{v}^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \tag{59}$$

$$\overset{\epsilon_0^{(n)}}{\approx} Q(x^n, w^n, u^n, v^n, \hat{u}^n, \hat{w}_1^n, \hat{v}^n, \hat{w}_2^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega). \tag{60}$$

Note that we have not yet included y^n and z^n in the above pmf's.

The next step is to find the conditions under which the Slepian-Wolf decoders of Protocol A work well with high probability. By the Slepian-Wolf theorem we need

$$\begin{aligned} \tilde{R}_0 + \tilde{R}_1 + R &> H(WU|\tilde{Y}), \\ \tilde{R}_1 &> H(U|W\tilde{Y}), \\ \tilde{R}_0 + \tilde{R}_2 + R &> H(WV|\tilde{Z}), \\ \tilde{R}_2 &> H(V|W\tilde{Z}). \end{aligned} \tag{61}$$

Then for an asymptotically vanishing sequence $\epsilon_1^{(n)}$, we have

$$\begin{aligned} &Q(x^n, w^n, u^n, v^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega, \hat{w}_1^n, \hat{u}^n, \hat{w}_2^n, \hat{v}^n) \\ &\overset{\epsilon_1^{(n)}}{\approx} Q(x^n, w^n, u^n, v^n, g_{[1:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\quad \times \mathbf{1}\{\hat{w}_1^n = \hat{w}_2^n = w^n, \hat{u}^n = u^n, \hat{v}^n = v^n\}. \end{aligned} \tag{62}$$

Using (60) and (62) and [9, Lemma 9] we have

$$\begin{aligned} &\hat{Q}(x^n, w^n, u^n, \hat{w}_1^n, v^n, \hat{u}^n, \hat{w}_2^n, \hat{v}^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\overset{\epsilon_0^{(n)}+\epsilon_1^{(n)}}{\approx} Q(x^n, u^n, v^n, w^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\quad \times \mathbf{1}\{\hat{w}_1^n = \hat{w}_2^n = w^n, \hat{u}^n = u^n, \hat{v}^n = v^n\}. \end{aligned} \tag{63}$$

Moreover, the third part of [9, Lemma 3] implies that

$$\begin{aligned} &\hat{Q}(x^n, w^n, u^n, v^n, \hat{w}_1^n, \hat{w}_2^n, \hat{u}^n, \hat{v}^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\qquad \times p(z^n|\tilde{z}^n, \hat{v}^n, \hat{w}_2^n)p(y^n|\tilde{y}^n, \hat{u}^n, \hat{w}_1^n) \\ &\overset{\epsilon_0^{(n)}+\epsilon_1^{(n)}}{\approx} Q(x^n, u^n, w^n, v^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\qquad \times \mathbf{1}\{\hat{w}_1^n = \hat{w}_2^n = w^n, \hat{u}^n = u^n, \hat{v}^n = v^n\} \\ &\qquad \times p(z^n|\tilde{z}^n, \hat{w}_2^n, \hat{v}^n)p(y^n|\tilde{y}^n, \hat{w}_1^n, \hat{u}^n) \\ &= Q(x^n, u^n, w^n, v^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\qquad \times \mathbf{1}\{\hat{w}_1^n = \hat{w}_2^n = w^n, \hat{u}^n = u^n, \hat{v}^n = v^n\} \\ &\qquad \times p(z^n|\tilde{z}^n w^n v^n)p(y^n|\tilde{y}^n w^n u^n). \end{aligned} \tag{64}$$

Therefore,

$$\begin{aligned} &\hat{Q}(x^n, y^n, z^n, w^n, u^n, \hat{w}_1^n, v^n, \hat{w}_2^n, \hat{u}^n, \hat{v}^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\overset{\epsilon_0^{(n)}+\epsilon_1^{(n)}}{\approx} \\ &Q(x^n, y^n, z^n, w^n, u^n, v^n, g_{[0:2]}, \tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \omega) \\ &\quad \times \mathbf{1}\{\hat{w}_1^n = \hat{w}_2^n = w^n, \hat{u}^n = u^n, \hat{v}^n = v^n\}. \end{aligned} \tag{65}$$

Finally, using the second item in part 1 of [9, Lemma 3] we conclude that

$$\hat{Q}(g_{[0:2]}, x^n, y^n, z^n) \overset{\epsilon_0^{(n)}+\epsilon_1^{(n)}}{\approx} Q(g_{[0:2]}, x^n, y^n, z^n). \tag{66}$$

In particular, the marginal pmf of (X^n, Y^n, Z^n) of the right hand side of this expression is equal to $p(x^n, y^n, z^n)$, which is the desired distribution.

Part(3): In the above protocol we assumed that the sender and receivers have access to external randomnesses $G_{[0:2]}$ which are not present in the model. To eliminate these extra common randomnesses, we will fix particular instances $g_{[0:2]}$ of $G_{[0:2]}$ and show that the same protocol works even if we fix $G_{[0:2]} = g_{[0:2]}$. To prove this note that by letting $G_{[0:2]} = g_{[0:2]}$, the induced pmf $\hat{Q}(x^n, y^n, z^n)$ changes to the conditional pmf $\hat{Q}(x^n, y^n, z^n|g_{[0:2]})$. But if $G_{[0:2]}$ is almost independent of (X^n, Y^n, Z^n), the conditional pmf $\hat{Q}(x^n, y^n, z^n|g_{[0:2]})$ would be close to the desired distribution as well. To obtain the independence, we again use [9, Th. 1]. Substituting $T = 3$, $X_1 \leftarrow W$, $X_2 \leftarrow WU$, $X_3 \leftarrow WV$, and $Z \leftarrow XYZ$ in [9, Th. 1], we find that if

$$\begin{aligned} \tilde{R}_0 &< H(W|XYZ), \\ \tilde{R}_0 + \tilde{R}_1 &< H(WU|XYZ), \\ \tilde{R}_0 + \tilde{R}_2 &< H(WV|XYZ), \\ \tilde{R}_0 + \tilde{R}_1 + \tilde{R}_2 &< H(WUV|XYZ), \end{aligned} \tag{67}$$

then $Q(x^n, y^n, z^n, g_{[0:2]}) \overset{\epsilon_2^{(n)}}{\approx} p^{\mathsf{U}}(g_{[0:2]})p(x^n, y^n, z^n)$, for some asymptotically vanishing $\epsilon_2^{(n)}$. Thus, by triangular inequality for total variation, we have $\hat{Q}(x^n, y^n, z^n, g_{[0:2]}) \overset{\epsilon^{(n)}}{\approx} p^U(g_{[0:2]})p(x^n, y^n, z^n)$, where $\epsilon^{(n)} = \sum_{i=0}^{2} \epsilon_i^{(n)}$. From the definition of total variation distance random pmf's, the average of the total variation distance between $\hat{q}(x^n, y^n, z^n, g_{[0:2]})$ and $p^{\mathsf{U}}(g_{[0:2]})p(x^n, y^n, z^n)$ over all random binning is small. Thus, there exists a fixed binning with the corresponding pmf $\hat{q}$ such that $\hat{q}(x^n, y^n, z^n, g_{[0:2]}) \overset{\epsilon^{(n)}}{\approx} p^{\mathsf{U}}(g_{[0:2]})p(x^n, y^n, z^n)$. Next, [9, Lemma 3] guarantees the existance of an instance $g_{[0:2]}$ such that

$$\hat{q}(x^n, y^n, z^n|g_{[0:2]}) \overset{2\epsilon^{(n)}}{\approx} p(x^n, y^n, z^n).$$

Then the extra shared randomness $G_{[0:2]}$ can be eliminated by fixing it to be $G_{[0:2]} = g_{[0:2]}$.

Finally, observe that the rate region in the theorem is seen to be equivalent to that given by equations (58), (61) and (67) after eliminating $\tilde{R}_0, \tilde{R}_1, \tilde{R}_2$ using Fourier-Motzkin elimination. ■

E. An Infeasibility Result for Exact Channel Simulation

Proof of Theorem 10: Let Φ be a function that takes in an arbitrary discrete channel and returns a real number. Assume that $\Phi(p(y|x))$ that satisfies additivity and data processing properties, namely,

$$\Phi\big(\prod_{i=1}^{n} p(y_i|x_i)\big) = n\Phi\big(p(y|x)\big)$$

and

$$\Phi(p(y|x)) \geq \Phi(p(b|a)),$$

if $p(b|a) = \sum_{x,y} p(b|y)p(y|x)p(x|a)$ for some $p(x|a)$ *and* $p(b|y)$.

We claim that if $p(y|x)$ can be exactly simulated from $p(\tilde{y}|\tilde{x})$ with no shared randomness, we have $\Phi(p(y|x)) \geq \Phi(p(\tilde{y}|\tilde{x}))$. To show this assume that there is n and encoding and decoding maps which result in a joint pmf $q(x^n, y^n, \tilde{x}^n, \tilde{y}^n)$ such that (23) holds. We then have

$$\begin{aligned} n\Phi(p(y|x)) &= \Phi(p(y^n|x^n)) && (68)\\ &\geq \Phi(p(\tilde{y}^n|\tilde{x}^n)) && (69)\\ &= n\Phi(p(\tilde{y}|\tilde{x})), && (70)\end{aligned}$$

where (68) and (70) follow from the additivity of Φ for product channels, and (69) follows from the data processing property of Φ.

Next, assume that $\Phi(p(y|x))$ is also quasi-convex in $p(y|x)$. We claim that if $p(y|x)$ can be exactly simulated from $p(\tilde{y}|\tilde{x})$ with infinite shared randomness, we have $\Phi(p(y|x)) \geq \Phi(p(\tilde{y}|\tilde{x}))$. Assume that there is n and encoding and decoding maps which result in a joint pmf $q(x^n, y^n, \tilde{x}^n, \tilde{y}^n, \omega)$ such that (23) holds. Since $p(y^n|x^n) = \sum_\omega p(y^n|x^n, \omega)$, by quasi-convexity of Φ, there is some choice for $\omega = \omega^*$ such that

$$\Phi(p(y^n|x^n)) \geq \Phi(p(y^n|x^n, \omega^*)).$$

Then, using the fact that $p(\tilde{y}^n|\tilde{x}^n, \omega^*) = p(\tilde{y}^n|\tilde{x}^n)$, we can similarly write

$$\begin{aligned} n\Phi(p(y|x)) &= \Phi(p(y^n|x^n))\\ &\geq \Phi(p(y^n|x^n, \omega^*))\\ &\geq \Phi(p(\tilde{y}^n|\tilde{x}^n))\\ &= n\Phi(p(\tilde{y}|\tilde{x})). && (71)\end{aligned}$$

Note that $\Phi(p(y|x)) = C_\alpha(p(y|x))$ satisfies the additivity by Theorem 1, data processing by Theorem 2 and quasi-convexity by Lemma 1. This concludes the proof for capacity of order α.

It remains to show that $\Phi(p(y|x)) = \mathsf{Diam}_\alpha(p(y|x))$ satisfies the additivity, data processing and quasi-convexity properties:

Data processing: If $p(z|x) = \sum_y p(y|x)p(z|y)$, then by the data processing property of α-Rényi divergence we have

$$\begin{aligned} \mathsf{Diam}_\alpha(p(y|x)) &= \max_{p(x),q(x)} D_\alpha(p(y)\|q(y))\\ &\geq \max_{p(x),q(x)} D_\alpha(p(z)\|q(z))\\ &= \mathsf{Diam}_\alpha(p(z|x)).\end{aligned}$$

If $p(y|w) = \sum_x p(x|w)p(y|x)$, then any $p(w)$ and $q(w)$ correspond to some $p(x)$ and $q(x)$. Therefore,

$$\begin{aligned} \mathsf{Diam}_\alpha(p(y|w)) &= \max_{p(w),q(w)} D_\alpha(p(y)\|q(y))\\ &\leq \max_{p(x),q(x)} D_\alpha(p(y)\|q(y))\\ &= \mathsf{Diam}_\alpha(p(y|x)).\end{aligned}$$

Additivity: First, observe that by the quasi-convexity of D_α in its arguments (see e.g., [17]), we have

$$\begin{aligned} \mathsf{Diam}_\alpha(p(y|x)) &= \max_{p(x),q(x)} D_\alpha(p(y)\|q(y))\\ &= \max_{x_1,x_2} D_\alpha\big(p(y|x_1)\big\|p(y|x_2)\big). && (72)\end{aligned}$$

Therefore,

$$\begin{aligned} \mathsf{Diam}_\alpha(p(y^n|x^n)) &= \max_{x_1^n,x_2^n} D_\alpha\big(p(y^n|x_1^n)\big\|p(y^n|x_2^n)\big)\\ &= \max_{x_1^n,x_2^n} \sum_{i=1}^n D_\alpha\big(p(y_i|x_{1i})\big\|p(y_i|x_{2i})\big) && (73)\\ &= n\cdot\max_{x_1,x_2} D_\alpha\big(p(y|x_1)\big\|p(y|x_2)\big)\\ &= n\cdot\mathsf{Diam}_\alpha(p(y|x)),\end{aligned}$$

where in (73) we use the fact that $p(y^n|x_1^n) = \prod_{i=1}^n p(y_i|x_{1i})$ and similarly for $p(y^n|x_2^n)$.

Quasi-convexity: This follows from the quasi-convexity of D_α in its arguments (see e.g., [17]). If $p(y|x) = \sum_w p(y|x, w)p(w)$, then for every x_1, x_2, we have

$$D_\alpha\big(p(y|x_1)\big\|p(y|x_2)\big) \leq \max_w D_\alpha\big(p(y|x_1, w)\big\|p(y|x_2, w)\big).$$

Hence, characterization of channel diameter in (72), we have

$$\begin{aligned}&\max_{x_1,x_2} D_\alpha\big(p(y|x_1)\big\|p(y|x_2)\big)\\ &\quad\leq \max_w \max_{x_1,x_2} D_\alpha\big(p(y|x_1, w)\big\|p(y|x_2, w)\big).\end{aligned}$$

■

F. Exact Simulation of a BSC Channel From a BEC Channel

Proof of Theorem 11: By Theorem 10, BSC(p) can be exactly simulated from BEC(ϵ) with infinite shared randomness only if

$$C_\infty(\text{BEC}(\epsilon)) \geq C_\infty(\text{BSC}(p)).$$

Using equation (2), it is easy to verify that $C_\infty(\text{BEC}(\epsilon)) = \log(2-\epsilon)$ and

$$C_\infty(\text{BSC}(p)) = \log(2\max\{p, \bar{p}\}).$$

Thus, we should have $2\max(p, \bar{p}) \leq 2-\epsilon$. Since $p \in [0, 1/2]$, we get $2(1-p) \leq 2-\epsilon$, or $p \geq \epsilon/2$. On the other hand, a degradation strategy shows that any $p \geq \epsilon/2$ is achievable (without any need for shared randomness). This completes the proof. ■

G. Exact Simulation of a BIBO Channel From a BIBO Channel

Proof of Theorem 12:

Acheivability: We first show that any point (r, s) inside any of the two parallelograms is achievable. The parallelogram with vertices $\{(r, s), (\bar{r}, \bar{s}), (0, 0), (1, 1)\}$ is achievable as follows: fix $\tilde{X} = X$. Then there are decoder strategies for

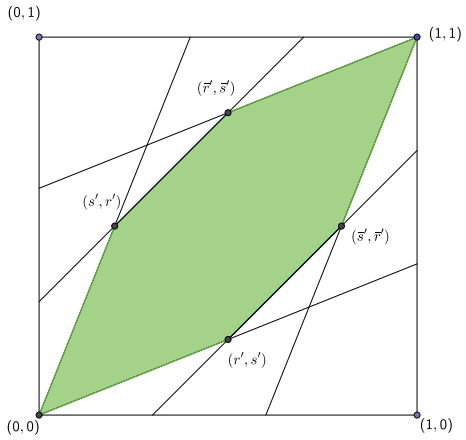

Fig. 10. The boundary of the simulation region for the exact channel simulation of BIBO channel from another BIBO channel.

achieving any of these four vertices of the parallelogram if we use $Y = \tilde{Y}$, $Y = 1 - \tilde{Y}$, $Y = 0$ and $Y = 1$. The whole parallelogram is achievable by time-sharing between these vertices using *private* randomness at the decoder. The parallelogram with vertices $\{(s, r), (\bar{s}, \bar{r}), (0, 0), (1, 1)\}$ is achievable in a similar way if we fix $\tilde{X} = 1 - X$ instead.

Thus, if shared randomness is not available, the union of the two parallelograms is achievable. If shared randomness is available, the convex hull of the region, which is the convex polygon is achievable.

Converse: It suffices to prove the converse for the case of infinite shared randomness. It is clear that the simulation region when there is no shared randomness cannot exceed that when shared randomness exists.

By (2) for a BIBO channel $p(y|x)$ with parameters (r, s) we have

$$\begin{aligned} C_\infty(p(y|x)) &= \log(\max\{r, s\} + \max\{\bar{r}, \bar{s}\}) \\ &= \log\left(\max\{r + \bar{s}, s + \bar{r}\}\right). \end{aligned}$$

Thus, by Theorem 10, the possibility of simulation gives

$$\max\{r + \bar{s}, s + \bar{r}\} \le \max\{r' + \bar{s}', s' + \bar{r}'\},$$

or $\max\{r - s, s - r\} \le \max\{r' - s', s' - r'\}$. Equivalently, we have

$$|r - s| \le |r' - s'|. \tag{74}$$

Similarly, using (72), for such a channel $p(y|x)$ we have

$$\begin{aligned} \mathsf{Diam}_\infty(p(y|x)) &= \max_{x_1, x_2} D_\infty(p(y|x_1) \| q(y|x_2)) \\ &= \max\left\{D_\infty((r, \bar{r}) \| (s, \bar{s})), D_\infty((s, \bar{s}) \| (r, \bar{r}))\right\} \\ &= \log \max\left\{\frac{r}{s}, \frac{\bar{r}}{\bar{s}}, \frac{s}{r}, \frac{\bar{s}}{\bar{r}}\right\}. \end{aligned}$$

Therefore, again by Theorem 10, the possibility of channel simulation gives

$$\max\left\{\frac{r}{s}, \frac{s}{r}, \frac{\bar{r}}{\bar{s}}, \frac{\bar{s}}{\bar{r}}\right\} \le \max\left\{\frac{r'}{s'}, \frac{s'}{r'}, \frac{\bar{r}'}{\bar{s}'}, \frac{\bar{s}'}{\bar{r}'}\right\}. \tag{75}$$

Equations (74) and (75) imply that (r, s) is in the area depicted in Fig. 10 for the given pair (r', s'). In particular, equation (74) gives the two edges that are parallel to the line $r = s$, and equation (75) gives the four side boundaries of the region (see Fig. 10). This completes the proof. ■

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewer for valuable comments and suggestions to improve the quality of the paper.

REFERENCES

[1] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002.

[2] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1393–1397.

[3] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2926–2959, May 2014.

[4] P. Cuff and C. Schieler, "Hybrid codes needed for coordination over the point-to-point channel," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Sep. 2011, pp. 235–239.

[5] G. R. Kumar, C. T. Li, and A. El Gamal, "Exact common information," in *Proc. IEEE Symp. Inf. Theory (ISIT)*, Jul. 2014, pp. 161–165.

[6] M. Abroshan, A. Gohari, and S. Jaggi. (May 2015). "Zero error coordination." [Online]. Available: https://arxiv.org/abs/1505.01110

[7] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 3053–3057.

[8] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, "Zero-error channel capacity and simulation assisted by non-local correlations," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5509–5523, Aug. 2011.

[9] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 6760–6786, Nov. 2014.

[10] P. Cuff, "Communication in networks for coordinating behavior," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ. Serra Mall, Stanford, CA, USA, 2009.

[11] R. Sibson, "Information radius," *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, vol. 14, no. 2, pp. 149–160, 1969.

[12] S. Verdú, "α-mutual information," in *Proc. Inf. Theory Appl. Workshop*, 2015. [Online]. Available: http://ita.ucsd.edu/workshop/15/files/paper/paper_374.pdf

[13] S. Arimoto, "Information measures and capacity of order for discrete memoryless channels," in *Proc. Colloq. Math. Topics Inf. Theory Soc.*, Keszthely, Hungary, 1975, pp. 41–52.

[14] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proc. 48th Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2010, pp. 1327–1333.

[15] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[16] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Feb. 1975.

[17] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.

[18] D. L. Neuhoff and P. C. Shields, "Channels with almost finite memory," *IEEE Trans. Inf. Theory*, vol. 25, no. 4, pp. 440–447, Apr. 1979.

[19] D. L. Neuhoff and P. C. Shields, "Channel entropy and primitive approximation," *Ann. Probab.*, vol 10, no. 1, pp. 188–198, Jan. 1982.

[20] Y. Steinberg and S. Verdú, "Channel simulation and coding with side information," *IEEE Trans. Inf. Theory*, vol. 40, no. 3, pp. 634–646, Mar. 1994.

[21] Y. Altug and A. B. Wagner, "Source and channel simulation using arbitrary randomness," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1345–1360, Mar. 2012.

[22] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6351–6355, Oct. 2011.

[23] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.

[24] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.
[25] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inf. Control*, vol. 10, no. 1, pp. 65–103, Jan. 1967.
[26] C. E. Shannon, "A note on a partial ordering for communication channels," *Inf. Control*, vol. 1, no. 4, pp. 390–397, 1958.
[27] H. J. Helgert, "A partial ordering of discrete, memoryless channels," *IEEE Trans. Inf. Theory*, vol. 13, no. 3, pp. 360–365, 1967.

Farzin Haddadpour received the B.Sc. degree in electrical engineering from University of Tabriz, Tabriz, Iran, in 2010 and his M.Sc. degree from Sharif University of Technology, Tehran, Iran in 2012, respectively. He is currently working towards the Ph.D. degree in the Department of Electrical Engineering and Computer Science, Pennsylvania State University, State College, PA. He was awarded the Trust scholarship from the Cambridge University in 2014. He has worked in the areas of Network information Theory with special focus on communication rates for having coordinated behavior.

Mohammad Hossein Yassaee is a postdoctoral research associate at the electrical engineering department, Princeton university. He received his Ph.D. from Sharif University of Technology in Dec., 2014. His research interests include network information theory and non-asymptotic information theory. He is a recipient of the 2013 Jack Keil Wolf ISIT Student Paper Award. He also was a finalist for the 2012 ISIT Student Paper Award.

Salman Beigi received his BS at Sharif University of Technology, Tehran in 2004. He finished his Ph.D. at the MIT Math Department in 2009 under the supervision of Peter Shor, and continued his research as a postdoc at the Institute for Quantum Information at Caltech. He is now a researcher at the Institute for Research in Fundamental Sciences (IPM), Tehran. His interests include quantum complexity theory, quantum coding theory, zeroerror channel capacity and quantum information theory.

Amin Gohari is an associate professor at Sharif University of Technology, Tehran, Iran. Dr. Gohari received his M.Sc. and Ph.D. degree in electrical engineering in 2010 from the University of California, Berkeley, and his B.Sc. degree in 2004 from Sharif University of Technology, Iran. He received the 2010 Eli Jury Award from UC Berkeley, Department of Electrical Engineering, for "outstanding achievement in the area of communication networks," and the 2009-2010 Bernard Friedman Memorial Prize in Applied Mathematics from UC Berkeley, Department of Mathematics, for "demonstrated ability to do research in applied mathematics." He also received the Gold Medal from the 41st International Mathematical Olympiad (IMO 2000) and the First Prize from the 9th International Mathematical Competition for University Students (IMC 2002). He is also a co-author of a paper that received the 2013 Jack Keil Wolf ISIT Student Paper Award. He was selected as an exemplary reviewer for *Transactions on Communications* in 2016.

Mohammad Reza Aref received the B.Sc. degree in 1975 from the University of Tehran, Iran, and the M.Sc. and Ph.D. degrees in 1976 and 1980, respectively, from Stanford University, Stanford, CA, USA, all in electrical engineering. He returned to Iran in 1980 and was actively engaged in academic affairs. He was a Faculty member of Isfahan University of Technology from 1982 to 1995. He has been a Professor of electrical engineering at Sharif University of Technology, Tehran, since 1995, and has published more than 270 technical papers in communication and information theory and cryptography in international journals and conferences proceedings. His current research interests include areas of communication theory, information theory, and cryptography.