# Improving on the Cut-Set Bound via Geometric Analysis of Typical Sets

Xiugang Wu, *Member, IEEE*, Ayfer Özgür, *Member, IEEE*, and Liang-Liang Xie, *Fellow, IEEE*

***Abstract*—We consider the discrete memoryless symmetric primitive relay channel, where, a source $X$ wants to send information to a destination $Y$ with the help of a relay $Z$ and the relay can communicate to the destination via an error-free digital link of rate $R_0$, while $Y$ and $Z$ are conditionally independent and identically distributed given $X$. We develop two new upper bounds on the capacity of this channel that are tighter than existing bounds, including the celebrated cut-set bound. Our approach significantly deviates from the standard information-theoretic approach for proving upper bounds on the capacity of multi-user channels. We build on the blowing-up lemma to analyze the probabilistic geometric relations between the typical sets of the $n$-letter random variables associated with a reliable code for communicating over this channel. These relations translate to new entropy inequalities between the $n$-letter random variables involved. As an application of our bounds, we study an open question posed by (Cover, 1987), namely, what is the minimum rate $R_0^*$ needed for the $Z$–$Y$ link in order for the capacity of the relay channel to be equal to that of the broadcast cut. We consider the special case when the $X$–$Y$ and $X$–$Z$ links are both binary symmetric channels. Our tighter bounds on the capacity of the relay channel immediately translate to tighter lower bounds for $R_0^*$. More interestingly, we show that when $p \to 1/2$, $R_0^* \geq 0.1803$; even though the broadcast channel becomes completely noisy as $p \to 1/2$ and its capacity, and therefore the capacity of the relay channel, goes to zero, a strictly positive rate $R_0$ is required for the relay channel capacity to be equal to the broadcast bound. Existing upper bounds on the capacity of the relay channel, and the cut-set bound in particular, would rather imply $R_0^* \to 0$, while achievability schemes require $R_0^* \to 1$. We conjecture that $R_0^* \to 1$ as $p \to 1/2$.**

***Index Terms*—Relay channel, cut-set bound, converse, information inequality, geometry of typical set.**

## I. INTRODUCTION

CHARACTERIZING the capacity of relay channels [3] has been a long-standing open problem in network information theory. The seminal work of Cover and El Gamal [4] has introduced two basic achievability schemes: Decode-and-Forward and Compress-and-Forward, and derived a general upper bound on the capacity of this channel, now known as the cut-set bound. Over the last decade, significant progress has been made on the achievability side: these schemes have been extended and unified to multi-relay networks [5]–[7] and many new relaying strategies have been discovered, such as Amplify-and-Forward, Quantize-Map-and-Forward, Compute-and-Forward, Noisy Network Coding, Hybrid Coding etc [8]–[12]. However, the progress on developing upper bounds that are tighter than the cut-set bound has been relatively limited. In particular, in most of the special cases where the capacity is known, the converse is given by the cut-set bound [4], [13]–[15].

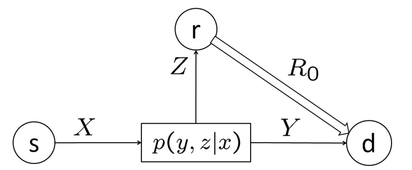


Fig. 1. Primitive relay channel.

In general, however, the cut-set bound is known to be not tight. Specifically, consider the primitive relay channel depicted in Fig. 1, where the source's input $X$ is received by the relay $Z$ and the destination $Y$ through a channel $p(y,z|x)$, and the relay $Z$ can communicate to the destination $Y$ via an error-free digital link of rate $R_0$. When $Y$ and $Z$ are conditionally independent given $X$, and $Y$ is a stochastically degraded version of $Z$, Zhang [16] uses the blowing-up lemma [17] to show that the capacity can be strictly smaller than the cut-set bound in certain regimes of this channel. However, Zhang's result does not provide any information regarding the gap or suggest a way to compute it. For a special case of the primitive relay channel where the noise for the $X$-$Y$ link is modulo additive and $Z$ is a corrupted version of this noise, Aleksic, Razaghi and Yu characterize the capacity and show that it is strictly lower than the cut-set bound [18]. While this result provides an exact capacity characterization for a non-trivial special case, it builds strongly on the peculiarity of the channel model and in this respect its scope is more limited than Zhang's result.

More recently, a new upper bound demonstrating an explicit gap to the cut-set bound was developed by Xue [19] for general primitive relay channels. In Xue's bound, the gap to the cut-set bound is related to the reliability function of the $X$-$Y$ link. In particular, it builds on the blowing-up lemma to obtain a lower bound on the successful decoding probability based only

Manuscript received February 26, 2016; revised November 20, 2016; accepted January 6, 2017. Date of publication February 6, 2017; date of current version March 15, 2017. This work was supported in part by NSF under Award CIF-1514538 and in part by the Center for Science of Information, an NSF Science and Technology Center, under Grant CCF-0939370. This paper was presented in part at the 2015 IEEE International Symposium on Information Theory [1] and the 2016 International Zurich Seminar on Communications [2].

X. Wu and A. Özgür are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: x23wu@stanford.edu; aozgur@stanford.edu).

L.-L. Xie is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: llxie@uwaterloo.ca).

Communicated by J. Chen, Associate Editor for Shannon Theory.

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.2017.2664811

0018-9448 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

on $Y$ and then compares this lower bound to the reliability function of the single-user channel $X$-$Y$. Unlike Zhang's result, Xue's bound identifies an explicit gap to the cut-set bound that can be numerically computed. However, compared to the cut-set bound, Xue's bound only considers the information flow from the source and the relay to the destination (the multiple-access cut) and does not bound the flow from the source to the relay and the destination (the broadcast cut). Therefore it can be looser than the cut-set bound since it does not capture the inherent trade-off between maximizing the information flow across these two different cuts of the network.

In this paper, we present two new upper bounds on the capacity of the primitive relay channel that are tighter than the cut-set bound. To simplify exposition, we concentrate on the symmetric case ($Y$ and $Z$ are conditionally independent and identically distributed given $X$) in this paper, however our results can be extended to the asymmetric case via channel simulation arguments [32]. Just like Zhang and Xue, we critically build on the blowing up lemma, however we develop novel ways for utilizing it which lead to simpler arguments and tighter results. In general, proving an upper bound on the capacity of a multi-user channel involves dealing with entropy relations between the various $n$-letter random variables induced by the reliable code and the channel structure (together with using Fano's inequality). In order to prove the desired relations between the entropies of the $n$-letter random variables involved, in this paper we consider their $B$-length i.i.d. extensions (leading to length $B$ i.i.d. sequences of $n$-letter random variables). We then use the blowing-up lemma to analyze the geometry of the typical sets associated with these $B$-length sequences. The key step in our development is to translate the (probabilistic) geometric relations between these typical sets into new entropy relations between the random variables involved. While both of our bounds are based on this same approach, they use different arguments to translate the geometry of the typical sets to entropy relations, and eventually lead to two different bounds on the capacity of the channel which do not include each other in general.

As an application of our bounds, we consider the binary symmetric channel, i.e., we assume both the $X$-$Y$ and $X$-$Z$ links are binary symmetric channels with crossover probability, say, $p$. We demonstrate that both our bounds perform strictly better than the cut-set bound and Xue's bound, and particularly, our second bound provides considerable gain over these earlier bounds. We then use our bounds to investigate an open question posed by Cover [20] which asks for the minimum required $Z$-$Y$ link rate $R_0^*$ in order for the capacity of the relay channel to be equal to the capacity of the broadcast cut, i.e. $\max_{p(x)} I(X; Y, Z)$. Obviously as $R_0^*$ becomes larger the capacity of the relay channel does approach the capacity of the broadcast cut. For example, in the binary symmetric case if $R_0 = 1$, the relay can convey its noisy observation as it is to the destinaton, therefore the broadcast cut capacity is trivially achievable. In this sense, Cover's open problem asks how smaller $R_0$ can be made than 1 without decreasing the capacity of the relay channel. Interestingly, there is a striking dichotomy between the currently available upper and lower bounds for $R_0^*$ when $p \to 1/2$, i.e. when the broadcast channel becomes completely noisy and its capacity, and therefore the capacity of the relay channel, goes to zero. Achievability schemes, Hash-and-Forward in particular, require $R_0 \to 1$ even though the capacity itself tends to zero. The cut-set bound and Xue's bound, on the other hand, require $R_0 \to 0$ in order for the capacity to be equal to the broadcast capacity. By strengthening our second bound in this specific case, we show that $R_0^* \geq 0.1803$; indeed a strictly positive rate $R_0$ is needed in order to achieve the vanishing broadcast capacity. We conjecture that $R_0^* \to 1$ when $p \to 1/2$; to achieve the broadcast capacity the relay has no choice but to forward its observation, which is almost pure noise, as it is to the destination.[1]

### A. Organization of the Paper

The remainder of the paper is organized as follows. Section II and III introduces the channel model and reviews the existing upper bounds for primitive relay channels, respectively. Section IV discusses our new upper bounds for symmetric primitive relay channels in detail, followed by a treatment on the binary symmetric channel case in Section V. Sections VI–IX are then dedicated to the proofs of our bounds. Finally, some concluding remarks are included in Section X.

## II. Channel Model

Consider a primitive relay channel as depicted in Fig. 1. The source's input $X$ is received by the relay $Z$ and the destination $Y$ through a channel

$$(\Omega_X, p(y, z|x), \Omega_Y \times \Omega_Z)$$

where $\Omega_X, \Omega_Y$ and $\Omega_Z$ are finite sets denoting the alphabets of the source, the destination and the relay, respectively, and $p(y, z|x)$ is the channel transition probability; the relay $Z$ can communicate to the destination $Y$ via an error-free digital link of rate $R_0$.

For this channel, a code of rate $R$ for $n$ channel uses, denoted by

$$(\mathcal{C}_{(n,R)}, f_n(z^n), g_n(y^n, f_n(z^n))), \text{ or simply, } (\mathcal{C}_{(n,R)}, f_n, g_n),$$

consists of the following:

1) A codebook at the source $X$,
$$\mathcal{C}_{(n,R)} = \{x^n(m) \in \Omega_X^n, m \in \{1, 2, \ldots, 2^{nR}\}\};$$
2) An encoding function at the relay $Z$,
$$f_n : \Omega_Z^n \to \{1, 2, \ldots, 2^{nR_0}\};$$
3) A decoding function at the destination $Y$,
$$g_n : \Omega_Y^n \times \{1, 2, \ldots, 2^{nR_0}\} \to \{1, 2, \ldots, 2^{nR}\}.$$

The average probability of error of the code is defined as

$$P_e^{(n)} = \Pr(g_n(Y^n, f_n(Z^n)) \neq M),$$

[1] Recently we have solved Cover's problem in the Gaussian case [21]. In particular, we show that for a symmetric Gaussian primitive relay channel, the broadcast capacity can be achieved if and only if the relay–destination link is of infinite capacity.

where the message $M$ is assumed to be uniformly drawn from the message set $\{1, 2, \ldots, 2^{nR}\}$. A rate $R$ is said to be achievable if there exists a sequence of codes

$$\{(\mathcal{C}_{(n,R)}, f_n, g_n)\}_{n=1}^{\infty}$$

such that the average probability of error $P_e^{(n)} \to 0$ as $n \to \infty$.

The capacity of the primitive relay channel is the supremum of all achievable rates, denoted by $C(R_0)$. Also, denote by $C_{XY}, C_{XZ}$ and $C_{XYZ}$ the capacities of the channels $X$-$Y$, $X$-$Z$, and $X$-$YZ$, respectively. Obviously, we have $C(0) = C_{XY}$ and $C(\infty) = C_{XYZ}$.

### A. Symmetric Primitive Relay Channel

In this paper, we focus on the symmetric case of the primitive relay channel, that is, when $Y$ and $Z$ are conditionally independent and identically distributed given $X$. Formally, a primitive relay channel is said to be symmetric if

1) $p(y, z|x) = p(y|x)p(z|x)$,
2) $\Omega_Y = \Omega_Z := \Omega$, and $\Pr(Y = \omega|X = x) = \Pr(Z = \omega|X = x)$ for any $\omega \in \Omega$ and $x \in \Omega_X$.

In this case, we also use $p(\omega|x)$ to denote the transition probability of both the $X$-$Y$ and $X$-$Z$ channels.

## III. Existing Upper Bounds for Primitive Relay Channels

For general primitive relay channels, the well-known cut-set bound can be stated as follows.

*Proposition 1 (Cut-Set Bound):* For the general primitive relay channel, if a rate $R$ is achievable, then there exists some $p(x)$ such that

$$\begin{cases} R \le I(X; Y, Z) & (1) \\ R \le I(X; Y) + R_0. & (2) \end{cases}$$

Inequalities (1) and (2) are generally known as the broadcast bound and multiple-access bound, since they correspond to the broadcast channel $X$-$YZ$ and multiple-access channel $XZ$-$Y$, respectively.

Note that although the cut-set bound in (1)–(2) is tight for most of the cases where the capacity is determined [4], [13]–[15], it is known to be not tight in general. The first counterexample was given by Zhang in [16], where he considered a class of stochastically degraded primitive relay channels. Using the blowing-up lemma [17], he showed that the capacity of the channel can be strictly smaller than the cut-set bound.

*Proposition 2 (Zhang [16]):* For a primitive relay channel, if $Y$ and $Z$ are conditionally independent given $X$, and $Y$ is a stochastically degraded version of $Z$ (i.e. there exists some $q(y|z)$ such that $p(y|x) = \sum_z p(z|x)q(y|z)$), then the capacity $C(R_0)$ of the channel satisfies

$$C(R_0) < C_{XY} + R_0 \tag{3}$$

when

$$R_0 > \max_{p(x): I(X;Y)=C_{XY}} I(X; Z) - C_{XY}. \tag{4}$$

In the regime where $R_0$ satisfies both (4) and the condition

$$C_{XY} + R_0 < \max_{p(x): I(X;Y)=C_{XY}} I(X; Y, Z),$$

the cut-set bound becomes $C_{XY} + R_0$, however the strictness of the inequality in (3) implies that the cut-set bound is loose with some positive gap. Roughly speaking this corresponds to the regime where the cut-set bound is limited by the multiple-access bound but the source–relay channel is not strong enough to enable the relay to trivially decode the transmitted message. However, note that Zhang's result does not provide any information about how large the gap can be.

Recently, a new upper bound demonstrating an explicit gap to the cut-set bound was developed by Xue [19]. This new bound was first established for the symmetric case, and then extended to the general asymmetric case employing channel simulation theory [22], [23]. The proof uses a generalized version of the blowing-up lemma [19] to characterize the successful decoding probability based only on $Y$ and then compares it with the reliability function for the single-user channel $X$-$Y$. Xue's bound specialized for the symmetric case is given as follows.

*Proposition 3 (Xue's Bound):* For the symmetric primitive relay channel, if a rate $R$ is achievable, then there exists some $a \in [0, R_0]$ such that

$$\begin{cases} R \le C_{XY} + R_0 - a & (5) \\ E(R) \le H_2(\sqrt{a}) + \sqrt{a} \log |\Omega| & (6) \end{cases}$$

where $H_2(r)$ is the binary entropy function and $E(R)$ is the reliability function for the $X$-$Y$ link defined as

$$E(R) := \max_{\rho \in [-1,0)} (-\rho R + \min_{p(x)} E_0(\rho, p(x))) \tag{7}$$

with

$$E_0(\rho, p(x)) := -\log \left[ \sum_y \left( \sum_x p(x) p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right].$$

It can be seen that Xue's bound modifies the original multiple-access bound (2) by introducing an additional term "$-a$" in (5), where "$a$" is a non-negative auxiliary variable subject to the constraint (6). As noted in [19], this implies that the capacity of the symmetric primitive relay channel is strictly less than $C_{XY} + R_0$ for any $R_0 > 0$. To see this, consider any rate $R > C_{XY}$. Then it follows from [24] that $E(R) > 0$, which forces $a$ to be strictly positive in light of (6), and thus $R < C_{XY} + R_0$ by (5). Since $a$ here is numerically computable, Xue's bound improves over Zhang's result as it provides an explicit lower bound on the gap between the capacity and the cut-set bound.

Nevertheless, Xue's bound has its own limitations: i) compared to the cut-set bound, it lacks a constraint on the broadcast cut and therefore decouples the information flow over the broadcast and multiple-access cuts of the channel (note that it can be always amended by including the bound $R \le C_{XYZ}$, however with such a straightforward amendment the bound $R \le C_{XYZ}$ would have no coupling with those in (5) and (6), which can be potentially coupled through $p(x)$ as done in the cut-set bound); ii) there is no coupling between (5) and (6) which can also benefit from a coupling through the input distribution $p(x)$. Our bounds presented in the next section overcome these limitations and further improve on Xue's bound. They are also structurally different from Xue's bound as they involve only classical information measures and do not involve the reliability function.

## IV. New Upper Bounds for Symmetric Primitive Relay Channels

This section presents two new upper bounds on the capacity of symmetric primitive relay channels that are generally tighter than the cut-set bound. Before stating our main theorems, in the following subsection we first explain the relation of our new bounds to the cut-set bound.

### A. Improving on the Cut-Set Bound

Let the relay's transmission be denoted by $I_n = f_n(Z^n)$. Let us recall the derivation of the cut-set bound. The first step in deriving (1)–(2) is to use Fano's inequality to conclude that

$$nR \leq I(X^n; Y^n, I_n) + n\epsilon.$$

We can then either proceed as

$$\begin{aligned} nR &\leq I(X^n; Y^n, I_n) + n\epsilon \\ &\leq I(X^n; Y^n, Z^n) + n\epsilon \\ &\leq nI(X; Y, Z) + n\epsilon \end{aligned}$$

to obtain the broadcast bound (1), where the second inequality follows from the data processing inequality and the single letterization in the third line can be either done with a time-sharing or fixed composition code argument[2]; or we can proceed as

$$\begin{aligned} nR &\leq I(X^n; Y^n, I_n) + n\epsilon \\ &\leq I(X^n; Y^n) + H(I_n|Y^n) - H(I_n|X^n) + n\epsilon \qquad (8) \\ &\leq nI(X; Y) + nR_0 + n\epsilon \qquad (9) \end{aligned}$$

to obtain the multiple-access bound (2), where to obtain the last inequality we upper bound $H(I_n|Y^n)$ by $nR_0$ and use the fact that $H(I_n|X^n)$ is non-negative.

Instead of simply lower bounding $H(I_n|X^n)$ by 0 in the last step, our bounds presented in the next two subsections are based on letting $H(I_n|X^n) = na_n$ and proving a third inequality that forces $a_n$ to be strictly non-zero. This new inequality is based on capturing the tension between how large $H(I_n|Y^n)$ and how small $H(I_n|X^n)$ can be. Intuitively, it is easy to see this tension. Specifically, suppose $H(I_n|X^n) \approx 0$, then roughly speaking, this implies that given the transmitted codeword $X^n$, there is no ambiguity about $I_n$, or equivalently, all the $Z^n$ sequences jointly typical with $X^n$ are mapped to the same $I_n$. Since $Y^n$ and $Z^n$ are statistically equivalent given $X^n$ (they share the same typical set given $X^n$) this would further imply that $I_n$ can be determined based on $Y^n$, and therefore $H(I_n|Y^n) \approx 0$. This would force the rate to be even smaller than $I(X; Y)$.

Equivalently, rewriting (8) as

$$R \leq nI(X; Y) + I(I_n; X^n) - I(I_n; Y^n) + n\epsilon, \qquad (10)$$

our approach can be thought of as fixing the first $n$-letter mutual information to be

$$I(I_n; X^n) = H(I_n) - na_n$$

and controlling the second $n$-letter mutual information $I(I_n; Y^n)$. In doing so, we only build on the Markov chain structure $I_n - Z^n - X^n - Y^n$ and the fact that $Z^n$ and $Y^n$ are conditionally i.i.d. given $X^n$. In particular, we do not employ the fact that these random variables are associated with a reliable code. Note that this approach of directly studying the relation between the $n$-letter information measures involved significantly deviates from the standard approach in network information theory for developing converses, where one usually seeks to single letterize such $n$-letter expressions.

More precisely, we proceed as follows. We fix $H(I_n|X^n) = na_n$ and leave this term as it is in (8), yielding

$$R \leq I(X; Y) + R_0 - a_n + \epsilon.$$

We then prove the following two upper bounds (11) and (12) on $I(I_n; X^n) - I(I_n; Y^n)$ in terms of $a_n$:

$$I(I_n; X^n) - I(I_n; Y^n) \leq nV\left(\sqrt{\frac{a_n \ln 2}{2}}\right) - na_n, \qquad (11)$$

where

$$V(r) := \begin{cases} \log|\Omega| & \text{if } r > \frac{|\Omega|-1}{|\Omega|} \\ H_2(r) + r\log(|\Omega|-1) & \text{if } r \leq \frac{|\Omega|-1}{|\Omega|}; \end{cases}$$

and

$$I(I_n; X^n) - I(I_n; Y^n) \leq n\Delta\left(p(x), \sqrt{\frac{a_n \ln 2}{2}}\right), \qquad (12)$$

where $\Delta\left(p(x), \sqrt{\frac{a_n \ln 2}{2}}\right)$ is a quantity that depends on the distribution $p(x)$ and $a_n$, which we will formally define in Section IV-C. These two bounds are obtained via bounding $H(I_n|Y^n)$ and $H(Y^n|I_n)$ in terms of $a_n$ respectively, and combined with (10) they immediately yield new constraints on $R$.

The heart of our argument is therefore to prove the two bounds in (11) and (12). To accomplish this, we suggest a new set of proof techniques. In particular, we look at the $B$-letter i.i.d. extensions of the random variables $X^n, Y^n, Z^n$ and $I_n$ and study the geometric relations between their typical sets by using the generalized blowing-up lemma. While we use this same general approach for obtaining (11) and (12), we build on different arguments in each case, which eventually leads to two different bounds on the capacity of the relay channel that do not include each other in general.

### B. Via Bounding $H(I_n|Y^n)$

Our first bound builds on bounding $H(I_n|Y^n)$ and it is given by the following theorem.

*Theorem 4:* For the symmetric primitive relay channel, if a rate $R$ is achievable, then there exists some $p(x)$ and

$$a \in \left[0, \min\left\{R_0, H(Z|X), \frac{2}{\ln 2}\left(\frac{|\Omega|-1}{|\Omega|}\right)^2\right\}\right] \qquad (13)$$

[2]Note that the time-sharing or the fixed composition code argument for single letterization is needed to preserve the coupling to the second inequality in (9) via $X$.

such that

$$\begin{cases} R \le I(X;Y,Z) & (14) \\ R \le I(X;Y) + R_0 - a & (15) \\ R \le I(X;Y) + H_2\left(\sqrt{\frac{a \ln 2}{2}}\right) + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1) - a. & (16) \end{cases}$$

Clearly our bound in Theorem 4 implies the cut-set bound in Proposition 1. In fact, it can be checked that our bound is *strictly* tighter than the cut-set bound for any $R_0 > 0$. For this, note that (15) will reduce to (2) only if $a = 0$; however, if $a = 0$ then (16) will constrain $R$ by the rate $I(X;Y)$ which is lower than the cut-set bound.

Our bound is also generally tighter than Xue's bound, and since Xue's bound implies Zhang's result [16], so does our bound. In particular, our bound overcomes the limitations of Xue's bound that are observed in Section III, and furthermore tightens the constraint (6) on $a$ to (16). By contrasting (14)–(16) to (5)–(6), we note the following improvements:

1) Our bound introduces the missing constraint on the broadcast cut (14) and couples it with (15)–(16) through the input distribution $p(x)$.
2) The term $C_{XY}$ in (5) is replaced by $I(X;Y)$ in (15). Since the distribution $p(x)$ in Theorem 4 has to be chosen to satisfy all the constraints (14)–(16), it may not necessarily maximize $I(X;Y)$, and thus (15) is in general stricter than (5).
3) The constraint (6) on $a$ is replaced by (16). To show that the latter is stricter, rewrite it as

$$\begin{aligned} &R - I(X;Y) \\ &\le H_2\left(\sqrt{\frac{a \ln 2}{2}}\right) + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1) - a. \end{aligned} \tag{17}$$

Note that (6) is active only if $R > C_{XY}$. In Appendix A we show that in this case the L.H.S. (left-hand-side) of (17) is generally greater than that of (6), while the the R.H.S. (right-hand-side) of (17) is obviously less than that of (6) for any $a > 0$. Therefore, the constraint (16) is also stricter than (6).

A simple example demonstrating the above improvements is given in Appendix B. The improvements 1) and 2) come from fixed composition code analysis [25] (or alternatively a time-sharing argument), while the key to improvement 3), which accounts for the structural change from (6) to (16), is a new argument for bounding $H(I_n|Y^n)$ instead of analyzing the successful decoding probability based only on $Y$ as done in [19].

### C. Via Bounding $H(Y^n|I_n)$

Before presenting our second upper bound, we first define a parameter that will be used in stating the theorem.

*Definition 5:* Given a fixed channel transition probability $p(\omega|x)$, for any $p(x)$ and $d \ge 0$, $\Delta(p(x), d)$ is defined as

$$\begin{aligned} &\Delta(p(x), d) \\ &:= \max_{\tilde{p}(\omega|x)} H(\tilde{p}(\omega|x)|p(x)) + D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) \\ &\quad - H(p(\omega|x)|p(x)) \end{aligned} \tag{18}$$

$$\text{s.t. } \frac{1}{2} \sum_{(x,\omega)} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \le d. \tag{19}$$

In the above, we adopt the notation in [26]. Specifically, $D(\tilde{p}(\omega|x)||p(\omega|x)|p(x))$ is the conditional relative entropy defined as

$$D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) := \sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \frac{\tilde{p}(\omega|x)}{p(\omega|x)}, \tag{20}$$

$H(\tilde{p}(\omega|x)|p(x))$ is the conditional entropy defined with respect to the joint distribution $p(x)\tilde{p}(\omega|x)$, i.e.,

$$H(\tilde{p}(\omega|x)|p(x)) := -\sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \tilde{p}(\omega|x), \tag{21}$$

and $H(p(\omega|x)|p(x))$ is the conditional entropy similarly defined with respect to $p(x)p(\omega|x)$.

$\Delta(p(x), d)$ can be interpreted as follows: given a random variable $X \sim p(x)$, assume we want to describe a related random variable $Y$. We use a code designed for the conditional distribution $p(w|x)$ for $Y$ given $X$, while $Y$ actually comes from a distribution $\tilde{p}(w|x)$. The distribution $\tilde{p}(w|x)$ cannot be too different than the assumed distribution in the sense that the total variation distance between the two joint distributions $p(x)p(w|x)$ and $p(x)\tilde{p}(w|x)$ is bounded by $d$. $\Delta(p(x), d)$ captures the maximal inefficiency we would incur for having $Y$ come from a different distribution than the one assumed, i.e. it is the maximal number of extra bits we would use when compared to the case where $Y$ comes from the assumed distribution.

It can be easily seen that $\Delta(p(x), d) \ge 0$ for all $p(x)$ and $d \ge 0$, and $\Delta(p(x), d) = 0$ when $d = 0$. Moreover, for any fixed $p(x)$ and $d > 0$, $\Delta(p(x), d) = \infty$ if and only if there exists some $x$ with $p(x) > 0$, and some $\omega$ such that $p(\omega|x) = 0$. Thus, a sufficient condition for $\Delta(p(x), d) < \infty$ for all $p(x)$ and $d > 0$ is that the channel transition matrix is *fully connected*, i.e., $p(\omega|x) > 0, \forall (x, \omega) \in \Omega_X \times \Omega$. In this case, $\Delta(p(x), d) \to 0$ as $d \to 0$ for any $p(x)$.

*Example 6:* Suppose $p(\omega|x)$ corresponds to a binary symmetric channel with crossover probability $p < 1/2$. We derive $\Delta(p(x), d)$ according to Definition 5 in Appendix C and obtain

$$\Delta(p(x), d) = \min\{d, 1 - p\} \log \frac{1 - p}{p}. \tag{22}$$

Interestingly, in this case $\Delta(p(x), d)$ has a simple expression that is independent of $p(x)$.

We are now ready to state our second new upper bound, which is proved by bounding $H(Y^n|I_n)$.

*Theorem 7:* For the symmetric primitive relay channel, if a rate $R$ is achievable, then there exists some $p(x)$ and

$a \in [0, \min\{R_0, H(Z|X)\}]$ such that

$$\begin{cases} R \leq I(X;Y,Z) & (23) \\ R \leq I(X;Y) + R_0 - a & (24) \\ R \leq I(X;Y) + \Delta\left(p(x), \sqrt{\frac{a \ln 2}{2}}\right). & (25) \end{cases}$$

Theorem 7 also implies the cut-set bound in Propositions 1. In particular, when the channels $X$-$Y$ and $X$-$Z$ have a fully connected transition matrix, our new bound is *strictly* tighter than the cut-set bound since $\Delta(p(x), d) \to 0$ as $d \to 0$ for any $p(x)$ in this case.

It should be pointed out that the bounds in Theorems 4 and 7 are proved based on essentially different arguments, and they do not include each other in general. For instance, in the case when $X$-$Y$ and $X$-$Z$ are binary erasure channels (i.e. $\Pr(Y = x|x) = 1 - p$, and $\Pr(Y = \text{erasure}|x) = p$, $\forall x \in \{0, 1\}$), $\Delta(p(x), d) = \infty$ for all $p(x)$ and $d > 0$, and thus our second bound reduces to the cut-set bound, but the first bound is still strictly tighter than the cut-set bound; whereas in the case when $X$-$Y$ and $X$-$Z$ are binary symmetric channels, our second bound is significantly tighter than both the cut-set bound and the first bound as we will show in the next section.

## V. Binary Symmetric Channel

As an application of the upper bounds stated in Sections III and IV, we consider the case where the channel is binary symmetric, i.e., both the $X$-$Y$ and $X$-$Z$ links are binary symmetric channels with crossover probability $p$. The various upper bounds can be specialized to this case as follows (see Appendix D for derivations).

- Cut-set bound (Prop. 1):

$$C(R_0) \leq \min\{1 + H_2(p * p) - 2H_2(p), \\ 1 - H_2(p) + R_0\},$$

  where $p_1 * p_2 := p_1(1 - p_2) + p_2(1 - p_1)$.
- Xue's bound (Prop. 3):

$$C(R_0) \leq \max_{a \in [0, R_0]} \min\{1 - H_2(p) + R_0 - a, \\ E^{-1}(H_2(\sqrt{a}) + \sqrt{a})\},$$

  where $E^{-1}(\cdot)$ is the inverse function of $E(R)$.
- Our first bound (Thm. 4):

$$C(R_0) \leq \max_{a \in \left[0, \min\left\{R_0, H_2(p), \frac{1}{2\ln 2}\right\}\right]} \min\left\{1 + H_2(p * p) - 2H_2(p), \\ 1 - H_2(p) + R_0 - a, \\ 1 - H_2(p) + H_2\left(\sqrt{\frac{a \ln 2}{2}}\right) - a\right\}.$$

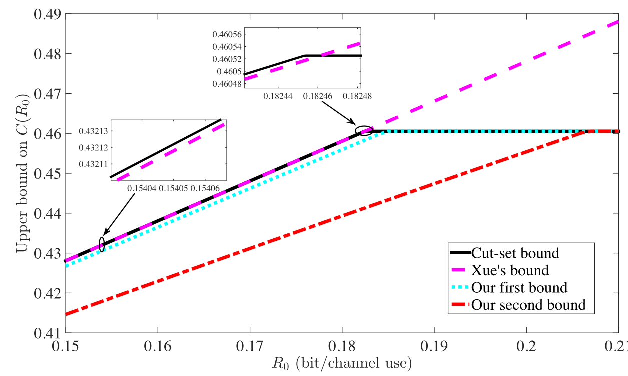


Fig. 2. Upper bounds on $C(R_0)$ for binary symmetric case with $p = 0.2$.

- Our second bound (Thm. 7):

$$C(R_0) \leq \max_{a \in \left[0, \min\left\{R_0, H_2(p), \frac{2}{\ln 2}(1-p)^2\right\}\right]} \min\left\{1 + H_2(p * p) - 2H_2(p), \\ 1 - H_2(p) + R_0 - a, \\ 1 - H_2(p) + \sqrt{\frac{a \ln 2}{2}} \log \frac{1 - p}{p}\right\}.$$

Fig. 2 plots the above bounds for $p = 0.2$ and $R_0 \in [0.15, 0.21]$. As can be seen, both of our bounds perform strictly better than the cut-set bound and Xue's bound, where the latter two are quite close to each other. Particularly, our second bound provides considerable gain over the other three bounds.

### A. Cover's Open Problem on the Critical $R_0$

Now suppose we want to achieve the rate $C_{XYZ}$ for the relay channel. What is the minimum rate needed for the relay–destination link? This question was posed by Cover [20] and has been open for decades. Formally, we are interested in the critical value

$$R_0^* = \inf\{R_0 : C(R_0) = C_{XYZ}\}.$$

The upper bounds on the capacity of the primitive relay channel presented in the previous sections can be immediately used to develop lower bounds on $R_0^*$. Note that since Xue's bound is always dominated by our first bound, in the following we only compare the lower bounds on $R_0^*$ implied by our two bounds with that implied by the cut-set bound.

- Cut-set bound (Prop. 1):

$$R_0^* \geq H_2(p * p) - H_2(p).$$

- Our first bound (Thm. 4):

$$R_0^* \geq \min_{\substack{H_2\left(\sqrt{\frac{a \ln 2}{2}}\right) - a \\ \geq H_2(p * p) - H_2(p)}} H_2(p * p) - H_2(p) + a.$$

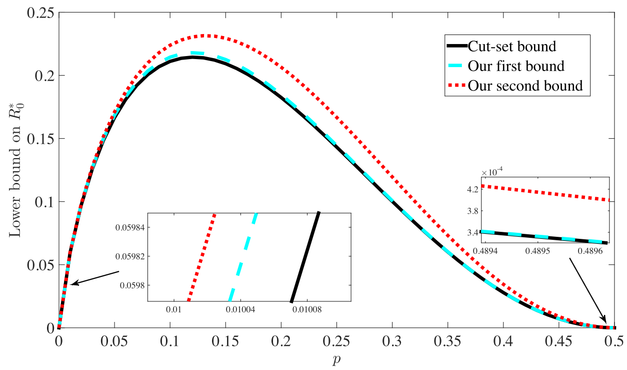


Fig. 3. Lower bounds on $R_0^*$ for binary symmetric case.

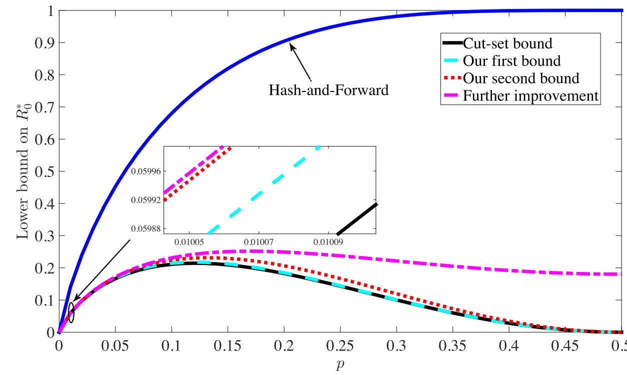


Fig. 4. Improved lower bound on $R_0^*$ for the binary symmetric case.

- Our second bound (Thm. 7):

$$R_0^* \geq H_2(p*p) - H_2(p) + \frac{2}{\ln 2}\left(\frac{H_2(p*p) - H_2(p)}{\log \frac{1-p}{p}}\right)^2.$$

Fig. 3 plots these lower bounds on $R_0^*$ versus the crossover probability $p$. We see again that our second bound provides more gain over the cut-set bound than our first bound does.

From Fig. 3 we observe that all these lower bounds on $R_0^*$ converge to 0 as $p \to 0$ or $p \to 1/2$. On the other hand, to achieve $C_{XYZ}$, a natural way is to use a simple Compress-and-Forward scheme with only Slepian-Wolf binning, a.k.a. Hash-and-Forward (H-F) [27], to faithfully transfer the relay's observation $Z^n$ to the destination so that the joint decoding based on $Z^n$ and $Y^n$ can be performed.[3] This leads to an upper bound on $R_0^*$, namely $R_0^* \leq H_2(p*p)$, where $H_2(p*p)$ is the conditional entropy $H(Z|Y)$ induced by the uniform input distribution. Interestingly, this H-F upper bound also converges to 0 as $p \to 0$; but as $p \to 1/2$, it converges to 1 even though $C_{XYZ}$ is diminishing in this regime, which is in sharp contrast to the above lower bounds on $R_0^*$ that all converge to 0.

This leads to an interesting dichotomy: as $p \to 1/2$ while achievability requires a full bit of $R_0$ to support the diminishing $C_{XYZ}$ rate, the converse results allow for a diminishing $R_0$. Building on our second upper bound on the capacity of the primitive relay channel, in Section IX we prove the following improved lower bound on $R_0^*$, which deviates from 0 as $p \to 1/2$, thus suggesting that a positive $R_0$ is needed to achieve $C_{XYZ}$ even when $C_{XYZ} \to 0$. The proof of this result follows the argument for proving our second bound, however it also critically incorporates the fact that the rate of the codebook is approximately $C_{XYZ}$ in this case as well as the fact that the channel is binary symmetric, which allows us to do a combinatorial geometric analysis of the typical sets in Hamming space.

[3]Note that Decode-and-Forward can not achieve $C_{XYZ}$ here and Compress-and-Forward with binary compression can achieve $C_{XYZ}$ only when the compression process is lossless, in which case it reduces to Hash-and-Forward. One can also check that due to the symmetry of the relay channel model under consideration, a combination of Decode-and-Forward and Compress-and-Forward as in [4, Thm 7] falls back to Compress-and-Forward itself.

*Theorem 8:* For the binary symmetric channel case,

$$R_0^* \geq H_2(p*p) - H_2(p) + \frac{2}{\ln 2}\left(\frac{H_2(p*p) - H_2(p)}{(1-2p)\log \frac{1-p}{p}}\right)^2.$$

Fig. 4 shows this further improved lower bound on $R_0^*$ as well as the H-F upper bound. Clearly, this improved lower bound is tighter than all other lower bounds, and in particular, it converges to a strictly positive value, 0.1803, as $p \to 1/2$ while all the other lower bounds converge to 0. This also shows that $R_0^*$ is discontinuous since when $p = 1/2$, the capacity of the relay channel is 0, and therefore trivially $R_0^* = 0$. We indeed believe that $R_0^* \to 1$ as $p \to 1/2$ but proving this currently remains out of reach.

From Fig. 4, one can observe that in the other extreme, as $p \to 0$, upper and lower bounds on $R_0^*$ do indeed match and all approach 0. One can indeed check that the speed at which they approach 0 is also not too different. In particular, we can show that H-F is approximately optimal within a multiplicative factor of 2 in the regime where $p \to 0$. More precisely, letting $R_0^{\text{H-F}} = H_2(p*p)$ and $R_0^{\text{C-S}} = H_2(p*p) - H_2(p)$ denote the H-F bound and the cut-set bound on $R_0^*$ respectively, it can be shown (see Appendix E) that

$$\frac{R_0^{\text{H-F}}}{R_0^{\text{C-S}}} \to 2 \text{ as } p \to 0. \tag{26}$$

## VI. Preliminary Results

Before proceeding to the proofs of the theorems, we state a few preliminary results that will be used in the sequel.

### A. Fixed Composition Code

We start by overviewing the notion of fixed composition code [25]. This notion will be useful for coupling constraints (14)–(16) together in Theorem 4, and similarly (23)–(25) in Theorem 7, through the input distribution $p(x)$. For the purpose of showing Theorem 4, this notion can be replaced by a time sharing argument [30], however the latter technique is not sufficient in deriving the bound in Theorem 7. Therefore for consistency, this paper employs the fixed composition code argument for proving both Theorems 4 and 7.

*Definition 9:* The composition $Q_{x^n}$ (or empirical probability distribution) of a sequence $x^n$ is the relative proportion of occurrences of each symbol of $\mathbf{\Omega}_X$, i.e., $Q_{x^n}(a) = N(a|x^n)/n$ for all $a \in \mathbf{\Omega}_X$, where $N(a|x^n)$ is the number of times the symbol $a$ occurs in the sequence $x^n$.

*Definition 10:* A code for the primitive relay channel is said to be of fixed composition $Q$, denoted by $(\mathcal{C}_{(n,R)}^{[Q]}, f_n, g_n)$, if all the codewords in $\mathcal{C}_{(n,R)}$ have the same composition $Q$.

The following lemma says that if a rate $R$ is achievable by some sequence of codes, then there exists a sequence of fixed composition codes that can achieve essentially the same rate.

*Lemma 11:* Suppose a rate $R$ is achievable over the primitive relay channel. Then for any $\tau > 0$, there exists a sequence of fixed composition codes with rate $R_\tau := R - \tau$

$$\{(\mathcal{C}_{(n,R_\tau)}^{[Q_n]}, f_n, g_n)\}_{n=1}^{\infty} \tag{27}$$

such that the average probability of error $P_e^{(n)} \to 0$ as $n \to \infty$.

The proof of this lemma relies on the property that there are only a polynomial number of compositions and can be easily extended from the proof for a single user channel [25].

Justified by the above lemma, in deriving the upper bounds in Theorems 4 and 7 we will assume that a reliable fixed composition code, as given in (27), is used to communicate over the relay channel. A benefit of this is that now the various $n$-letter information quantities possess single-letter characterizations or bounds, as demonstrated in the following.

*Lemma 12:* For the $n$-channel use code with fixed composition $Q_n$, we have

$$\begin{aligned} H(Y^n|X^n) = H(Z^n|X^n) = nH(Y|X) = nH(Z|X)\\ H(Y^n, Z^n|X^n) = nH(Y, Z|X), \end{aligned}$$

and

$$\begin{aligned} I(X^n; Y^n) = I(X^n; Z^n) \le nI(X; Y) = nI(X; Z)\\ I(X^n; Y^n, Z^n) \le nI(X; Y, Z) \end{aligned}$$

where $H(Y|X)$, $H(Z|X)$ and $I(X;Y)$, $I(X;Z)$ are calculated based on $Q_n(x)p(\omega|x)$, and $H(Y,Z|X)$ and $I(X;Y,Z)$ are calculated based on $Q_n(x)p(y|x)p(z|x)$.

The proof of this lemma is given in Appendix F.

### *B. Blowing-Up Lemma*

We next recall the generalized blowing-up lemma [29, Lemma 12] which will be the key geometric ingredient in our proofs.

*Lemma 13 (Generalized Blowing-Up Lemma):* Let $U_1, U_2, \ldots, U_n$ be $n$ independent random variables taking values in a finite set $\mathcal{U}$. Then for any $A \subseteq \mathcal{U}^n$ with $\Pr(U^n \in A) \ge 2^{-na_n}$,

$$\Pr(U^n \in \Gamma_{n(\sqrt{\frac{a_n \ln 2}{2}}+t)}(A)) \ge 1 - e^{-2nt^2}, \quad \forall\, t > 0,$$

in which $\Gamma_r(A)$ is the blown-up set of $A$ with radius $r$ defined as

$$\Gamma_r(A) := \left\{u^n \in \mathcal{U}^n : \exists\ v^n \in A \quad \text{s.t. } d_H(u^n, v^n) \le r\right\},$$

where $d_H(u^n, v^n)$ denotes the Hamming distance between the two sequences $u^n$ and $v^n$.

Note that in the above lemma there are no assumptions on the individual probability distribution of $U_1, U_2, \ldots, U_n$; in particular, they are not required to be identically distributed.

### *C. Strong Typicality*

We finally recall the notion of strong typicality and some of the properties of strongly typical sequences that will be used in the sequel. For a more detailed discussion and proofs see [31].

Consider any discrete random variable (or vector) $U \in \mathcal{U}$ with distribution $p(u)$. Let $\mathbf{u} = (u_1, u_2, \ldots, u_B)$ be a $B$-length sequence with elements drawn from $\mathcal{U}$. The set of sequences $\mathbf{u} \in \mathcal{U}^B$ that are $\epsilon$-typical with respect to $U$ is defined as

$$\mathcal{T}_\epsilon^{(B)}(U) = \{\mathbf{u} : |Q_{\mathbf{u}}(u) - p(u)| \le \epsilon p(u), \forall u \in \mathcal{U}\}, \tag{28}$$

where $Q_{\mathbf{u}}(u)$ is the empirical distribution of $\mathbf{u}$ as defined in Definition 9.

The notion of typicality can be readily extended to two random variables $(U, V)$ by treating $(U, V)$ as a single random variable, leading to the jointly $\epsilon$-typical set $\mathcal{T}_\epsilon^{(B)}(U, V)$. Based on this one can further define the conditionally $\epsilon$-typical set

$$\mathcal{T}_\epsilon^{(B)}(V|\mathbf{u}) = \{\mathbf{v} : (\mathbf{u}, \mathbf{v}) \in \mathcal{T}_\epsilon^{(B)}(U, V)\}.$$

Finally, note that the above notions of typicality, joint typicality and conditional typicality extend to three or more random variables.

We now summarize some of the properties of strongly typical sequences that will be used in our proofs. Let $(\mathbf{U}, \mathbf{V}, \mathbf{W})$ be i.i.d. generated according to $p(u, v, w)$, i.e.,

$$p(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \prod_{b=1}^{B} p(u_b, v_b, w_b).$$

Then we have:

1) For any $(\mathbf{u}, \mathbf{v}) \in \mathcal{T}_\epsilon^{(B)}(U, V)$,
$$2^{-B(H(V|U)+\delta)} \le p(\mathbf{v}|\mathbf{u}) \le 2^{-B(H(V|U)-\delta)} \tag{29}$$
for some $\delta \to 0$ as $\epsilon \to 0$;
2) For any $\mathbf{u} \in \mathcal{T}_{\epsilon_0}^{(B)}(U)$ and $\epsilon > \epsilon_0$,
$$\Pr(\mathbf{V} \in \mathcal{T}_\epsilon^{(B)}(V|\mathbf{u})|\mathbf{U} = \mathbf{u}) \to 1 \text{ as } B \to \infty; \tag{30}$$
3) For any $\mathbf{u} \in \mathcal{T}_\epsilon^{(B)}(U)$ and $B$ sufficiently large,
$$2^{B(H(V|U)-\delta)} \le \left|\mathcal{T}_\epsilon^{(B)}(V|\mathbf{u})\right| \le 2^{B(H(V|U)+\delta)} \tag{31}$$
for some $\delta \to 0$ as $\epsilon \to 0$.
4) For any $(\mathbf{u}, \mathbf{v}) \in \mathcal{T}_\epsilon^{(B)}(U, V)$ and $B$ sufficiently large,
$$\begin{aligned} \Pr(\mathbf{W} \in \mathcal{T}_\epsilon^{(B)}(W|\mathbf{u}, \mathbf{v})|\mathbf{U} = \mathbf{u}) \ge 2^{-B(I(V;W|U)+\delta)}\\ \text{and } \Pr(\mathbf{W} \in \mathcal{T}_\epsilon^{(B)}(W|\mathbf{u}, \mathbf{v})|\mathbf{U} = \mathbf{u}) \le 2^{-B(I(V;W|U)-\delta)} \end{aligned} \tag{32}$$
for some $\delta \to 0$ as $\epsilon \to 0$.

## VII. Proof of Theorem 4

We are now ready to prove Theorem 4. In particular, we will prove bounds (14)–(16) sequentially with the focus on showing (16).

### A. Proof of (14)–(15)

We assume that a reliable fixed composition code as given in (27) is used to communicate over the relay channel. Let the relay's transmission be denoted by $I_n = f_n(Z^n)$. Using Lemma 12, we have

$$\begin{aligned} nR_\tau &= n(R-\tau) = H(M) \\ &= I(M; Y^n, I_n) + H(M|Y^n, I_n) \\ &\le I(X^n; Y^n, I_n) + n\epsilon \\ &\le I(X^n; Y^n, Z^n) + n\epsilon \\ &\le n(I(X; Y, Z) + \epsilon) \end{aligned} \tag{33}$$

i.e.,

$$R \le I(X; Y, Z) + \tau + \epsilon \tag{34}$$

for any $\tau, \epsilon > 0$ and sufficiently large $n$, where (33) follows from Fano's inequality.

Moreover, for any $\tau, \epsilon > 0$ and sufficiently large $n$, continuing with (33) we have

$$\begin{aligned} &n(R-\tau) \\ &\quad\le I(X^n; Y^n, I_n) + n\epsilon \\ &\quad= I(X^n; Y^n) + I(X^n; I_n|Y^n) + n\epsilon \\ &\quad= I(X^n; Y^n) + H(I_n|Y^n) - H(I_n|X^n) + n\epsilon && (35) \\ &\quad\le n(I(X; Y) + R_0 - a_n + \epsilon) && (36) \end{aligned}$$

i.e.,

$$R \le I(X; Y) + R_0 - a_n + \tau + \epsilon \tag{37}$$

where

$$a_n := \frac{1}{n} H(I_n|X^n) \tag{38}$$

is subject to the following constraint

$$0 \le a_n \le \min\left\{R_0, \frac{1}{n} H(Z^n|X^n)\right\} = \min\{R_0, H(Z|X)\}. \tag{39}$$

### B. Proof of (16)

The proof of (16) is based on the following key lemma, whose proof is provided at the end of this section. Note that this lemma provides an entropy relation for random variables (of arbitrary dimension) satisfying certain conditions and can be of interest in its own right, decoupled from the specific relay channel problem considered in this paper.

*Lemma 14:* For any fixed $n$, let $I_n$ be an integer random variable and $X^n$, $Y^n$ and $Z^n$ be $n$-length discrete random vectors which form the Markov chain $I_n - Z^n - X^n - Y^n$. Assume moreover that $Z^n$ and $Y^n$ are conditionally i.i.d. and memoryless given $X^n$, i.e., $p_{Y|X}$ are $p_{Z|X}$ are such that

$$p_{Y|X}(\omega|x) = p_{Z|X}(\omega|x) := p(\omega|x), \forall \omega \in \Omega, x \in \Omega_X,$$

where $\Omega_X$ and $\Omega$ denote the alphabet of $X$ and the common alphabet of $Y$ and $Z$ respectively and

$$p_{Y^n,Z^n|X^n}(y^n, z^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i) p(z_i|x_i),$$

for all $y^n, z^n \in \Omega^n, x^n \in \Omega_X^n$; and $I_n = f_n(Z^n)$ is a deterministic mapping of $Z^n$ to a set of integers. Let $H(I_n|X^n) = na_n$ for some $a_n \ge 0$. Then

$$H(I_n|Y^n) \le nV\left(\sqrt{\frac{a_n \ln 2}{2}}\right), \tag{40}$$

with

$$V(r) := \begin{cases} \log|\Omega| & \text{if } r > \frac{|\Omega|-1}{|\Omega|} \quad (41) \\ H_2(r) + r\log(|\Omega|-1) & \text{if } r \le \frac{|\Omega|-1}{|\Omega|} \quad (42) \end{cases}$$

where $H_2(r)$ is the binary entropy function defined as $H_2(r) = -r\log r - (1-r)\log(1-r)$.

To apply the above lemma and prove (16), we continue with (35). Instead of upper bounding $H(I_n|Y^n)$ by $nR_0$ as in (36), we use relation (40) to upper bound $H(I_n|Y^n)$. It is trivial to observe that the random variables $(I_n, Z^n, X^n, Y^n)$ in the relay channel satisfy the technical conditions of the lemma. Thus, recalling the definition of $a_n$ in (38) and plugging the bound (40) on $H(I_n|Y^n)$ into (35), we have for any $\tau, \epsilon > 0$ and sufficiently large $n$,

$$R \le I(X; Y) + V\left(\sqrt{\frac{a_n \ln 2}{2}}\right) - a_n + \tau + \epsilon. \tag{43}$$

Combining (34), (37), (43) and (39), we have that if a rate $R$ is achievable, then for any $\delta > 0$ and sufficiently large $n$,

$$\begin{cases} R \le I(X; Y, Z) + \delta \\ R \le I(X; Y) + R_0 - a_n + \delta \\ R \le I(X; Y) + V\left(\sqrt{\frac{a_n \ln 2}{2}}\right) - a_n + \delta \end{cases}$$

where

$$a_n \in [0, \min\{R_0, H(Z|X)\}].$$

Since $\delta$ can be made arbitrarily small, we arrive at the following proposition.

*Proposition 15:* If a rate $R$ is achievable, then there exists some $p(x)$ and $a \in [0, \min\{R_0, H(Z|X)\}]$ such that

$$\begin{cases} R \le I(X; Y, Z) & (44) \\ R \le I(X; Y) + R_0 - a & (45) \\ R \le I(X; Y) + V\left(\sqrt{\frac{a \ln 2}{2}}\right) - a & (46) \end{cases}$$

where

$$V\left(\sqrt{\frac{a \ln 2}{2}}\right) = \begin{cases} \log|\Omega| & \text{if } a > \frac{2}{\ln 2}\left(\frac{|\Omega|-1}{|\Omega|}\right)^2 \\ H_2\left(\sqrt{\frac{a\ln 2}{2}}\right) + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega|-1) & \text{otherwise.} \end{cases}$$

Now we show that Proposition 15 is in fact equivalent to Theorem 4.

Theorem 4 → Proposition 15: Suppose Theorem 4 is true. Then, for any $R$ achievable, there exists some

$$a \in \left[0, \min\left\{R_0, H(Z|X), \frac{2}{\ln 2}\left(\frac{|\Omega|-1}{|\Omega|}\right)^2\right\}\right]$$

satisfying (14)–(16). For such $a \leq \frac{2}{\ln 2}\left(\frac{|\Omega|-1}{|\Omega|}\right)^2$, (46) reduces to (16) and thus Proposition 15 is also true.

Proposition 15 $\rightarrow$ Theorem 4: Suppose Proposition 15 is true. Then, for any $R$ achievable, there exists some $a \in [0, \min\{R_0, H(Z|X)\}]$ satisfying (44)–(46). If such $a \leq \frac{2}{\ln 2}\left(\frac{|\Omega|-1}{|\Omega|}\right)^2$, then (14)–(16) hold with this $a$; otherwise, (14)–(16) hold with the choice of $a' = \frac{2}{\ln 2}\left(\frac{|\Omega|-1}{|\Omega|}\right)^2$. In either case, Theorem 4 is also true.

This establishes the equivalence between Proposition 15 and Theorem 4, and thus completes the proof of Theorem 4 apart from the proof of Lemma 14. We next provide the proof for this lemma.

*Proof of Lemma 14:* To prove Lemma 14, we lift the $n$-dimensional random variables to a higher dimensional, say $nB$ dimensional space, and invoke the concept of strong typicality. Specifically, consider the $B$-length i.i.d. extensions of the random variables $X^n, Y^n, Z^n$ and $I_n$, i.e.,

$$\{(X^n(b), Y^n(b), Z^n(b), I_n(b))\}_{b=1}^B, \tag{47}$$

where for any $b \in [1:B]$, $(X^n(b), Y^n(b), Z^n(b), I_n(b))$ has the same distribution as $(X^n, Y^n, Z^n, I_n)$. For notational convenience, in the sequel we write the $B$-length vector $[X^n(1), X^n(2), \ldots, X^n(B)]$ as $\mathbf{X}$ and similarly define $\mathbf{Y}, \mathbf{Z}$ and $\mathbf{I}$; note here we have $\mathbf{I} = [f_n(Z^n(1)), f_n(Z^n(2)), \ldots, f_n(Z^n(B))] =: f(\mathbf{Z})$.

Consider now the $\epsilon$-jointly typical set $\mathcal{T}_\epsilon^{(B)}(X^n, I_n)$ with respect to $(X^n, I_n)$, defined similarly as in (28). Due to property (29), for any $(\mathbf{x}, \mathbf{i}) \in \mathcal{T}_\epsilon^{(B)}(X^n, I_n)$ we have

$$p(\mathbf{i}|\mathbf{x}) \geq 2^{-B(H(I_n|X^n)+\epsilon_1)} \geq 2^{-nB(a_n+\epsilon_1)},$$

for some $\epsilon_1 \to 0$ as $\epsilon \to 0$, i.e.,

$$\Pr(\mathbf{Z} \in f^{-1}(\mathbf{i})|\mathbf{x}) \geq 2^{-nB(a_n+\epsilon_1)},$$

where $f^{-1}(\mathbf{i}) := \{\underline{\omega} \in \Omega^{nB} : f(\underline{\omega}) = \mathbf{i}\}$ with $\Omega$ denoting the common alphabet of $Z$ and $Y$.

We will now blow-up the set $f^{-1}(\mathbf{i})$. Note that by the assumptions of the lemma, $\mathbf{Z}$ is an $nB$-length sequence of independent random variables given $\mathbf{x}$, and thus by applying Lemma 13 we obtain (48), shown at the bottom of this page, for sufficiently large $B$. Since $\mathbf{Y}$ and $\mathbf{Z}$ are identically distributed given $\mathbf{X}$, we have

$$t\Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) \geq 1-\sqrt{\epsilon_1},$$

and thus,

$$\begin{aligned}
&\Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{I}))) \\
&= \sum_{(\mathbf{x},\mathbf{i})} \Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x},\mathbf{i})p(\mathbf{x},\mathbf{i}) \\
&= \sum_{(\mathbf{x},\mathbf{i})} \Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x})p(\mathbf{x},\mathbf{i}) \qquad (49) \\
&\geq \sum_{(\mathbf{x},\mathbf{i}) \in \mathcal{T}_\epsilon^{(B)}(X^n,I_n)} \Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) \\
&\quad \times p(\mathbf{x},\mathbf{i}) \\
&\geq (1-\sqrt{\epsilon_1}) \sum_{(\mathbf{x},\mathbf{i}) \in \mathcal{T}_\epsilon^{(B)}(X^n,I_n)} p(\mathbf{x},\mathbf{i}) \\
&\geq (1-\sqrt{\epsilon_1})^2 \\
&\geq 1-2\sqrt{\epsilon_1} \qquad (50)
\end{aligned}$$

for sufficiently large $B$, where (49) follows due to the Markov chain: $\mathbf{Y} - \mathbf{X} - \mathbf{Z} - \mathbf{I}$, and (50) follows since $\Pr(\mathcal{T}_\epsilon^{(B)}(X^n, I_n)) \to 1$ as $B \to \infty$. Finally, choosing $\delta$ to be $2\sqrt{\epsilon_1}$ we arrive at the following proposition.

*Proposition 16:* For any $\delta > 0$ and $B$ sufficiently large,

$$\Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+\delta)}(f^{-1}(\mathbf{I}))) \geq 1-\delta.$$

With the above proposition, we now upper bound $H(\mathbf{I}|\mathbf{Y})$. Let

$$E = \mathbb{I}(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+\delta)}(f^{-1}(\mathbf{I})))$$

where $\mathbb{I}(\cdot)$ is the indicator function defined as

$$\mathbb{I}(A) = \begin{cases} 1 & \text{if } A \text{ holds} \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned}
H(\mathbf{I}|\mathbf{Y}) &\leq H(\mathbf{I}, E|\mathbf{Y}) \\
&= H(E|\mathbf{Y}) + H(\mathbf{I}|\mathbf{Y}, E) \\
&\leq H(\mathbf{I}|\mathbf{Y}, E) + 1 \\
&= \Pr(E=1)H(\mathbf{I}|\mathbf{Y}, E=1) \\
&\quad +\Pr(E=0)H(\mathbf{I}|\mathbf{Y}, E=0) + 1 \\
&\leq H(\mathbf{I}|\mathbf{Y}, E=1) + \delta nBR_0 + 1. \qquad (51)
\end{aligned}$$

To bound $H(\mathbf{I}|\mathbf{Y}, E=1)$, consider a Hamming ball centered

$$\begin{aligned}
\Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) &= \Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{(a_n+\epsilon_1)\ln 2}{2}}+[\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1}-\sqrt{\frac{(a_n+\epsilon_1)\ln 2}{2}}])}(f^{-1}(\mathbf{i}))|\mathbf{x}) \\
&\geq \Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{(a_n+\epsilon_1)\ln 2}{2}}+[\sqrt{\frac{a_n \ln 2}{2}}+2\sqrt{\epsilon_1}-\sqrt{\frac{a_n \ln 2}{2}}-\sqrt{\frac{\epsilon_1 \ln 2}{2}}])}(f^{-1}(\mathbf{i}))|\mathbf{x}) \\
&\geq \Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{(a_n+\epsilon_1)\ln 2}{2}}+\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) \\
&\geq 1-e^{-2nB\epsilon_1} \\
&\geq 1-\sqrt{\epsilon_1}
\end{aligned} \tag{48}$$

at $\mathbf{Y}$ of radius $nB\left(\sqrt{\frac{a_n \ln 2}{2}}+\delta\right)$, which we denote as[4]

$$\begin{aligned}&\mathrm{Ball}\left(\mathbf{Y}, nB\left(\sqrt{\frac{a_n \ln 2}{2}}+\delta\right)\right)\\&\quad:=\left\{\underline{\omega} : d_H(\underline{\omega},\mathbf{Y}) \le nB\left(\sqrt{\frac{a_n \ln 2}{2}}+\delta\right)\right\}.\end{aligned}$$

The condition $E = 1$, i.e., $\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}}+\delta)}(f^{-1}(\mathbf{I}))$, ensures that there is at least one point $\underline{\omega} \in f^{-1}(\mathbf{I})$ belonging to this ball, and therefore, given $E = 1$ and $\mathbf{Y}$ the number of different possibilities for $\mathbf{I}$ is bounded by $\left|\mathrm{Ball}\left(\mathbf{Y}, nB\left(\sqrt{\frac{a_n \ln 2}{2}}+\delta\right)\right)\right|$, the number of sequences in this Hamming ball, leading to the following upper bound on $H(\mathbf{I}|\mathbf{Y}, E=1)$,

$$\begin{aligned}H(\mathbf{I}|\mathbf{Y}, E=1) &\le \log\left|\mathrm{Ball}\left(\mathbf{Y}, nB\left(\sqrt{\frac{a_n \ln 2}{2}}+\delta\right)\right)\right|\\&= nBV\left(\sqrt{\frac{a_n \ln 2}{2}}+\delta\right) \qquad (52)\\&\le nB\left[V\left(\sqrt{\frac{a_n \ln 2}{2}}\right)+\delta_1\right] \qquad (53)\end{aligned}$$

for some $\delta_1 \to 0$ as $\delta \to 0$, where the function $V(\cdot)$ is defined as in (41)–(42), (52) follows from the characterization of the volume of a Hamming ball (see Appendix G for details), and (53) follows from the continuity of the function $V(\cdot)$. Plugging (53) into (51), we have

$$H(\mathbf{I}|\mathbf{Y}) \le nB\left[V\left(\sqrt{\frac{a_n \ln 2}{2}}\right)+\delta_1\right]+\delta nBR_0+1.$$

Dividing $B$ at both sides of the above inequality and noting that

$$H(\mathbf{I}|\mathbf{Y}) = \sum_{b=1}^{B} H(I_n(b)|Y^n(b)) = BH(I_n|Y^n),$$

we have

$$H(I_n|Y^n) \le n\left(V\left(\sqrt{\frac{a_n \ln 2}{2}}\right)+\delta_1+\delta R_0+\frac{1}{nB}\right). \qquad (54)$$

Since $\delta, \delta_1$ and $\frac{1}{nB}$ in (54) can all be made arbitrarily small by choosing $B$ sufficiently large, we obtain

$$H(I_n|Y^n) \le nV\left(\sqrt{\frac{a_n \ln 2}{2}}\right). \qquad (55)$$

This finishes the proof of Lemma 14.

[4]The Hamming ball here should be distinguished from the notion of Hamming sphere that will be used later in Section IX. Specifically, a Hamming ball centered at $\mathbf{c}$ of radius $r$, denoted by $\mathrm{Ball}(\mathbf{c}, r)$, is defined as the set of points that are within Hamming distance $r$ of $\mathbf{c}$, whereas a corresponding Hamming sphere, denoted by $\mathrm{Sphere}(\mathbf{c}, r)$, is the set of points that are at a Hamming distance equal to $r$ from $\mathbf{c}$.

## VIII. Proof of Theorem 7

We now prove Theorem 7. The bounds (23)–(24) are the same as (14)–(15), which have been proved in Section VII. To show (25), we still assume that we use a reliable fixed composition code for communicating over the relay channel as given in (27). However instead of using Lemma 14 to upper bound $H(I_n|Y^n)$ as we did in the previous section, we will now use the following lemma, which upper bounds the conditional entropy $H(Y^n|I_n)$ and whose proof is given in Section VIII-A. Note that just like Lemma 14, the statement of this lemma is decoupled from the relay channel problem.

*Lemma 17:* For any fixed $n$, let $I_n$ be an integer random variable and $X^n$, $Y^n$ and $Z^n$ be $n$-length discrete random vectors which form the Markov chain $I_n - Z^n - X^n - Y^n$. Assume moreover that $Z^n$ and $Y^n$ are conditionally i.i.d. and memoryless given $X^n$, i.e., $p_{Y|X}$ are $p_{Z|X}$ are such that

$$p_{Y|X}(\omega|x) = p_{Z|X}(\omega|x) := p(\omega|x), \quad \forall \omega \in \boldsymbol{\Omega}, x \in \boldsymbol{\Omega}_X,$$

where $\boldsymbol{\Omega}_X$ and $\boldsymbol{\Omega}$ denote the alphabet of $X$ and the common alphabet of $Y$ and $Z$ respectively and

$$p_{Y^n,Z^n|X^n}(y^n, z^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)p(z_i|x_i),$$

for all $y^n, z^n \in \boldsymbol{\Omega}^n, x^n \in \boldsymbol{\Omega}_X^n$; and $I_n = f_n(Z^n)$ is a deterministic mapping of $Z^n$ to a set of integers; and $X^n$ has a fixed composition $Q_n(x)$. Let $H(I_n|X^n) = na_n$ for some $a_n \ge 0$. Then

$$\begin{aligned}H(Y^n|I_n) &\le H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X)\\&\quad + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right), \qquad (56)\end{aligned}$$

where $H(Y|X)$ is calculated based on $Q_n(x)p(\omega|x)$, and $\Delta(\cdot,\cdot)$ is as defined in (18)–(19).

It is straightforward to observe that the random variables associated with the $n$-blocklength reliable fixed composition code satisfy the assumptions of the lemma. Therefore for the relay channel, we have

$$\begin{aligned}&n(R-\tau)\\&\le I(X^n; Y^n, I_n) + n\epsilon\\&= I(X^n; I_n) + I(X^n; Y^n|I_n) + n\epsilon\\&= H(X^n) - H(X^n|I_n) + H(Y^n|I_n) - H(Y^n|X^n) + n\epsilon\\&\le H(X^n) - H(X^n|I_n) + H(X^n|I_n) - H(X^n|Z^n)\\&\quad + nH(Y|X) + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) - H(Y^n|X^n) + n\epsilon\\&= I(X^n; Z^n) + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + n\epsilon \qquad (57)\\&\le n\left[I(X;Y) + \Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + \epsilon\right] \qquad (58)\end{aligned}$$

for any $\tau, \epsilon > 0$ and $n$ sufficiently large, where in (57) we have used the fact that $H(Y^n|X^n) = nH(Y|X)$ (cf. Lemma 12), and (58) follows from the symmetry between $Y^n$ and $Z^n$ and Lemma 12 again. This proves the bound (25) and hence Theorem 7. In the rest of this section we prove Lemma 17.

### *A. Proof of Lemma 17*

The remaining step then is to show the entropy inequality (56) in Lemma 17. For this, again we look at the $B$-length i.i.d. sequence of $(X^n, Y^n, Z^n, I_n)$, i.e.,

$$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{I}) := \{(X^n(b), Y^n(b), Z^n(b), I_n(b))\}_{b=1}^B.$$

The following lemma is crucial for proving inequality (56), and its own proof will be given in the next subsection.

*Lemma 18:* For any $\delta > 0$ and $B$ sufficiently large, there exists a set $\mathcal{S}(Y^n, I_n)$ of $(\mathbf{y}, \mathbf{i})$ pairs such that

$$\Pr((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n)) \geq 1 - \delta,$$

and for any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$p(\mathbf{y}|\mathbf{i}) \geq 2^{-B\left(H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + \delta\right)}.$$

We will now use Lemma 18 to prove Lemma 17. Letting $E = \mathbb{I}((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n))$, we have for any $\delta > 0$ and $B$ sufficiently large,

$$\begin{aligned}
H(\mathbf{Y}|\mathbf{I}) &\leq H(\mathbf{Y}, E|\mathbf{I}) \\
&= H(E|\mathbf{I}) + H(\mathbf{Y}|\mathbf{I}, E) \\
&\leq H(\mathbf{Y}|\mathbf{I}, E) + 1 \\
&= \Pr(E = 1)H(\mathbf{Y}|\mathbf{I}, E = 1) \\
&\quad + \Pr(E = 0)H(\mathbf{Y}|\mathbf{I}, E = 0) + 1 \\
&\leq H(\mathbf{Y}|\mathbf{I}, E = 1) + \delta nB \log|\Omega| + 1 \\
&= -\sum_{(\mathbf{y},\mathbf{i}) \in \mathcal{S}(Y^n, I_n)} p(\mathbf{y}, \mathbf{i}|E = 1) \log p(\mathbf{y}|\mathbf{i}, E = 1) \\
&\quad + \delta nB \log|\Omega| + 1 \\
&\leq -\sum_{(\mathbf{y},\mathbf{i}) \in \mathcal{S}(Y^n, I_n)} p(\mathbf{y}, \mathbf{i}|E = 1) \log p(\mathbf{y}|\mathbf{i}) \\
&\quad + \delta nB \log|\Omega| + 1 \qquad (59) \\
&\leq B\bigg[H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) \\
&\quad + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + \delta\bigg] + \delta nB \log|\Omega| + 1, \qquad (60)
\end{aligned}$$

where (59) follows because for any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$, we have

$$\begin{aligned}
p(\mathbf{y}|\mathbf{i}, E = 1) &= \frac{p(\mathbf{y}|\mathbf{i})p(E = 1|\mathbf{y}, \mathbf{i})}{p(E = 1|\mathbf{i})} \\
&= \frac{p(\mathbf{y}|\mathbf{i})}{p(E = 1|\mathbf{i})} \\
&\geq p(\mathbf{y}|\mathbf{i}),
\end{aligned}$$

where the second equality holds since $(\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n)$ implies $E = 1$. Dividing $B$ at both sides of (60) and noticing that $H(\mathbf{Y}|\mathbf{I}) = BH(Y^n|I_n)$, we have

$$\begin{aligned}
H(Y^n|I_n) &\leq H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) \\
&\quad + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + \delta + \delta n \log|\Omega| + \frac{1}{B}.
\end{aligned}$$

Since both $\delta$ and $\frac{1}{B}$ in the above inequality can be made arbitrarily small by choosing $B$ sufficiently large, Lemma 17 is thus proved.

### *B. Proof of Lemma 18*

Let $\mathcal{S}(Y^n, I_n)$ be defined as

$$\mathcal{S}(Y^n, I_n) := \{(\mathbf{y}, \mathbf{i}) : \mathbf{y} \in \Gamma_{nB\left(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon\right)}(\mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i}))\}. \qquad (61)$$

We first show that for any $\epsilon > 0$ and $B$ sufficiently large,

$$\Pr((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n)) \geq 1 - \epsilon. \qquad (62)$$

For this, consider any $(\mathbf{x}, \mathbf{i}) \in \mathcal{T}_{\tilde{\epsilon}}^{(B)}(X^n, I_n)$, $\tilde{\epsilon} > 0$. From property (32) of jointly typical sequences, we have

$$\begin{aligned}
\Pr(\mathbf{Z} \in \mathcal{T}_{\tilde{\epsilon}}^{(B)}(Z^n|\mathbf{x}, \mathbf{i})|\mathbf{x}) &\geq 2^{-B(I(Z^n;I_n|X^n) + \tilde{\epsilon}_1)} \\
&= 2^{-B(H(I_n|X^n) + \tilde{\epsilon}_1)} \\
&\geq 2^{-nB(a_n + \tilde{\epsilon}_1)},
\end{aligned}$$

where $\tilde{\epsilon}_1 \to 0$ as $\tilde{\epsilon} \to 0$ and $B \to \infty$. Since $T_{\tilde{\epsilon}}^{(B)}(Z^n|\mathbf{x}, \mathbf{i}) \subseteq T_{\tilde{\epsilon}}^{(B)}(Z^n|\mathbf{i})$, we further have

$$\Pr(\mathbf{Z} \in \mathcal{T}_{\tilde{\epsilon}}^{(B)}(Z^n|\mathbf{i})|\mathbf{x}) \geq 2^{-nB(a_n + \tilde{\epsilon}_1)}.$$

Then, by applying Lemma 13 along the same lines as the proof of Proposition 16, we can obtain

$$\Pr(\mathbf{Y} \in \Gamma_{nB\left(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\tilde{\epsilon}_1}\right)}(\mathcal{T}_{\tilde{\epsilon}}^{(B)}(Z^n|\mathbf{i}))) \geq 1 - 2\sqrt{\tilde{\epsilon}_1}$$

for sufficiently large $B$. Choosing $\epsilon$ to be $\max\{2\sqrt{\tilde{\epsilon}_1}, \tilde{\epsilon}\}$ then proves (62).

Consider any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$. By the definition of $\mathcal{S}(Y^n, I_n)$, we can find one $\mathbf{z} \in \mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})$ such that

$$d_H(\mathbf{y}, \mathbf{z}) \leq nB\left(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon\right). \qquad (63)$$

Then,

$$\begin{aligned}
p(\mathbf{y}|\mathbf{i}) &= \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i}) \\
&\geq \sum_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i}) \qquad (64) \\
&\geq 2^{-B(H(X^n|I_n) + \epsilon_1)} \sum_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x}) \qquad (65) \\
&\geq 2^{-B(H(X^n|I_n) + \epsilon_1)} \left|\mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})\right| \min_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x}) \qquad (66) \\
&\geq 2^{-B(H(X^n|I_n) + \epsilon_1)} 2^{B(H(X^n|Z^n) - \epsilon_2)} \min_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x}), \qquad (67)
\end{aligned}$$

for some $\epsilon_1, \epsilon_2 \to 0$ as $\epsilon \to 0$ and $B \to \infty$, where the $\mathbf{z}$ throughout (64)–(67) is the one belonging to $\mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})$ and satisfying (63), and (65) and (67) follow from properties (29) and (31) of jointly typical sequences respectively.

We now lower bound $p(\mathbf{y}|\mathbf{x})$ for any $\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})$. Since $\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})$, we have $(\mathbf{x}, \mathbf{z}) \in \mathcal{T}_\epsilon^{(B)}(X^n, Z^n)$, i.e., $(\mathbf{x}, \mathbf{z})$ are jointly typical with respect to the $n$-letter random variables $(X^n, Z^n)$. Due to the fixed composition code assumption and the discrete memoryless property of the channel, this can be shown (see Appendix H) to further imply that $(\mathbf{x}, \mathbf{z})$ are also jointly typical with respect to the single-letter random variables $(X, Z)$, i.e.,

$$|P_{(\mathbf{x},\mathbf{z})}(x,\omega) - Q_n(x)p(\omega|x)| \leq \epsilon_3 Q_n(x)p(\omega|x), \forall (x,\omega) \tag{68}$$

for some $\epsilon_3 \to 0$ as $\epsilon \to 0$, where $P_{(\mathbf{x},\mathbf{z})}(x,\omega)$ denotes the joint empirical distribution of $(\mathbf{x}, \mathbf{z})$ with respect to $(X, Z)$, defined as

$$P_{(\mathbf{x},\mathbf{z})}(x,\omega) = \tfrac{1}{nB} N(x,\omega|\mathbf{x},\mathbf{z})$$

where $N(x,\omega|\mathbf{x},\mathbf{z})$ denotes the number of times the symbols $(x,\omega)$ occur in the sequences $(\mathbf{x}, \mathbf{z})$. On the other hand, we show in Appendix I that the bound (63) on the Hamming distance between $\mathbf{y}$ and $\mathbf{z}$ can translate to a bound on the total variation distance between the two empirical distributions $P_{(\mathbf{x},\mathbf{y})}(x,\omega)$ and $P_{(\mathbf{x},\mathbf{z})}(x,\omega)$ for any $\mathbf{x}$, namely,

$$\begin{aligned}\sum_{(x,\omega)} |P_{(\mathbf{x},\mathbf{y})}(x,\omega) - P_{(\mathbf{x},\mathbf{z})}(x,\omega)| &\leq \frac{2}{nB} d_H(\mathbf{y},\mathbf{z}) \\ &\leq 2\left(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon\right).\end{aligned} \tag{69}$$

Combining (68) and (69), we have for some $\epsilon_4 \to 0$ as $\epsilon \to 0$,

$$\sum_{(x,\omega)} |P_{(\mathbf{x},\mathbf{y})}(x,\omega) - Q_n(x)p(\omega|x)| \leq 2\sqrt{\frac{a_n \ln 2}{2}} + \epsilon_4, \tag{70}$$

or equivalently expressed as

$$\frac{1}{2}\sum_{(x,\omega)} |Q_n(x)P_{\mathbf{y}|\mathbf{x}}(\omega|x) - Q_n(x)p(\omega|x)| \leq \sqrt{\frac{a_n \ln 2}{2}} + \frac{\epsilon_4}{2}, \tag{71}$$

where we have used the fact that the empirical distribution $P_{\mathbf{x}}(x) = Q_n(x)$ due to the fixed composition code assumption, and $P_{\mathbf{y}|\mathbf{x}}(\omega|x)$ is the conditional empirical distribution satisfying

$$N(x,\omega|\mathbf{x},\mathbf{y}) = N(x|\mathbf{x})P_{\mathbf{y}|\mathbf{x}}(\omega|x).$$

To bound $p(\mathbf{y}|\mathbf{x})$, we have

$$\begin{aligned}
&-\frac{1}{nB}\log p(\mathbf{y}|\mathbf{x}) \\
&= -\frac{1}{nB}\sum_{i=1}^{nB} \log p(y_i|x_i) \\
&= -\frac{1}{nB}\sum_{(x,\omega)} N(x,\omega|\mathbf{x},\mathbf{y}) \log p(\omega|x) \\
&= -\sum_{(x,\omega)} P_{(\mathbf{x},\mathbf{y})}(x,w)\log p(\omega|x) \\
&= \sum_{(x,\omega)} [-P_{(\mathbf{x},\mathbf{y})}(x,w)\log p(\omega|x) \\
&\quad + P_{(\mathbf{x},\mathbf{y})}(x,w)\log P_{\mathbf{y}|\mathbf{x}}(w|x) \\
&\quad - P_{(\mathbf{x},\mathbf{y})}(x,w)\log P_{\mathbf{y}|\mathbf{x}}(w|x)] \\
&= -\sum_{(x,\omega)} P_{(\mathbf{x},\mathbf{y})}(x,w)\log P_{\mathbf{y}|\mathbf{x}}(w|x) \\
&\quad + \sum_{(x,\omega)} P_{(\mathbf{x},\mathbf{y})}(x,w)\log \frac{P_{\mathbf{y}|\mathbf{x}}(w|x)}{p(\omega|x)} \\
&= H(P_{\mathbf{y}|\mathbf{x}}(\omega|x)|P_{\mathbf{x}}(x)) + D(P_{\mathbf{y}|\mathbf{x}}(\omega|x)||p(\omega|x)|P_{\mathbf{x}}(x)) \\
&= H(P_{\mathbf{y}|\mathbf{x}}(\omega|x)|Q_n(x)) + D(P_{\mathbf{y}|\mathbf{x}}(\omega|x)||p(\omega|x)|Q_n(x)),
\end{aligned} \tag{72}$$

where the second equality follows from grouping $\log p(y_i|x_i)$ according to different $(x,\omega)$ and then summing over $(x,\omega)$, and $P_{\mathbf{y}|\mathbf{x}}(\omega|x)$ satisfies the constraint (71). For any $p(x)$ and $d \geq 0$, define $\Delta\,(p(x),d)$ as follows:

$$\begin{aligned}
&\Delta\,(p(x),d) \\
&\quad := \max_{\tilde{p}(\omega|x)} H(\tilde{p}(\omega|x)|p(x)) + D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) \\
&\qquad - H(p(\omega|x)|p(x)) &(73)\\
&\qquad \text{s.t. } \frac{1}{2}\sum_{(x,\omega)} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq d. &(74)
\end{aligned}$$

Comparing (72) and (71) to (73) and (74), we have

$$\begin{aligned}-\frac{1}{nB}\log p(\mathbf{y}|\mathbf{x}) &\leq \Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}} + \frac{\epsilon_4}{2}\right) + H(Y|X) \\ &\leq \Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + H(Y|X) + \epsilon_5\end{aligned} \tag{75}$$

for some $\epsilon_5 \to 0$ as $\epsilon \to 0$, where $H(Y|X)$ is calculated based on $Q_n(x)p(\omega|x)$, and the last inequality follows since $\Delta(p(x),d)$ is continuous in $d$ for $d > 0$. This combined with (67) yields that for any $(\mathbf{y},\mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$\begin{aligned}
p(\mathbf{y}|\mathbf{i}) &\geq 2^{-B(H(X^n|I_n)+\epsilon_1)} 2^{B(H(X^n|Z^n)-\epsilon_2)} \\
&\quad \times 2^{-nB(\Delta\left(Q_n,\sqrt{\frac{a_n \ln 2}{2}}\right)+H(Y|X)+\epsilon_5)} \\
&\geq 2^{-B(H(X^n|I_n)-H(X^n|Z^n)+nH(Y|X)+n\Delta\left(Q_n,\sqrt{\frac{a_n \ln 2}{2}}\right)+\epsilon_6)}
\end{aligned}$$

for some $\epsilon_6 \to 0$ as $\epsilon \to 0$ and $B \to \infty$. Finally, choosing $\delta = \max\{\epsilon, \epsilon_6\}$, we have

$$\Pr((\mathbf{Y},\mathbf{I}) \in \mathcal{S}(Y^n, I_n)) \geq 1 - \delta,$$

and for any $(\mathbf{y},\mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$\begin{aligned}
&p(\mathbf{y}|\mathbf{i}) \\
&\quad \geq 2^{-B(H(X^n|I_n)-H(X^n|Z^n)+nH(Y|X)+n\Delta\left(Q_n,\sqrt{\frac{a_n \ln 2}{2}}\right)+\delta)},
\end{aligned}$$

which concludes the proof of Lemma 18.

## IX. Proof of Theorem 8

The main idea for proving Theorem 8 follows that for Theorem 7. In order to highlight the difference, we first look at the the parameter $\Delta(p(x),d)$ that plays an important role in the bound in Theorem 7 more closely. In Section IV-C, we have indicated that $\Delta(p(x),d)$ can be interpreted as the maximal number of extra bits we would need to compress $Y$ given $X$, when $Y$ comes from a conditional distribution

$\tilde{p}(w|x)$ instead of the assumed distribution $p(w|x)$ and the total variation distance between the two joint distributions is bounded by $d$. An alternative role that emerges for this quantity in the context of the proof of Theorem 7 is the following.

Consider a pair $(\mathbf{x},\mathbf{z})$ of $nB$-length sequences that are jointly typical with respect to $p(x)p(\omega|x)$. We have[5]

$$p(\mathbf{z}|\mathbf{x}) \doteq 2^{-nBH(p(\omega|x)|p(x))}. \tag{76}$$

Let $\mathbf{y}$ be a sequence taking values in the same alphabet as $\mathbf{z}$ and bounded in its Hamming distance to $\mathbf{z}$ by $nBd$. Theorem 7 is based on obtaining a lower bound on the conditional probability of the sequence $\mathbf{y}$ given $\mathbf{x}$ under $p(x)p(\omega|x)$. In particular, in (75), we show that

$$p(\mathbf{y}|\mathbf{x}) \dot{\geq} 2^{-nB[H(p(\omega|x)|p(x))+\Delta(p(x),d)]}. \tag{77}$$

Comparing (76) and (77), we can see that $\Delta(p(x),d)$ characterizes the maximum possible exponential decrease from $p(\mathbf{z}|\mathbf{x})$ to $p(\mathbf{y}|\mathbf{x})$ where $(\mathbf{x},\mathbf{z})$ is jointly typical with respect to $p(x)p(\omega|x)$ and the Hamming distance between $\mathbf{y}$ and $\mathbf{z}$ is bounded by $nBd$.

For the binary symmetric channel, i.e. when the conditional distribution $p(\omega|x)$ corresponds to a binary symmetric channel with crossover probability $p < 1/2$, we show in Appendix C that we have the following explicit expression

$$\begin{aligned}\Delta(p(x),d) &= \min\left\{H_2(p) + d\log\frac{1-p}{p}, -\log p\right\} - H_2(p)\\ &= \min\{d, 1-p\}\log\frac{1-p}{p}. \end{aligned}\tag{78}$$

We next provide an alternative way to obtain this expression by resorting to the above interpretation of $\Delta(p(x),d)$. Note that when $p(\omega|x)$ corresponds to a binary symmetric channel with crossover portability $p < 1/2$, for a $(\mathbf{x},\mathbf{z})$ pair that is jointly typical with respect to $p(x)p(\omega|x)$, we have $d_H(\mathbf{x},\mathbf{z}) \leq nB(p+\epsilon)$ and

$$p(\mathbf{z}|\mathbf{x}) \doteq 2^{-nBH_2(p)}. \tag{79}$$

If $\mathbf{y}$ satisfies $d_H(\mathbf{y},\mathbf{z}) \leq nBd$, then by the triangle inequality we have

$$\begin{aligned}d_H(\mathbf{x},\mathbf{y}) &\leq d_H(\mathbf{x},\mathbf{z}) + d_H(\mathbf{y},\mathbf{z})\\ &\leq nB(p+d+2\epsilon)\end{aligned}$$

and therefore,

$$\begin{aligned}p(\mathbf{y}|\mathbf{x}) &\dot{\geq} p^{nB(p+d)}(1-p)^{nB-nB(p+d)}\\ &= 2^{-nB[H_2(p)+d\log\frac{1-p}{p}]}.\end{aligned}$$

Since we also trivially have $p(\mathbf{y}|\mathbf{x}) \geq p^{nB} = 2^{nB\log p}$, it follows that

$$p(\mathbf{y}|\mathbf{x}) \dot{\geq} 2^{-nB\min\left\{H_2(p)+d\log\frac{1-p}{p}, -\log p\right\}}. \tag{80}$$

Comparing (79) with (80), we have the maximum possible exponential decrease from $p(\mathbf{z}|\mathbf{x})$ to $p(\mathbf{y}|\mathbf{x})$ given by (78).

The above discussion reveals that the proof of Theorem 7 inherently uses the triangle inequality to obtain a worst case bound, equal to $nB(d+p)$, on the distance between $\mathbf{x}$ and $\mathbf{y}$. The new ingredient in the proof of Theorem 8 is a more precise analysis on the distance between $\mathbf{y}$ and $\mathbf{x}$ by building on the fact that the capacity of the primitive relay channel is equal to the broadcast bound in the context of Cover's open problem. Specifically, we show that most of the typical $\mathbf{x}$'s are within a distance $nB(d*p)$ from $\mathbf{y}$ in this case. The detailed proof of Theorem 8 is as follows, where we only emphasize the difference from that of Theorem 7.

We start by observing that Theorem 8 follows from the following proposition as a corollary.

*Proposition 19:* In the binary symmetric channel case, for any $\tau > 0$, if a rate $R = C_{XYZ} - \tau$ is achievable, then there exists some $a \geq 0$ such that

$$\begin{cases} C_{XYZ} - \tau \leq C_{XY} + R_0 - a & (81)\\ C_{XYZ} - \tau \leq C_{XY} + \Delta'\left(\sqrt{\dfrac{a\ln 2}{2}}\right) + \mu & (82)\end{cases}$$

where

$$\Delta'(d) := d(1-2p)\log\frac{1-p}{p} \tag{83}$$

and $\mu \to 0$ as $\tau \to 0$.

Specifically, we have by (81),

$$\begin{aligned}R_0 &\geq C_{XYZ} - C_{XY} + a - \tau\\ &= 1 + H_2(p*p) - 2H_2(p) - (1 - H_2(p)) + a - \tau\\ &= H_2(p*p) - H_2(p) + a - \tau, \end{aligned}\tag{84}$$

where we use the fact that $C_{XYZ} = 1 + H_2(p*p) - 2H_2(p)$ and $C_{XY} = 1 - H_2(p)$, and by (82),

$$\begin{aligned}\Delta'\left(\sqrt{\frac{a\ln 2}{2}}\right) &= \sqrt{\frac{a\ln 2}{2}}(1-2p)\log\frac{1-p}{p}\\ &\geq C_{XYZ} - C_{XY} - \tau - \mu\\ &= H_2(p*p) - H_2(p) - \tau - \mu\end{aligned}$$

so that

$$a \geq \frac{2}{\ln 2}\left(\frac{H_2(p*p) - H_2(p)}{(1-2p)\log\frac{1-p}{p}}\right)^2 - \mu_1 \tag{85}$$

for some $\mu_1 \to 0$ as $\tau \to 0$. Combining (84) and (85), we have

$$\begin{aligned}R_0 \geq\ & H_2(p*p) - H_2(p) + \frac{2}{\ln 2}\left(\frac{H_2(p*p) - H_2(p)}{(1-2p)\log\frac{1-p}{p}}\right)^2\\ &- \tau - \mu_1,\end{aligned}$$

and by the definition of $R_0^*$,

$$\begin{aligned}R_0^* &\geq \lim_{\tau\to 0} H_2(p*p) - H_2(p) + \frac{2}{\ln 2}\left(\frac{H_2(p*p)-H_2(p)}{(1-2p)\log\frac{1-p}{p}}\right)^2\\ &\quad - \tau - \mu_1\\ &= H_2(p*p) - H_2(p) + \frac{2}{\ln 2}\left(\frac{H_2(p*p) - H_2(p)}{(1-2p)\log\frac{1-p}{p}}\right)^2\end{aligned}$$

which is Theorem 8.

[5]Following [30], we say $a_m \doteq b_m$ if $\lim_{m\to 0}\frac{1}{m}\log\frac{a_m}{b_m} = 0$. Notations "$\dot{\geq}$" and "$\dot{\leq}$" are similarly defined.

We now show Proposition 19, whose proof builds on the technique developed to prove Theorem 7 but doesn't require fixed composition code analysis. To show (81), along the lines of the proof of (24), we have for any achievable rate $R = C_{XYZ} - \tau$, $\tau > 0$,

$$\begin{aligned} n(C_{XYZ} - \tau) &\leq I(X^n; Y^n) + nR_0 - na_n + n\epsilon \\ &\leq n(C_{XY} + R_0 - a_n + \epsilon), \end{aligned} \tag{86}$$

i.e.,

$$C_{XYZ} - \tau \leq C_{XY} + R_0 - a_n + \epsilon, \tag{87}$$

where (86) follows from the memoryless property of the channel and $\epsilon \to 0$ as $n \to \infty$. To show (82), we need the following lemma, whose proof is given in Section IX-A.

*Lemma 20:* In the binary symmetric channel case, for any $n$-channel use code with rate $R = C_{XYZ} - \tau$ and $P_e^{(n)} \to 0$,

$$\begin{aligned} H(Y^n|I_n) \leq{}& H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) \\ &+ n\Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right) + n\mu, \end{aligned} \tag{88}$$

where $\mu$ can be made arbitrarily small by choosing $n$ sufficiently large and $\tau$ sufficiently small, $H(Y|X) = H_2(p)$, $a_n = \frac{1}{n}H(I_n|X^n)$, and $\Delta'(\cdot)$ is as defined in (83).

With the above lemma, following the lines that lead to (57) from (56), we can show that for any achievable rate $R = C_{XYZ} - \tau$, $\tau > 0$,

$$\begin{aligned} n(C_{XYZ} - \tau) &\leq I(X^n; Y^n) + n\Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right) + n\mu + n\epsilon \\ &\leq n\left[C_{XY} + \Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right) + \mu + \epsilon\right], \end{aligned}$$

i.e.,

$$C_{XYZ} - \tau \leq C_{XY} + \Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right) + \mu + \epsilon, \tag{89}$$

where $\mu \to 0$ as $n \to \infty$ and $\tau \to 0$, and $\epsilon \to 0$ as $n \to \infty$. Combining (87) and (89) proves Proposition 19.

## A. *Proof of Lemma 20*

To show the entropy inequality (88) in Lemma 20, we again look at the $B$-length i.i.d. sequence of the $n$-letter random variables $(X^n, Y^n, Z^n, I_n)$ that are induced by the $n$-channel use reliable code of rate $C_{XYZ} - \tau$, denoted by

$$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{I}) := \{(X^n(b), Y^n(b), Z^n(b), I_n(b))\}_{b=1}^B.$$

The following Lemma 21 is crucial for establishing Lemma 20, which allows one to essentially only consider the $(\mathbf{y}, \mathbf{i})$ pairs belonging to the high probability set $\mathcal{S}(Y^n, I_n)$ with the desired property (90).

*Lemma 21:* Given any $\delta > 0$, for $\tau$ sufficiently small and $n, B$ sufficiently large, there exists a set $\mathcal{S}(Y^n, I_n)$ of $(\mathbf{y}, \mathbf{i})$ pairs such that

$$\Pr((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n)) \geq 1 - \delta,$$

and for any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$p(\mathbf{y}|\mathbf{i}) \geq 2^{-B\left(H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right) + n\delta\right)}. \tag{90}$$

With Lemma 21, along the same lines as in the proof of Lemma 17, we can show that

$$\begin{aligned} H(Y^n|I_n) \leq{}& H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) \\ &+ n\Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right) + n\delta + \delta n \log|\Omega| + \frac{1}{B}, \end{aligned}$$

where $\delta$ can be made arbitrarily small by choosing $\tau$ sufficiently small and $n, B$ sufficiently large. This proves the entropy inequality (88).

We are now in a position to show Lemma 21.

*Proof of Lemma 21:* The only difference of Lemma 21 from Lemma 18 is that here the lower bound on $p(\mathbf{y}|\mathbf{i}), \forall(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$ is sharpened to that of (90). In particular, assume $\mathcal{S}(Y^n, I_n)$ is defined exactly as in (61). Then, for any specific $(\mathbf{y}_0, \mathbf{i}_0) \in \mathcal{S}(Y^n, I_n)$, we can find one $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})$ such that

$$d_H(\mathbf{y}_0, \mathbf{z}_0) \leq nB\left(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon\right). \tag{91}$$

The key to the aforementioned sharpening is a tighter upper bound on the distance between $\mathbf{y}_0$ and the $\mathbf{x}$'s typical with $\mathbf{z}_0$, as stated in the following lemma. The proof of this lemma is based on a combinatorial geometric argument, and is deferred until we finish the proof of Lemma 21.

*Lemma 22:* Consider any $\mathbf{y}_0$ such that $d_H(\mathbf{y}_0, \mathbf{z}_0) = nBd_0$ for some $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$. There exists some $\epsilon' \to 0$ as $\epsilon \to 0$ such that

$$\Pr(d_H(\mathbf{X}, \mathbf{y}_0) \leq nB(d_0 * p + \epsilon')|\mathbf{z}_0) \geq 1 - \upsilon$$

where $\upsilon$ can be made arbitrarily small by choosing $n, B$ sufficiently large and $\tau$ sufficiently small.

Roughly speaking, Lemma 22 says that if the distance between $\mathbf{y}_0$ and $\mathbf{z}_0$ is $nBd_0$ and $\mathbf{z}_0$ is typical, then given $\mathbf{z}_0$ the typical $\mathbf{x}$'s are within a distance $nB(d_0 * p + \epsilon')$ from $\mathbf{y}_0$. Building on this, we further have for some $\epsilon' \to 0$ as $\epsilon \to 0$,

$$\begin{aligned} &\Pr\Big(\mathbf{X} \in \mathrm{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \\ &\qquad\qquad \bigcap \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)\Big|\mathbf{z}_0\Big) \\ &\geq 1 - \Pr(\mathbf{X} \notin \mathrm{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon'))|\mathbf{z}_0) \\ &\quad - \Pr(\mathbf{X} \notin \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)|\mathbf{z}_0) \\ &\geq 1 - \upsilon - \upsilon \\ &= 1 - 2\upsilon, \end{aligned}$$

where we have used the fact that

$$\Pr(\mathbf{X} \notin \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)|\mathbf{z}_0) \to 0 \text{ as } B \to \infty.$$

Since for any $\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)$, $p(\mathbf{x}|\mathbf{z}_0) \leq 2^{-B(H(X^n|Z^n) - \epsilon_1)}$ for some $\epsilon_1 \to 0$ as $\epsilon \to 0$, we can lower bound the number of $\mathbf{x}$'s that belong to

$\text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \bigcap \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)$ as follows:

$$\begin{aligned}&\left|\text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \bigcap \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)\right| \\ &\quad \geq (1-2\upsilon)2^{B(H(X^n|Z^n)-\epsilon_1)} \\ &\quad \geq 2^{B(H(X^n|Z^n)-\epsilon_1-\upsilon_1)}, \qquad (92)\end{aligned}$$

where $\upsilon_1 \to 0$ as $\tau \to 0$ and $n, B \to \infty$.

We now lower bound the conditional probability $p(\mathbf{y}_0|\mathbf{i}_0)$ for $(\mathbf{y}_0, \mathbf{i}_0) \in \mathcal{S}(Y^n, I_n)$. In particular, we have

$$\begin{aligned}p(\mathbf{y}_0|\mathbf{i}_0) &= \sum_{\mathbf{x}} p(\mathbf{y}_0|\mathbf{x})p(\mathbf{x}|\mathbf{i}_0) \\ &\geq \sum_{\substack{\mathbf{x} \in \text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \\ \bigcap \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)}} p(\mathbf{y}_0|\mathbf{x})p(\mathbf{x}|\mathbf{i}_0) \\ &\geq 2^{B(H(X^n|Z^n)-\epsilon_1-\upsilon_1)} 2^{-B(H(X^n|I_n)+\epsilon_2)} \\ &\quad \times \min_{\mathbf{x} \in \text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon'))} p(\mathbf{y}_0|\mathbf{x}) \qquad (93)\end{aligned}$$

for some $\epsilon_2 \to 0$ as $\epsilon \to 0$, where the equality follows from the law of total probability, and the second inequality follows because the number of $\mathbf{x}$'s belonging to

$$\text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \bigcap \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)$$

is lower bounded by (92) and for any $\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}_0, \mathbf{i}_0)$, $p(\mathbf{x}|\mathbf{i}_0)$ can be lower bounded by $2^{-B(H(X^n|I_n)+\epsilon_2)}$. Finally, to lower bound the last term in (93), we have for any $\mathbf{x}$ such that

$$\begin{aligned}d_H(\mathbf{x}, \mathbf{y}_0) &\leq nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon') \\ &=: nB(\sqrt{\frac{a_n \ln 2}{2}} * p + \epsilon_3),\end{aligned}$$

it follows that

$$\begin{aligned}p(\mathbf{y}_0|\mathbf{x}) &= (1-p)^{nB-d_H(\mathbf{x},\mathbf{y}_0)} p^{d_H(\mathbf{x},\mathbf{y}_0)} \\ &= (1-p)^{nB(1-p)} p^{nBp} \cdot \frac{(1-p)^{nB-d_H(\mathbf{x},\mathbf{y}_0)} p^{d_H(\mathbf{x},\mathbf{y}_0)}}{(1-p)^{nB(1-p)} p^{nBp}} \\ &= 2^{-nBH_2(p)} \cdot \left(\frac{p}{1-p}\right)^{d_H(\mathbf{x},\mathbf{y}_0)-nBp} \\ &\geq 2^{-nBH_2(p)} \cdot \left(\frac{p}{1-p}\right)^{nB(\sqrt{\frac{a_n \ln 2}{2}} * p + \epsilon_3)-nBp} \\ &= 2^{-nBH_2(p)} \cdot \left(\frac{p}{1-p}\right)^{nB(\sqrt{\frac{a_n \ln 2}{2}}(1-2p)+\epsilon_3)} \\ &= 2^{-nB\left(H_2(p) + (\sqrt{\frac{a_n \ln 2}{2}}(1-2p)+\epsilon_3)\log\frac{1-p}{p}\right)} \\ &= 2^{-nB\left(H_2(p) + \Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right)+\epsilon_4\right)} \qquad (94)\end{aligned}$$

where $\epsilon_3, \epsilon_4 \to 0$ as $\epsilon \to 0$. Plugging (94) into (93), we obtain that

$$\begin{aligned}p(\mathbf{y}_0|\mathbf{i}_0) &\geq 2^{-B\left[H(X^n|I_n) - H(X^n|Z^n) + nH_2(p) + n\Delta'\left(\sqrt{\frac{a_n \ln 2}{2}}\right)\right]} \\ &\quad \times 2^{-B(\upsilon_1+\epsilon_1+\epsilon_2+n\epsilon_4)},\end{aligned}$$

which proves the lemma. ∎

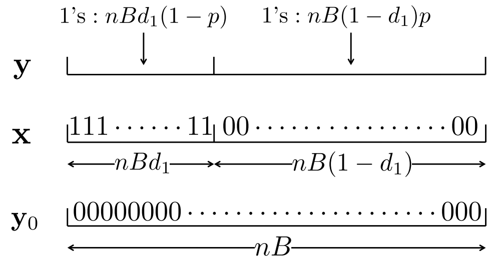


Fig. 5. Illustration of a specific pair $(\mathbf{x}, \mathbf{y}_0)$.

### *B. Proof of Lemma 22*

Consider a specific $(\mathbf{y}_0, \mathbf{z}_0)$ pair where $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and $d_H(\mathbf{y}_0, \mathbf{z}_0) = nBd_0$. Let

$$q_d := \Pr(d_H(\mathbf{X}, \mathbf{y}_0) \geq nBd|\mathbf{z}_0).$$

To show Lemma 22, we will show that $q_d$ diminishes for $d > d_0 * p$. More precisely, we will show that there exists some $\epsilon' \to 0$ as $\epsilon \to 0$ such that $d = d_0 * p + \epsilon'$ and $q_d \leq \upsilon$, where $\upsilon$ satisfies $\lim_{\tau\to 0}\lim_{n\to\infty}\lim_{B\to\infty} \upsilon = 0$. To this end, instead of directly upper bounding $q_d$, below we will look at the following intermediate probability involving $\mathbf{Y}$:

$$\begin{aligned}\Pr\Big(&d_H(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0), \\ &p(\mathbf{Y}|\mathbf{z}_0) \leq 2^{-B(H(Y^n|Z^n)-\epsilon_0)}, \\ &d_H(\mathbf{Y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)]\Big|\mathbf{z}_0\Big) \qquad (95)\end{aligned}$$

and show that it is essentially lower bounded by $q_d$; then via upper bounding this intermediate probability in (95) we can effectively upper bound $q_d$ as desired.

We first lower bound the probability in (95) in terms of $q_d$. Using the properties of jointly typical sequences, we can show (see Appendix J) that there exists some $\epsilon_0 \to 0$ as $\epsilon \to 0$ such that for any $\delta > 0$ and $B$ sufficiently large,

$$\begin{aligned}\Pr\Big(&p(\mathbf{Y}|\mathbf{z}_0) \leq 2^{-B(H(Y^n|Z^n)-\epsilon_0)}, \\ &\quad d_H(\mathbf{Y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)]\Big|\mathbf{z}_0\Big) \\ &\quad \geq 1 - \delta. \qquad (96)\end{aligned}$$

Consider the following inequalities:

$$\begin{aligned}&\Pr(d_H(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{z}_0) \\ &= \sum_{\mathbf{x}} \Pr(d_H(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{x})p(\mathbf{x}|\mathbf{z}_0) \\ &\geq \sum_{\mathbf{x}: d_H(\mathbf{x},\mathbf{y}_0) \geq nBd} \Pr(d_H(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{x})p(\mathbf{x}|\mathbf{z}_0) \\ &\geq q_d \cdot \min_{\mathbf{x}: d_H(\mathbf{x},\mathbf{y}_0) \geq nBd} \Pr(d_H(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{x}). \qquad (97)\end{aligned}$$

To bound the second term in (97), without loss of generality consider a specific pair $(\mathbf{x}, \mathbf{y}_0)$ as shown in Fig. 5, where $\mathbf{y}_0 = \mathbf{0}$ and $d_H(\mathbf{x}, \mathbf{y}_0) = nBd_1 \geq nBd$. By the law of large numbers, we have for any $\delta > 0$ and $B$ sufficiently large,

$$\begin{aligned}1-\delta &\le \Pr\bigg(\frac{1}{nB}N(0,1|\mathbf{x},\mathbf{Y}) \ge \frac{1}{nB}N(0|\mathbf{x})p - \frac{\epsilon_0}{2},\\ &\qquad \frac{1}{nB}N(1,1|\mathbf{x},\mathbf{Y}) \ge \frac{1}{nB}N(1|\mathbf{x})(1-p) - \frac{\epsilon_0}{2}\bigg|\mathbf{x}\bigg)\\ &\le \Pr(N(1|\mathbf{Y}) \ge nB(d_1 * p - \epsilon_0)|\mathbf{x})\\ &\le \Pr(N(1|\mathbf{Y}) \ge nB(d * p - \epsilon_0)|\mathbf{x})\\ &= \Pr(d_H(\mathbf{Y},\mathbf{y}_0) \ge nB(d * p - \epsilon_0)|\mathbf{x}), \qquad (98)\end{aligned}$$

where (98) follows since the event $N(1|\mathbf{Y}) \ge nB(d_1 * p - \epsilon_0)$ implies $N(1|\mathbf{Y}) \ge nB(d * p - \epsilon_0)$ due to the relation $d_1 \ge d$. Plugging this into (97), we obtain

$$\Pr(d_H(\mathbf{Y},\mathbf{y}_0) \ge nB(d * p - \epsilon_0)|\mathbf{z}_0) \ge q_d(1-\delta). \qquad (99)$$

Combining (96) and (99), we have for any $\delta > 0$ and $B$ sufficiently large,

$$\begin{aligned}&\Pr\Big(d_H(\mathbf{Y},\mathbf{y}_0) \ge nB(d * p - \epsilon_0),\\ &\qquad p(\mathbf{Y}|\mathbf{z}_0) \le 2^{-B(H(Y^n|Z^n)-\epsilon_0)},\\ &\qquad d_H(\mathbf{Y},\mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)]\Big|\mathbf{z}_0\Big)\\ &\ge 1 - (\delta + 1 - q_d(1-\delta))\\ &= q_d(1-\delta) - \delta\\ &\ge q_d - 2\delta.\end{aligned}$$

On the other hand,

$$\begin{aligned}&\Pr\Big(d_H(\mathbf{Y},\mathbf{y}_0) \ge nB(d * p - \epsilon_0),\\ &\qquad p(\mathbf{Y}|\mathbf{z}_0) \le 2^{-B(H(Y^n|Z^n)-\epsilon_0)},\\ &\qquad d_H(\mathbf{Y},\mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)]\Big|\mathbf{z}_0\Big)\\ &\le |\{\mathbf{y} : d_H(\mathbf{y},\mathbf{y}_0) \ge nB(d * p - \epsilon_0),\\ &\qquad d_H(\mathbf{y},\mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)]\}|\\ &\qquad \times 2^{-B(H(Y^n|Z^n)-\epsilon_0)}\\ &= \left|\bigcup_{r=d*p-\epsilon_0}^{1} \mathrm{Sphere}(\mathbf{y}_0, nBr) \bigcap \bigcup_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} \mathrm{Sphere}(\mathbf{z}_0, nB\rho)\right|\\ &\qquad \times 2^{-B(H(Y^n|Z^n)-\epsilon_0)}\\ &= \left|\bigcup_{r=d*p-\epsilon_0}^{1} \bigcup_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} \underbrace{\mathrm{Sphere}(\mathbf{y}_0, nBr) \bigcap \mathrm{Sphere}(\mathbf{z}_0, nB\rho)}_{\mathrm{Inter}(r,\rho)}\right|\\ &\qquad \times 2^{-B(H(Y^n|Z^n)-\epsilon_0)}.\end{aligned}$$

Therefore,

$$q_d \le 2\delta + 2^{-B(H(Y^n|Z^n)-\epsilon_0)} \left|\bigcup_{r=d*p-\epsilon_0}^{1} \bigcup_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} \mathrm{Inter}(r,\rho)\right|. \qquad (100)$$

Now we show that the second term on the R.H.S. of (100) vanishes if $d > d_0 * p$. Without loss of generality, consider a specific pair $(\mathbf{y}_0, \mathbf{z}_0)$ as shown in Fig. 6, where $\mathbf{y}_0 = \mathbf{0}$ and $d_H(\mathbf{y}_0, \mathbf{z}_0) = nd_0$. We first characterize the volume of $\mathrm{Inter}(r, \rho)$ for any $r$ and $\rho = p * p$, i.e.,

$$\left|\mathrm{Sphere}(\mathbf{y}_0, nBr) \bigcap \mathrm{Sphere}(\mathbf{z}_0, nBp * p)\right|. \qquad (101)$$

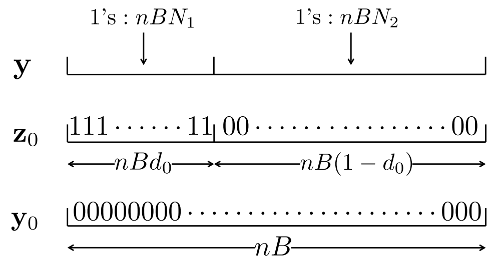


Fig. 6. Illustration of a specific pair $(\mathbf{y}_0, \mathbf{z}_0)$.

For any $\mathbf{y}$, let $nBN_1$ denote the number of 1's appearing in its first $nBd_0$ digits, and $nBN_2$ denote the number of 1's in the rest. Then the volume in (101) amounts to the number of $\mathbf{y}$'s such that the following two equalities hold

$$\begin{cases} N_1 + N_2 = r \\ d_0 - N_1 + N_2 = p * p \end{cases}$$

i.e.,

$$\begin{cases} N_1 = \dfrac{r + d_0 - p * p}{2} \\ N_2 = \dfrac{r + p * p - d_0}{2}. \end{cases}$$

Here, $N_1 \in [0, d_0]$ and $N_2 \in [0, (1 - d_0)]$, i.e.,

$$\begin{cases} r \in [p * p - d_0, p * p + d_0] & (102)\\ r \in [d_0 - p * p, 2 - p * p - d_0]. & (103)\end{cases}$$

Therefore, with $\rho = p * p$,

$$\begin{aligned}\mathrm{Inter}(r,\rho) &= \binom{nBd_0}{nB\frac{r+d_0-p*p}{2}}\binom{nB(1-d_0)}{nB\frac{r+p*p-d_0}{2}}\\ &\le 2^{nBd_0\left(H_2(\frac{r+d_0-p*p}{2d_0})\right)}2^{nB(1-d_0)\left(H_2(\frac{r+p*p-d_0}{2(1-d_0)})\right)}\\ &= 2^{nB\left(d_0H_2(\frac{r+d_0-p*p}{2d_0})+(1-d_0)H_2(\frac{r+p*p-d_0}{2(1-d_0)})\right)} \qquad (104)\end{aligned}$$

for sufficiently large $B$, where (104) follows from the bound $\binom{n}{nk} \le \frac{1}{\sqrt{n\pi k(1-k)}}2^{nH_2(k)}$ for any $k \in (0, 1)$, as stated in [30, Lemma 17.5.1].

Let $f(r) = d_0H_2(\frac{r+d_0-p*p}{2d_0}) + (1 - d_0)H_2(\frac{r+p*p-d_0}{2(1-d_0)})$. It can be verified that $f(r)$ attains the maximum $H_2(p * p)$ if and only if $r = d_0 * p * p$; see Appendix K. Thus, when

$$\begin{aligned}\rho &= p * p\\ d &= d_0 * p + \epsilon'\\ r &\ge d * p - \epsilon_0\\ &= (d_0 * p + \epsilon') * p - \epsilon_0\\ &= d_0 * p * p + \epsilon'(1 - 2p) - \epsilon_0,\end{aligned}$$

we have

$$\mathrm{Inter}(r,\rho) \le 2^{nB(H_2(p*p)-\epsilon_1)}$$

for some $\epsilon_1 > 0$ provided $\epsilon'(1-2p)-\epsilon_0 > 0$. Further, due to the continuity of $\mathrm{Inter}(r, \rho)$ in $\rho$, for any $\rho \in [p * p - \epsilon_0, p * p + \epsilon_0]$, $d = d_0 * p + \epsilon'$, $r \ge d * p - \epsilon_0$,

$$\mathrm{Inter}(r,\rho) \le 2^{nB(H_2(p*p)-\epsilon_1+\epsilon_2)} \qquad (105)$$

for some $\epsilon_2 \to 0$ as $\epsilon_0 \to 0$.

Plugging (105) into (100), we have for $d = d_0 * p + \epsilon'$ and sufficiently large $B$,

$$\begin{aligned} q_d &\le 2\delta + 2^{-B(H(Y^n|Z^n)-\epsilon_0)} \\ &\quad \times \sum_{r=d*p-\epsilon_0}^{1} \sum_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} 2^{nB(H_2(p*p)-\epsilon_1+\epsilon_2)} \\ &\le 2\delta + 2^{-B(H(Y^n|Z^n)-\epsilon_0)} 2^{nB(H_2(p*p)-\epsilon_1+\epsilon_0+\epsilon_2)} \\ &\le 2\delta + 2^{-nB(\frac{1}{n}H(Y^n|Z^n)-H_2(p*p)+\epsilon_1-2\epsilon_0-\epsilon_2)}, \end{aligned}$$

with

$$\begin{aligned} \frac{1}{n}H(Y^n|Z^n) &= \frac{1}{n}(H(Y^n, Z^n) - H(Z^n)) \\ &= \frac{1}{n}(H(X^n) + H(Y^n, Z^n|X^n) \\ &\qquad - H(X^n|Y^n, Z^n) - H(Z^n)) \\ &= \frac{1}{n}(H(M) - H(M|X^n) + H(Y^n, Z^n|X^n) \\ &\qquad - H(X^n|Y^n, Z^n) - H(Z^n)) \\ &\ge \frac{1}{n}(nR - n\epsilon_0 + 2nH_2(p) - n\epsilon_0 - n) \\ &= C_{XYZ} - \tau + 2H_2(p) - 1 - 2\epsilon_0 \\ &= H_2(p * p) - \tau - 2\epsilon_0, \end{aligned} \tag{106}$$

for $n$ sufficiently large, where in (106) we have used Fano's inequality. Thus, when $\tau, \epsilon_0$ are sufficiently small and $n$, $B$ are sufficiently large, we have for any $\epsilon' > 0$, $d = d_0 * p + \epsilon'$,

$$\begin{aligned} q_d &\le 2\delta + 2^{-nB(\epsilon_1-\tau-4\epsilon_0-\epsilon_2)} \\ &\le 3\delta. \end{aligned}$$

We finally conclude that for a $(\mathbf{y}_0, \mathbf{z}_0)$ pair where $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and $d_H(\mathbf{y}_0, \mathbf{z}_0) = nBd_0$, there exists some $\epsilon' \to 0$ as $\epsilon \to 0$ such that

$$\Pr(d_H(\mathbf{X}, \mathbf{y}_0) \le nB(d_0 * p + \epsilon')|\mathbf{z}_0) \ge 1 - \upsilon.$$

where $\upsilon$ can be made arbitrarily small by choosing $n$, $B$ sufficiently large and $\tau$ sufficiently small.

## X. Conclusion

We consider the symmetric primitive relay channel, and develop two new upper bounds on its capacity that are tighter than existing bounds, including the celebrated cut-set bound. Our approach uses measure concentration (the blowing-up lemma in particular) to analyze the probabilistic geometric relations between the typical sets of the $n$-letter random variables associated with a reliable code for communicating over this channel. We then translate these relations to new entropy inequalities between the $n$-letter random variables involved.

Information theory and geometry are indeed known to be inherently related; for example the differential entropy of a continuous random variable can be regarded as the exponential growth rate of the volume of its typical set. Therefore, entropy relations can, in principle, be developed by studying the relative geometry of the typical sets of the random variables. However, we are not aware of many examples where such geometric techniques have been successfully used to develop converses for problems in network information theory. It would be interesting to see if the approach we develop in this paper, i.e. deriving information inequalities by studying the geometry of typical sets, in particular using measure concentration, can be used to make progress on other long-standing open problems in network information theory.

While we have exclusively focused on the symmetric relay channel in this paper, our results can be extended to asymmetric primitive relay channels [32] using the idea of channel simulation. An extension of these ideas to the Gaussian case has been provided in [33].

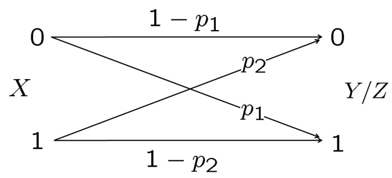


Fig. 7. Binary asymmetric channel.

## Appendix A

To see $E(R) \le R - C_{XY}$ for any $R > I(X;Y)$, recall that $E(R)$ has the following alternative form [28]:

$$\begin{aligned} E(R) = \min_{p(x)} \min_{\tilde{p}(y|x)} & D(\tilde{p}(y|x)||p(y|x)|p(x)) \\ & +|R - I(p(x), \tilde{p}(y|x))|^+ \end{aligned} \tag{107}$$

where $|t|^+ := \max\{0, t\}$, $D(\tilde{p}(y|x)||p(y|x)|p(x))$ is the conditional relative entropy defined as

$$D(\tilde{p}(y|x)||p(y|x)|p(x)) := \sum_{(x,y)} p(x)\tilde{p}(y|x) \log \frac{\tilde{p}(y|x)}{p(y|x)},$$

and $I(p(x), \tilde{p}(y|x))$ is the mutual information defined with respect to the joint distribution $p(x)\tilde{p}(y|x)$, i.e.,

$$I(p(x), \tilde{p}(y|x)) := \sum_{(x,y)} p(x)\tilde{p}(y|x) \log \frac{\tilde{p}(y|x)}{\sum_x p(x)\tilde{p}(y|x)}.$$

In the regime of $R > C_{XY}$, simply choosing the $p(x)$ and $\tilde{p}(y|x)$ in (107) to be capacity-achieving distribution $p^*(x)$ and $p(y|x)$ respectively would make the objective function equal to $R - C_{XY}$, and thus $E(R) \le R - C_{XY}$.

## Appendix B

We demonstrate the improvements of our bound in Theorem 4 over Xue's bound using the following simple example.

*Example 23:* Suppose both $X$-$Y$ and $X$-$Z$ links are the binary asymmetric channels as depicted in Fig. 7, with parameters $p_1 = 0.01$ and $p_2 = 0.3$. For the input distribution

$$p(x) = \begin{cases} \alpha & x = 0 \\ 1 - \alpha & x = 1 \end{cases}$$

we have

$$\begin{aligned} I(X;Y) = {} & H_2(\alpha(1 - p_1) + (1 - \alpha)p_2) \\ & -(\alpha H_2(p_1) + (1 - \alpha)H_2(p_2)) \end{aligned}$$

and

$$\begin{aligned}I(X;Y,Z) = H([&\alpha(1-p_1)^2+(1-\alpha)p_2^2,\\ &\alpha(1-p_1)p_1+(1-\alpha)(1-p_2)p_2,\\ &\alpha(1-p_1)p_1+(1-\alpha)(1-p_2)p_2,\\ &\alpha p_1^2+(1-\alpha)(1-p_2)^2])\\ &-2(\alpha H_2(p_1)+(1-\alpha)H_2(p_2)).\end{aligned}$$

With $p_1 = 0.01$ and $p_2 = 0.3$, numerical evaluation of $I(X;Y)$ and $I(X;Y,Z)$ yields that $C_{XY} = \max_\alpha I(X;Y) = 0.46432$ with the maximizer $\alpha^*_{XY} = 0.58$, and $C_{XYZ} = \max_\alpha I(X;Y,Z) = 0.72022$ with the maximizer $\alpha^*_{XYZ} = 0.54$.

Suppose we want to achieve a rate $R = C_{XYZ}$, and we use Proposition 3 and Theorem 4 to derive a lower bound on $R_0$, respectively. First consider Proposition 3. Numerically, we have $E(R) = 0.05951$ for $R = C_{XYZ} = 0.72022$, and the minimum $a$ to satisfy (6) is $a = 0.00008$. Thus, by (5), we have

$$\begin{aligned}R_0 &\geq R - C_{XY} + a\\ &\geq 0.72022 - 0.46432 + 0.00008\\ &= 0.25598. \qquad (108)\end{aligned}$$

We then apply Theorem 4 and demonstrate that the improvements mentioned in Section IV-B result in a tighter bound on $R_0$. In Theorem 4, $p(x)$ has to be chosen such that $\alpha = \alpha^*_{XYZ} = 0.54$ due to the constraint (14). Under such a distribution of $p(x)$, numerically, we have $I(X;Y) = 0.46223 < C_{XY}$, and $C_{XYZ} - I(X;Y) = 0.25799 > E(R)$. Noting the R.H.S. of (6) is also sharpened to that of (17), we can calculate the minimum $a$ satisfying (17) to be $a = 0.00546$. Thus, by (15), we have

$$\begin{aligned}R_0 &\geq R - I(X;Y) + a\\ &\geq 0.72022 - 0.46223 + 0.00546\\ &= 0.26345, \qquad (109)\end{aligned}$$

where it is easy to see that the last two terms in (109) are both sharpened compared to those in (108).

Therefore, in order to achieve the rate $R = C_{XYZ}$, the lower bounds on $R_0$ yielded by Proposition 3 and Theorem 4 are

$$R_0 \geq 0.25598$$

and

$$R_0 \geq 0.26345$$

respectively. Viewed from another perspective, for $R_0 \in [0.25598, 0.26345)$, the bound in Theorem 4 asserts that the capacity of the relay channel $C(R_0) < C_{XYZ}$, which excludes the possibility of achieving $R = C_{XYZ}$ while Xue's bound in Proposition 3 cannot.

## Appendix C
## $\Delta\,(p(x), d)$ for Binary Symmetric Channels

For a binary symmetric channel with crossover probability $p < 1/2$, the objective function in (18) can be expressed as

$$\begin{aligned}&H(\tilde{p}(\omega|x)|p(x)) + D(\tilde{p}(\omega|x)||p(\omega|x)|p(x))\\ &\quad -H(p(\omega|x)|p(x))\\ &= -\sum_{(x,\omega)} p(x)\tilde{p}(\omega|x)\log\tilde{p}(\omega|x)\\ &\quad +\sum_{(x,\omega)} p(x)\tilde{p}(\omega|x)\log\frac{\tilde{p}(\omega|x)}{p(\omega|x)}\\ &\quad -\sum_{(x,\omega)} p(x)p(\omega|x)\log\frac{1}{p(\omega|x)}\\ &= \sum_{(x,\omega)}[p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)]\log\frac{1}{p(\omega|x)}\\ &= \sum_{x\neq\omega}\big[p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)\big]\log\frac{1}{p}\\ &\quad +\sum_{x=\omega}[p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)]\log\frac{1}{1-p} \qquad (110)\\ &= \sum_{x\neq\omega}\big[p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)\big]\log\frac{1}{p}\\ &\quad +\sum_{x\neq\omega}[p(x)(1-\tilde{p}(\omega|x)) - p(x)(1-p(\omega|x))]\log\frac{1}{1-p} \qquad (111)\\ &= \sum_{x\neq\omega}\big[p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)\big]\log\frac{1}{p}\\ &\quad -\sum_{x\neq\omega}[p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)]\log\frac{1}{1-p} \qquad (112)\\ &= \sum_{x\neq\omega}\big[p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)\big]\log\frac{1-p}{p}. \qquad (113)\end{aligned}$$

We now show that under the constraint

$$\frac{1}{2}\sum_{(x,\omega)}|p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq d, \qquad (114)$$

the function in (113), and thus $\Delta\,(p(x), d)$, are upper bounded by

$$\min\{d, 1-p\}\log\frac{1-p}{p}.$$

Along the similar lines as in (110)–(112), we obtain

$$\begin{aligned}&\sum_{x=\omega}|p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)|\\ &\quad = \sum_{x\neq\omega}|p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)|,\end{aligned}$$

and thus the constraint (114) can be rewritten as

$$\sum_{x\neq\omega}|p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq d. \qquad (115)$$

On the other hand, we have

$$\begin{aligned}\sum_{x\neq\omega}&|p(x)\tilde{p}(\omega|x)-p(x)p(\omega|x)|\\&=\sum_{x\neq\omega}p(x)|\tilde{p}(\omega|x)-p(\omega|x)|\\&\leq\sum_{x\neq\omega}p(x)(1-p)\\&=1-p.\end{aligned}\tag{116}$$

Combining (113), (115) and (116) yields that

$$\Delta\left(p(x),d\right)\leq\min\{d,1-p\}\log\frac{1-p}{p}.$$

In fact, it can be easily checked that the equality sign in the above inequality can be attained by choosing

$$\tilde{p}(\omega|x)=p+\min\{d,1-p\}\,,\forall(x,\omega)\text{ with }x\neq\omega,$$

and thus we conclude that

$$\Delta\left(p(x),d\right)=\min\{d,1-p\}\log\frac{1-p}{p}.$$

## APPENDIX D
## UPPER BOUNDS FOR BINARY SYMMETRIC CHANNEL CASE

Various upper bounds are evaluated for the binary symmetric channel case as follows.

### A. Cut-Set Bound (Prop. 1)

The optimal distribution for Prop. 1 is $p^*(0)=p^*(1)=1/2$, under which,

$$\begin{cases}I^*(X;Y,Z)=C_{XYZ}=1+H_2(p*p)-2H_2(p)\\I^*(X;Y)=C_{XY}=1-H_2(p).\end{cases}$$

Therefore, the cut-set bound simplifies to

$$C(R_0)\leq\min\{1+H_2(p*p)-2H_2(p),1-H_2(p)+R_0\}\,.$$

### B. Xue's Bound (Prop. 3)

Since the function $E(R)$ is monotonic in $R$, its inverse function $E^{-1}(\cdot)$ exists and Xue's bound can be expressed as

$$C(R_0)\leq\max_{a\in[0,R_0]}\min\Big\{1-H_2(p)+R_0-a,\\E^{-1}(H_2(\sqrt{a})+\sqrt{a})\Big\}.$$

### C. Our First Bound (Thm. 4)

The uniform distribution of $X$ is also optimal for Thm. 4, under which our first bound reduces to

$$\begin{aligned}C(R_0)\leq&\max_{a\in[0,\min\{R_0,H_2(p),\frac{1}{2\ln 2}\}]}\\&\min\Big\{1+H_2(p*p)-2H_2(p),\\&1-H_2(p)+R_0-a,\\&1-H_2(p)+H_2\left(\sqrt{\frac{a\ln 2}{2}}\right)-a\Big\}.\end{aligned}$$

where the constraint of $a$ follows since $H(Z|X)=H_2(p)$ for any $p(x)$ and $\frac{2}{\ln 2}\left(\frac{|\Omega|-1}{|\Omega|}\right)^2=\frac{1}{2\ln 2}$, and the term $\sqrt{\frac{a\ln 2}{2}}\log(|\Omega|-1)$ disappears compared to Thm. 4 since it becomes 0 with $|\Omega|=2$.

### D. Our Second Bound (Thm. 7)

Recall that in the binary symmetric channel case, $\Delta\left(p(x),d\right)$ is independent of $p(x)$ and given by

$$\min\{d,1-p\}\log\frac{1-p}{p}:=\bar{\Delta}\left(d\right).$$

Therefore, our new bound becomes

$$\begin{aligned}C(R_0)\leq&\max_{a\in[0,\min\{R_0,H_2(p)\}]}\\&\min\Big\{1+H_2(p*p)-2H_2(p),\\&1-H_2(p)+R_0-a,\\&1-H_2(p)+\bar{\Delta}\left(\sqrt{\frac{a\ln 2}{2}}\right)\Big\}.\end{aligned}\tag{117}$$

It is not difficult to see that for the R.H.S. of (117), at least one of the maximizers must be no greater than $\frac{2}{\ln 2}(1-p)^2$, i.e., satisfying $\sqrt{\frac{a\ln 2}{2}}\leq 1-p$. Therefore, (117) can be equivalently stated as

$$\begin{aligned}C(R_0)\leq&\max_{a\in[0,\min\{R_0,H_2(p),\frac{2}{\ln 2}(1-p)^2\}]}\\&\min\Big\{1+H_2(p*p)-2H_2(p),\\&1-H_2(p)+R_0-a,\\&1-H_2(p)+\sqrt{\frac{a\ln 2}{2}}\log\frac{1-p}{p}\Big\}.\end{aligned}$$

## APPENDIX E
## PROOF OF (26)

To show (26), it suffices to show $H_2(p)/H_2(p*p)\to 1/2$ as $p\to 0$. For this, we have

$$\begin{aligned}\lim_{p\to 0}\frac{H_2(p)}{H_2(p*p)}&=\lim_{p\to 0}\frac{H_2'(p)}{H_2'(p*p)\cdot(p*p)'}\\&=\lim_{p\to 0}\frac{\log\frac{1-p}{p}}{\log\frac{1-p*p}{p*p}\cdot(2-4p)}\\&=\frac{1}{2}\lim_{p\to 0}\frac{\log'\frac{1-p}{p}\cdot(\frac{1-p}{p})'}{\log'\frac{1-p*p}{p*p}\cdot(\frac{1-p*p}{p*p})'}\\&=\frac{1}{2}\lim_{p\to 0}\frac{\frac{p}{1-p}\cdot(-\frac{1}{p^2})}{\frac{p*p}{1-p*p}\cdot(-\frac{1}{(p*p)^2})\cdot(2-4p)}\\&=\frac{1}{4}\lim_{p\to 0}\frac{(p*p)(1-p*p)}{p(1-p)}\\&=\frac{1}{4}\lim_{p\to 0}2(1-p*p)\\&=\frac{1}{2}\end{aligned}$$

which proves (26).

## APPENDIX F
## PROOF OF LEMMA 12

We only characterize $H(Y^n|X^n)$ and $I(X^n;Y^n)$ while the other information quantities can be characterized similarly. Consider $H(Y^n|x^n)$ for any specific $x^n$ with composition $Q_n$. We have

$$\begin{aligned}H(Y^n|x^n) &= \sum_{i=1}^{n} H(Y_i|x^n, Y^{i-1}) \\ &= \sum_{i=1}^{n} H(Y_i|x_i) \\ &= \sum_{x} nQ_n(x)H(Y|x) \\ &= nH(Y|X)\end{aligned}$$

where $H(Y|X)$ is calculated based on $Q_n(x)p(\omega|x)$. Therefore, for the code with fixed composition $Q_n$,

$$\begin{aligned}H(Y^n|X^n) &= \sum_{x^n} p(x^n)H(Y^n|x^n) \\ &= \sum_{x^n} p(x^n)[nH(Y|X)] \\ &= nH(Y|X).\end{aligned}$$

To bound $I(X^n;Y^n)$, it suffices to bound $H(Y^n)$. For any specific $x^n$ with composition $Q_n$, we have for some $\epsilon_1 \to 0$ as $n \to \infty$,

$$\Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)|x^n) \geq 1-\epsilon_1,$$

where $\mathcal{T}_{\epsilon_1}^{(n)}(Y)$ is the typical set with respect to $\sum_x Q_n(x)p(\omega|x)$. Therefore, for the code with fixed composition $Q_n$,

$$\begin{aligned}\Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)) &= \sum_{x^n} p(x^n)\Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)|x^n) \\ &\geq 1-\epsilon_1.\end{aligned}$$

Letting $W = \mathbb{I}(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y))$, we have

$$\begin{aligned}H(Y^n) &\leq H(Y^n, W) \\ &\leq 1 + H(Y^n|W) \\ &= 1 + \Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y))H(Y^n|Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)) \\ &\quad + \Pr(Y^n \notin \mathcal{T}_{\epsilon_1}^{(n)}(Y))H(Y^n|Y^n \notin \mathcal{T}_{\epsilon_1}^{(n)}(Y)) \\ &\leq 1 + \log|\mathcal{T}_{\epsilon_1}^{(n)}(Y)| + n\epsilon_1 \log|\Omega| \\ &\leq 1 + n(H(Y)+\epsilon_2) + n\epsilon_1 \log|\Omega| \\ &\leq n(H(Y)+\epsilon)\end{aligned}$$

where $\epsilon_1, \epsilon_2, \epsilon \to 0$ as $n \to \infty$, and $H(Y)$ is calculated based on $\sum_x Q_n(x)p(\omega|x)$. Combining this with the fact that $H(Y^n|X^n) = nH(Y|X)$, we have

$$I(X^n;Y^n) \leq n(I(X;Y)+\epsilon). \tag{118}$$

We now argue that the $\epsilon$ in (118) can be dropped. Given any fixed $n$, consider a length-$B$ sequence of i.i.d. random vector pairs $\{(X^n(b), Y^n(b))\}_{b=1}^{B}$, denoted by $(\mathbf{X}, \mathbf{Y})$, where $(X^n(b), Y^n(b))$ have the same distribution as $(X^n, Y^n)$ for any $b \in [1:B]$. Obviously the length-$nB$ vector $\mathbf{X}$ also has composition $Q_n$, and by (118) we have

$$I(\mathbf{X};\mathbf{Y}) \leq nB(I(X;Y)+\epsilon),$$

where $\epsilon \to 0$ as $B \to \infty$. Due to the i.i.d. property, we further have

$$BI(X^n;Y^n) \leq nB(I(X;Y)+\epsilon).$$

Dividing $B$ at both sides of the above equation and letting $B \to 0$, we obtain $I(X^n;Y^n) \leq nI(X;Y)$.

## APPENDIX G
## VOLUME OF A HAMMING BALL

Consider the volume of a general $n$-dimensional Hamming ball in $\Omega^n$ with radius $nr$. It is obvious that when $r \geq 1$, the volume

$$|\mathrm{Ball}(nr)| = |\Omega|^n = 2^{n\log|\Omega|}. \tag{119}$$

For $r < 1$, we have

$$\begin{aligned}|\mathrm{Ball}(nr)| &= 1 + \sum_{k=\frac{1}{n}}^{r} \binom{n}{nk}(|\Omega|-1)^{nk} \\ &\leq 1 + \sum_{k=\frac{1}{n}}^{r} \frac{2^{nH_2(k)}}{\sqrt{\pi nk(1-k)}}(|\Omega|-1)^{nk} \\ &\leq \sum_{k=0}^{r} 2^{n(H_2(k)+k\log(|\Omega|-1))} \\ &\leq (nr+1)\max_{k\in\{0,\frac{1}{n},\ldots,r\}} 2^{n(H_2(k)+k\log(|\Omega|-1))} \\ &\leq \max_{k\in\{0,\frac{1}{n},\ldots,r\}} 2^{n(H_2(k)+k\log(|\Omega|-1)+\epsilon)} \\ &= 2^{n\left(\max_{k\in\{0,\frac{1}{n},\ldots,r\}} H_2(k)+k\log(|\Omega|-1)+\epsilon\right)}\end{aligned}$$

for any $\epsilon > 0$ and sufficiently large $n$, where the first inequality follows from Stirling's formula [30, Lemma 17.5.1]. Moreover, by [30, Lemma 17.5.1] we also have for $k = \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}$,

$$\binom{n}{nk} \geq \frac{2^{nH_2(k)}}{\sqrt{8nk(1-k)}},$$

and therefore we can similarly lower bound $|\mathrm{Ball}(nr)|$ as

$$|\mathrm{Ball}(nr)| \geq 2^{n\left(\max_{k\in\{0,\frac{1}{n},\ldots,r\}} H_2(k)+k\log(|\Omega|-1)-\epsilon\right)}$$

for any $\epsilon > 0$ and sufficiently large $n$. Therefore, we have for $r < 1$,

$$\lim_{n\to\infty} \frac{1}{n}\log|\mathrm{Ball}(nr)| = \max_{t\in[0,r]} H_2(t) + t\log(|\Omega|-1). \tag{120}$$

Now we simplify the above expression. Let $\upsilon(t) = H_2(t) + t\log(|\Omega|-1)$ for $t \in (0,1)$. We have $\upsilon'(t) = \log\frac{(1-t)(|\Omega|-1)}{t}$, which is decreasing in $t$ for $t \in (0,1)$ and equals 0 when

$t = \frac{|\Omega|-1}{|\Omega|}$. Thus, the maximum of $v(t)$ is attained when $t = \frac{|\Omega|-1}{|\Omega|}$, and is given by

$$\begin{aligned} v^*(t) &= H_2\left(\frac{|\Omega|-1}{|\Omega|}\right) + \frac{|\Omega|-1}{|\Omega|}\log(|\Omega|-1) \\ &= -\frac{|\Omega|-1}{|\Omega|}\log\frac{|\Omega|-1}{|\Omega|} - \frac{1}{|\Omega|}\log\frac{1}{|\Omega|} \\ &\quad + \frac{|\Omega|-1}{|\Omega|}\log(|\Omega|-1) \\ &= \frac{|\Omega|-1}{|\Omega|}\log\frac{(|\Omega|-1)|\Omega|}{(|\Omega|-1)} - \frac{1}{|\Omega|}\log\frac{1}{|\Omega|} \\ &= \frac{|\Omega|-1}{|\Omega|}\log|\Omega| + \frac{1}{|\Omega|}\log|\Omega| \\ &= \log|\Omega|. \end{aligned} \tag{121}$$

Therefore, we have

$$\max_{t\in[0,r]} H_2(t) + t\log(|\Omega|-1) = \begin{cases} \log|\Omega| & \text{when } r \in \left(\frac{|\Omega|-1}{|\Omega|}, 1\right) \\ H_2(r) + r\log(|\Omega|-1) & \text{when } r \in \left[0, \frac{|\Omega|-1}{|\Omega|}\right]. \end{cases} \tag{122}$$

Combining (119), (120) and (122), we obtain that

$$\lim_{n\to\infty}\frac{1}{n}\log|\mathrm{Ball}(nr)| = \begin{cases} \log|\Omega| & \text{when } r > \frac{|\Omega|-1}{|\Omega|} \\ H_2(r) + r\log(|\Omega|-1) & \text{when } r \le \frac{|\Omega|-1}{|\Omega|}. \end{cases}$$

## APPENDIX H

For any $(\mathbf{x}, \mathbf{z}) \in \mathcal{T}_\epsilon^{(B)}(X^n, Z^n)$, we have

$$|P_{\mathbf{x}}(x^n) - p(x^n)| \le \epsilon p(x^n), \quad \forall x^n$$
$$|P_{(\mathbf{x},\mathbf{z})}(x^n, z^n) - p(x^n, z^n)| \le \epsilon p(x^n, z^n), \quad \forall (x^n, z^n)$$

and thus

$$|P_{\mathbf{z}|\mathbf{x}}(z^n|x^n) - p(z^n|x^n)| \le \epsilon_1 p(z^n|x^n), \quad \forall (x^n, z^n) \text{ with } P_{\mathbf{x}}(x^n) \ne 0, \tag{123}$$

for some $\epsilon_1 \to 0$ as $\epsilon \to 0$.

Therefore, we have for any $(x, \omega)$ that

$$\begin{aligned} &P_{(\mathbf{x},\mathbf{z})}(x, \omega) \\ &= \frac{1}{nB}N(x, \omega|\mathbf{x}, \mathbf{z}) \\ &= \frac{1}{nB}\sum_{(x^n,z^n)} N(x^n, z^n|\mathbf{x}, \mathbf{z}) \cdot N(x, \omega|x^n, z^n) \\ &= \sum_{(x^n,z^n)} P_{(\mathbf{x},\mathbf{z})}(x^n, z^n) \cdot P_{(x^n,z^n)}(x, \omega) \\ &= \sum_{x^n:P_{\mathbf{x}}(x^n)>0} P_{\mathbf{x}}(x^n)\sum_{z^n} P_{\mathbf{z}|\mathbf{x}}(z^n|x^n) \cdot P_{(x^n,z^n)}(x, \omega) \\ &\le \sum_{x^n:P_{\mathbf{x}}(x^n)>0} P_{\mathbf{x}}(x^n)\sum_{z^n} p(z^n|x^n)(1+\epsilon_1) \cdot P_{(x^n,z^n)}(x, \omega) \\ &= (1+\epsilon_1)\sum_{x^n:P_{\mathbf{x}}(x^n)>0} P_{\mathbf{x}}(x^n)E[P_{(x^n,Z^n)}(x, \omega)] \\ &= (1+\epsilon_1)\sum_{x^n:P_{\mathbf{x}}(x^n)>0} P_{\mathbf{x}}(x^n)E\left[\frac{1}{n}N(x, \omega|x^n, Z^n)\right] \\ &= (1+\epsilon_1)\sum_{x^n:P_{\mathbf{x}}(x^n)>0} P_{\mathbf{x}}(x^n)E\left[\frac{1}{n}\sum_{j:x_j=x}\mathbb{I}(Z_j=\omega)\right] \\ &= (1+\epsilon_1)\sum_{x^n:P_{\mathbf{x}}(x^n)>0} P_{\mathbf{x}}(x^n)P_{x^n}(x)p(\omega|x) \\ &= (1+\epsilon_1)P_{\mathbf{x}}(x)p(\omega|x) \\ &= (1+\epsilon_1)Q_n(x)p(\omega|x). \end{aligned}$$

Similarly,

$$P_{(\mathbf{x},\mathbf{z})}(x, \omega) \ge (1-\epsilon_1)Q_n(x)p(\omega|x), \quad \forall (x, \omega),$$

and thus

$$|P_{(\mathbf{x},\mathbf{z})}(x, \omega) - Q_n(x)p(\omega|x)| \le \epsilon_1 Q_n(x)p(\omega|x), \quad \forall (x, \omega).$$

## APPENDIX I

For any $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, consider the total variation distance between $P_{(\mathbf{x},\mathbf{y})}(x, \omega)$ and $P_{(\mathbf{x},\mathbf{z})}(x, \omega)$. We have

$$\begin{aligned} &nB\sum_{(x,\omega)}|P_{(\mathbf{x},\mathbf{y})}(x, \omega) - P_{(\mathbf{x},\mathbf{z})}(x, \omega)| \\ &= \sum_{(x,\omega)}\left|\sum_{i=1}^{nB}\mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega))\right| \\ &= \sum_{(x,\omega)}\left|\sum_{i:y_i=z_i}\mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega))\right. \\ &\qquad\left. + \sum_{i:y_i\ne z_i}\mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega))\right| \\ &= \sum_{(x,\omega)}\left|\sum_{i:y_i\ne z_i}\mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega))\right| \\ &\le \sum_{(x,\omega)}\sum_{i:y_i\ne z_i}\mathbb{I}((x_i, y_i) = (x, \omega)) \\ &\quad + \sum_{(x,\omega)}\sum_{i:y_i\ne z_i}\mathbb{I}((x_i, z_i) = (x, \omega)) \\ &= 2d_H(\mathbf{y}, \mathbf{z}), \end{aligned}$$

i.e.,

$$\sum_{(x,\omega)}|P_{(\mathbf{x},\mathbf{y})}(x, \omega) - P_{(\mathbf{x},\mathbf{z})}(x, \omega)| \le \frac{2}{nB}d_H(\mathbf{y}, \mathbf{z}).$$

## APPENDIX J

From property (30) of jointly typical sequences, for any $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and $\epsilon_1 > \epsilon$,

$$\Pr((\mathbf{X}, \mathbf{Y}, \mathbf{z}_0) \in \mathcal{T}_{\epsilon_1}^{(B)}(X^n, Y^n, Z^n)|\mathbf{z}_0) \to 1 \text{ as } B \to \infty. \tag{124}$$

For any $(\mathbf{y}, \mathbf{z}_0) \in \mathcal{T}_{\epsilon_1}^{(B)}(Y^n, Z^n)$,

$$p(\mathbf{y}|\mathbf{z}_0) \le 2^{-B(H(Y^n|Z^n)-\epsilon_2)}, \text{ for some } \epsilon_2 \to 0 \text{ as } \epsilon_1 \to 0. \tag{125}$$

Also, along the same lines as Appendix H, it can be shown that if $(\mathbf{x}, \mathbf{y}, \mathbf{z}_0)$ are jointly typical with respect to the $n$-letter random variables $(X^n, Y^n, Z^n)$, then $(\mathbf{x}, \mathbf{y}, \mathbf{z}_0)$ are also jointly typical with respect to the single-letter random variables $(X, Y, Z)$, i.e.,

$$\begin{aligned}|P_{(\mathbf{x},\mathbf{y},\mathbf{z}_0)}(x, y, z) - P_{\mathbf{x}}(x)p(y|x)p(z|x)| \\ \leq \epsilon_3 P_{\mathbf{x}}(x)p(y|x)p(z|x),\end{aligned}$$

for some $\epsilon_3 \to 0$ as $\epsilon_1 \to 0$. Then,

$$\begin{aligned}&P_{(\mathbf{y},\mathbf{z}_0)}(0, 1) \\ &\leq P_{\mathbf{x}}(0)(1-p)p(1+\epsilon_3) + P_{\mathbf{x}}(1)p(1-p)(1+\epsilon_3) \\ &= p(1-p)(1+\epsilon_3)\end{aligned}$$

and $P_{(\mathbf{y},\mathbf{z}_0)}(0, 1) \geq p(1-p)(1-\epsilon_3)$. Similarly, we also have

$$P_{(\mathbf{y},\mathbf{z}_0)}(1, 0) \in [p(1-p)(1-\epsilon_3), p(1-p)(1+\epsilon_3)],$$

and thus

$$\begin{aligned}d_H(\mathbf{y}, \mathbf{z}_0) &= nBP_{(\mathbf{y},\mathbf{z}_0)}(0, 1) + nBP_{(\mathbf{y},\mathbf{z}_0)}(1, 0) \\ &\in [2nBp(1-p)(1-\epsilon_3), 2nBp(1-p)(1+\epsilon_3)],\end{aligned}$$

i.e.,

$$d_H(\mathbf{y}, \mathbf{z}_0) \in [nB(p*p-\epsilon_4), nB(p*p+\epsilon_4)], \quad (126)$$

where $\epsilon_4 \to 0$ as $\epsilon_1 \to 0$.

Combining (124), (125) and (126), we conclude that for any $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and some $\epsilon_0 \to 0$ as $\epsilon \to 0$,

$$\begin{aligned}&\Pr\Big(p(\mathbf{Y}|\mathbf{z}_0) \leq 2^{-B(H(Y^n|Z^n)-\epsilon_0)}, \\ &\qquad d_H(\mathbf{Y}, \mathbf{z}_0) \in [nB(p*p-\epsilon_0), nB(p*p+\epsilon_0)]\Big|\mathbf{z}_0\Big) \\ &\to 1 \text{ as } B \to \infty.\end{aligned}$$

## Appendix K
## Property of $f(r)$

For notational convenience, let $q := p * p$. Taking the first derivative of $f(r)$, we have

$$\begin{aligned}f'(r) &= d_0 H_2'\left(\frac{r+d_0-q}{2d_0}\right)\cdot\left(\frac{r+d_0-q}{2d_0}\right)' \\ &\quad +(1-d_0)H_2'\left(\frac{r+q-d_0}{2(1-d_0)}\right)\cdot\left(\frac{r+q-d_0}{2(1-d_0)}\right)' \\ &= \frac{1}{2}\left[\log\frac{d_0-r+q}{r+d_0-q} + \log\frac{2-d_0-r-q}{r+q-d_0}\right].\end{aligned}$$

With $r = d_0 * q$, we have

$$\begin{aligned}&f(d_0 * q) \\ &= d_0 H_2\left(\frac{d_0*q+d_0-q}{2d_0}\right) \\ &\quad + (1-d_0)H_2\left(\frac{d_0*q+q-d_0}{2(1-d_0)}\right) \\ &= d_0 H_2\left(\frac{(d_0(1-q)+(1-d_0)q)+d_0-q}{2d_0}\right) \\ &\quad + (1-d_0)H_2\left(\frac{(d_0(1-q)+(1-d_0)q)+q-d_0}{2(1-d_0)}\right) \\ &= d_0 H_2(1-q) + (1-d_0)H_2(q) \\ &= H_2(q)\end{aligned}$$

and

$$\begin{aligned}&f'(d_0 * q) \\ &= \frac{1}{2}\left[\log\frac{d_0-d_0*q+q}{d_0*q+d_0-q} + \log\frac{2-d_0-d_0*q-q}{d_0*q+q-d_0}\right] \\ &= \frac{1}{2}\log\left[\frac{d_0-(d_0(1-q)+(1-d_0)q)+q}{(d_0(1-q)+(1-d_0)q)+d_0-q}\right. \\ &\qquad \left.\cdot\frac{2-d_0-(d_0(1-q)+(1-d_0)q)-q}{(d_0(1-q)+(1-d_0)q)+q-d_0}\right] \\ &= \frac{1}{2}\log\left[\frac{2d_0 q}{2d_0(1-q)}\cdot\frac{2(1-d_0)(1-q)}{2q(1-d_0)}\right] \\ &= 0.\end{aligned}$$

Further taking the second derivative of $f(r)$ yields that

$$\begin{aligned}f''(r) &= \frac{1}{2\ln 2}\left[\frac{r+d_0-q}{d_0-r+q}\left(\frac{d_0-r+q}{r+d_0-q}\right)'\right. \\ &\qquad \left.+\frac{r+q-d_0}{2-d_0-r-q}\left(\frac{2-d_0-r-q}{r+q-d_0}\right)'\right] \\ &= \frac{1}{2\ln 2}\left[\frac{2d_0}{(r+d_0-q)(r-d_0-q)}\right. \\ &\qquad \left.+\frac{2(d_0-1)}{(2-d_0-r-q)(r-d_0+q)}\right].\end{aligned}$$

Provided the following constraint on $r$ (cf. (102)–(103))

$$r \in (\max\{q-d_0, d_0-q\}, \min\{q+d_0, 2-q-d_0\})$$

it can be easily seen that $f''(r) < 0$. Therefore, $f(r)$ attains the maximum $H_2(p*p)$ if and only if $r = d_0 * q = d_0 * p * p$.

## References


[1] X. Wu, L.-L. Xie, and A. Özgür, "Upper bounds on the capacity of symmetric primitive relay channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 14–19.
[2] X. Wu and A. Ozgur, "Improving on the cut-set bound via geometric analysis of typical sets," in *Proc. Int. Zurich Seminar Commun.*, Mar. 2016, pp. 130–134.
[3] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, no. 1, pp. 120–154, 1971.
[4] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.
[5] L.-L. Xie and P. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.
[6] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
[7] X. Wu and L.-L. Xie, "A unified relay framework with both D-F and C-F relay nodes," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 586–604, Jan. 2014.
[8] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2000, pp. 1–22.
[9] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
[10] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.
[11] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
[12] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1509–1523, Apr. 2015.

[13] S. Zahedi, "On reliable communication over relay channels," Ph.D. dissertation, Dept. EE, Stanford Univ., Stanford, CA, USA, 2005.
[14] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, p. 536, May 1982.
[15] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.
[16] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1106–1110, Sep. 1988.
[17] K. Marton, "A simple proof of the blowing-up lemma (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 3, pp. 445–446, May 1986.
[18] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 921–930, Mar. 2009.
[19] F. Xue, "A new upper bound on the capacity of a primitive relay channel based on channel simulation," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4786–4798, Aug. 2014.
[20] T. M. Cover, "The capacity of the relay channel," in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, Eds. New York, NY, USA: Springer-Verlag, 1987, pp. 72–73.
[21] X. Wu, L. Barnes, and A. Ozgur, "Cover's open problem: 'The capacity of the relay channel,'" in *Proc. 54th Annu. Allerton Conf. Commun. Control, Comput.*, Monticello, IL, USA, Sep. 2016, pp. 1135–1142.
[22] T. S. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[23] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1393–1397.
[24] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 357–359, May 1973.
[25] R. G. Gallager. (1992). *Fixed Composition Arguments and Lower Bounds to Error Probability*. [Online]. Available: http://www.rle.mit.edu/rgallager/documents/notes5.pdf
[26] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[27] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. 45th Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Sep. 2007, pp. 130–135.
[28] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 25, no. 1, pp. 82–85, Jan. 1979.
[29] M. Raginsky and I. Sason. (Oct. 2013). *Concentration of Measure Inequalities in Information Theory, Communications and Coding*. [Online]. Available: http://arxiv.org/abs/1212.4663
[30] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
[31] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[32] X. Wu and A. Ozgur, "Improving on the cut-set bound for general primitive relay channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 1675–1679.
[33] X. Wu and A. Özgür, "Cut-set bound is loose for Gaussian relay networks," in *Proc. 53rd Annu. Allerton Conf. Commun. Control, Comput.*, vol. 1. Monticello, IL, USA, Sep. 2015, pp. 1135–1142.

**Xiugang Wu** (S'08–M'14) received the B.Eng. degree with honors in electronics and information engineering from Tongji University, Shanghai, China, in 2007, and the M.A.Sc and Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, Ontario, Canada, in 2009 and 2014, respectively. He is currently a postdoctoral research fellow in the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. His research interests are in information theory and wireless networks.

**Ayfer Özgür** (M'06) received her B.Sc. degrees in electrical engineering and physics from Middle East Technical University, Turkey, in 2001 and the M.Sc. degree in communications from the same university in 2004. From 2001 to 2004, she worked as hardware engineer for the Defense Industries Development Institute in Turkey. She received her Ph.D. degree in 2009 from the Information Processing Group at EPFL, Switzerland. In 2010 and 2011, she was a post-doctoral scholar with the Algorithmic Research in Network Information Group at EPFL. She is currently an Assistant Professor in the Electrical Engineering Department at Stanford University. Her research interests include network communications, wireless systems, and information and coding theory. Dr. Özgür received the EPFL Best Ph.D. Thesis Award in 2010 and a NSF CAREER award in 2013.

**Liang-Liang Xie** (M'03–SM'09–F'16) received the B.S. degree in mathematics from Shandong University, Jinan, China, in 1995 and the Ph.D. degree in control theory from the Chinese Academy of Sciences, Beijing, China, in 1999. He did postdoctoral research with the Automatic Control Group, Linköping University, Linköping, Sweden, during 1999-2000 and with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, during 2000-2002. He is currently a Professor at the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include network information theory, adaptive control and system identification.