# Improving on The Cut-Set Bound for General Primitive Relay Channels

Xiugang Wu
Stanford University
x23wu@stanford.edu

Ayfer Özgür
Stanford University
aozgur@stanford.edu

***Abstract*—Consider a 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 $R_0$. For the symmetric case, i.e., when $Y$ and $Z$ are conditionally i.i.d. given $X$, we have recently developed new upper bounds on the capacity of this channel that are tighter than existing bounds, including the celebrated cut-set bound. In this paper, we extend these bounds to the asymmetric case, where $Y$ and $Z$ are conditionally independent given $X$ with arbitrary conditional marginal distributions, for both discrete memoryless and Gaussian channels.**

## I. Introduction

Characterizing the capacity of relay channels [1] has been a long-standing open problem in information theory. The seminal work of Cover and El Gamal [2] introduced two basic achievability schemes: Decode-Forward and Compress-Forward, and derived a general upper bound on the capacity, now known as the cut-set bound. In most of the special cases where the capacity is known, the cut-set bound is tight. However, it is known to be not tight in general [3]–[5].

In our recent work [6]–[7], we developed upper bounds on the capacity of the primitive relay channel depicted in Fig. 1, which corresponds to a special case of the relay channel where the multiple access channel from the source and the relay to the destination has orthogonal components. We also assumed that the broadcast channel (BC) from the source to the relay and the destination is symmetric (i.e., $Y$ and $Z$ are conditionally i.i.d. given $X$ in Fig. 1). For this special case, when the BC is discrete and memoryless, our bounds [6]–[7] significantly improve over existing bounds for this channel [3], [5], and unlike [5] are always tighter than the cut-set bound.[1] In the case when the BC is Gaussian, our bound [8] provides the first example demonstrating that the cut-set bound is not tight for the Gaussian relay channel. Our approach builds on measure concentration to analyze the probabilistic geometric relations between the typical sets of the $n$-letter random variables associated with a reliable code. We translate these geometric relations to new information inequalities between the random variables involved.

This work was supported in part by the NSF CAREER award 1254786 and by the NSF award CCF-1514538.

[1]Note that [4] proves an exact capacity result but for a very specific instance 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.

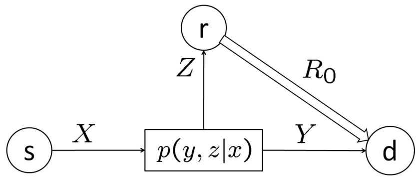


Fig. 1. Primitive relay channel.

Our arguments in [6]–[7] also critically build on the symmetry of the BC channel. In this paper, we extend our results [6]–[8] on symmetric discrete memoryless and symmetric Gaussian primitive relay channels to the general asymmetric case, and obtain new upper bounds that improve on the cut-set bound. In the discrete memoryless case, we use channel simulation ideas to connect the asymmetric setting to the symmetric case. In the Gaussian case, we extend our arguments in [8] to include asymmetric channels.

## II. Channel Model

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

$$(\Omega_X, p_{Y,Z|X}(y,z|x), \Omega_Y \times \Omega_Z)$$

where $\Omega_X, \Omega_Y$ and $\Omega_Z$ are finite sets denoting the alphabets of the source, the destination and the relay, respectively, and $p_{Y,Z|X}(y,z|x)$ is the channel transition probability; the relay $Z$ can communicate to the destination $Y$ via an error-free digital link of rate $R_0$. We assume that $Y$ and $Z$ are conditionally independent given $X$, i.e., $p_{Y,Z|X}(y,z|x) = p_{Y|X}(y|x)p_{Z|X}(z|x)$. The main interest in this paper is this general primitive relay channel. However, throughout the paper, we will also refer to the following special cases for this channel:

*1) Symmetric:* We say a primitive relay channel is symmetric, if $Y$ and $Z$ are further conditionally identically distributed given $X$, i.e., $\Omega_Y = \Omega_Z =: \Omega$ and $p_{Y|X}(\omega|x) = p_{Z|X}(\omega|x) =: p(\omega|x)$ for any $\omega \in \Omega$ and $x \in \Omega_X$.

*2) Stochastically degraded:* We say the channel is stochastically degraded if $Y$ is a stochastically degraded version of $Z$ (with respect to $X$), i.e., there exists some transition probability distribution $q_1(y|z)$ such that $p_{Y|X}(y|x) = \sum_z p_{Z|X}(z|x)q_1(y|z)$.

978-1-5090-1806-2/16/$31.00 ©2016 IEEE

*3) Reversely stochastically degraded:* We say the channel is reversely stochastically degraded if $Z$ is a stochastically degraded version of $Y$, i.e., there exists some transition probability distribution $q_2(z|y)$ such that $p_{Z|X}(z|x) = \sum_y p_{Y|X}(y|x)q_2(z|y)$.

We will be also interested in the Gaussian version of this channel, given by

$$\begin{cases} Z = X + W_1 \\ Y = X + W_2 \end{cases}$$

where $X \in \mathbb{R}$ denotes the source signal which is constrained to average power $P$, $Z \in \mathbb{R}$ and $Y \in \mathbb{R}$ denote the received signals of the relay and the destination, and $W_1$ and $W_2$ are Gaussian noises that are independent of each other and $X$, and have zero mean and variances $N_1$ and $N_2$ respectively. We say a Gaussian primitive relay channel is symmetric if $N_1 = N_2 =: N$.

## III. Main Result

The main result of this paper is to extend our previous results in [6] and [8] for the symmetric discrete memoryless and symmetric Gaussian primitive relay channels respectively, to the general case. In particular, in [6] we present two upper bounds on the capacity of the symmetric discrete memoryless primitive relay channel. The first of these bounds is given as follows:

*Proposition 3.1:* For the symmetric discrete memoryless primitive relay channel, if a rate $R$ is achievable, then there exists some $p_X(x)$ and $a \geq 0$ such that

$$\begin{cases} R \leq I(X;Y,Z) & (1) \\ R \leq I(X;Y) + R_0 - a & (2) \\ R \leq I(X;Y) + H\left(\sqrt{\frac{a \ln 2}{2}}\right) & \\ \qquad\qquad + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1) - a & (3) \end{cases}$$

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

In this paper, we extend this bound to general discrete memoryless primitive relay channels as follows:

*Theorem 3.1:* For the general primitive relay channel, if a rate $R$ is achievable, then there exists some $p_X(x)$ and $a \geq 0$ such that

$$\begin{cases} R \leq I(X;Y,Z) & (4) \\ R \leq I(X;Y) + R_0 - a & (5) \\ R \leq I(X;Y,\tilde{Z}) + H\left(\sqrt{\frac{a \ln 2}{2}}\right) & \\ \qquad\qquad + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega_Z| - 1) - a & (6) \end{cases}$$

for any random variable $\tilde{Z}$ with the same conditional distribution as $Z$ given $X$, i.e., $p_{\tilde{Z}|X}(z|x) = p_{Z|X}(z|x)$.

The evaluation of the bound in Theorem 3.1 involves optimizing over all the $\tilde{Z}$ random variables that have the same conditional distribution as $Z$. While this optimization may not be straightforward in general, note that any valid choice of $\tilde{Z}$ provides an upper bound on the capacity. In the following cases, it is trivial to find $\tilde{Z}$ that gives the tightest bound.

*1) Symmetric:* When the channel is symmetric, choosing $\tilde{Z} = Y$ makes the mutual information term $I(X;Y,\tilde{Z})$ in constraint (6) equal to $I(X;Y)$, which is the minimum possible value $I(X;Y,\tilde{Z})$ can achieve. In this case, Theorem 3.1 reduces to Proposition 3.1.

*2) Stochastically degraded:* It is easy to check that in the stochastically degraded case, the optimal $\tilde{Z}$ is such that $p_{Y,\tilde{Z}|X}(y,z|x) = p_{Z|X}(z|x)q_1(y|z)$, i.e., $X \leftrightarrow \tilde{Z} \leftrightarrow Y$ form a Markov chain. To see this, observe that $I(X;Y,\tilde{Z}) \geq I(X;\tilde{Z}) = I(X;Z)$, where the inequality holds with equality if and only if $X \leftrightarrow \tilde{Z} \leftrightarrow Y$ form a Markov chain.

*3) Reversely stochastically degraded:* In this case it is optimal to choose $\tilde{Z}$ such that $p_{Y,\tilde{Z}|X}(y,z|x) = p_{Y|X}(y|x)q_2(z|y)$, i.e., $X \leftrightarrow Y \leftrightarrow \tilde{Z}$ form a Markov chain. With this choice of $\tilde{Z}$, $I(X;Y,\tilde{Z})$ becomes $I(X;Y)$ and the constraints of Theorem 3.1 reduce to those as stated in Proposition 3.1 except that the $|\Omega|$ in (3) is replaced by $|\Omega_Z|$.

It is trivial to observe that the above bound is in general tighter than the cut-set bound given in the following proposition.

*Proposition 3.2 (Cut-set Bound):* For the general primitive relay channel, if a rate $R$ is achievable, then there exists some $p_X(x)$ such that

$$\begin{cases} R \leq I(X;Y,Z) & (7) \\ R \leq I(X;Y) + R_0. & (8) \end{cases}$$

Since $a \geq 0$ in Theorem 3.1, the constraint (5) is in general tighter than (8) and therefore our bound in Theorem 3.1 is in general tighter than the cut-set bound. In fact, the bound in Proposition 3.1, therefore Theorem 3.1 in the special case of symmetric channels, is *strictly* tighter than the cut-set bound. To see this, note that (2) will reduce to (8) only if $a = 0$; however, if $a = 0$ then (3) will constrain $R$ by the rate $I(X;Y)$ which is lower than the cut-set bound. Given this fact, it is easy to create examples of asymmetric primitive relay channels for which Theorem 3.1 is tighter than the cut-set bound. For example, as observed in 3) above, when the channel is reversely stochastically degraded, Theorem 3.1 has the same constraints as in Proposition 3.1 and is therefore strictly tighter than the cut-set bound. One can also create examples of stochastically degraded channels for which Theorem 3.1 is tighter than the cut-set bound by simply taking any symmetric channel and adding a sufficiently small degradation to the $X$-$Y$ link so that $I(X;Y) \approx I(X;Z)$.

In [6], we also provide a second bound on the capacity of the symmetric discrete memoryless primitive relay channel which does not include the bound in Theorem 3.1 in general and vice versa. The advantage of this second bound is that it can be significantly tighter than the one in Theorem 3.1 in some cases, as shown in [7] for the case of the binary symmetric channel, and can be extended to the continuous case, the Gaussian channel in particular as shown in [8]. Instead of recalling this second bound and providing its corresponding extension to the general case, due to space constraints, in this paper we only provide an extension of [8] to general Gaussian primitive relay channels. In particular, we have the following result which includes the result of [8] for symmetric Gaussian primitive relay channels as a special case.

*Theorem 3.2:* For the Gaussian primitive relay channel, if a rate $R$ is achievable, then there exists some $a \geq 0$ such that

$$\begin{cases} R \leq \frac{1}{2}\log\left(1+\frac{P}{N_1}+\frac{P}{N_2}\right) & (9) \\ R \leq \frac{1}{2}\log\left(1+\frac{P}{N_2}\right)+R_0-a & (10) \\ R \leq \frac{1}{2}\log\left(1+\frac{P}{N_1}\right)+\frac{N_1}{N_2}a+\left(\frac{N_1}{N_2}-1\right)^+\log e \\ \quad +\sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}-1\right)^+}(1+\sqrt{2a\ln 2})\log e \\ \quad +\sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a\ln 2+\left(1-\frac{N_1}{N_2}\right)^+\right)}\log e & (11) \end{cases}$$

where $(t)^+ = \max\{t, 0\}$.

## IV. Discrete Memoryless Primitive Relay Channels

In this section, we sketch the proof of Theorem 3.1 for discrete memoryless primitive relay channels. We first prove the theorem for the special cases of reversely stochastically degraded and stochastically degraded primitive relay channels where the arguments are relatively simple. We then prove the result for the general case. We build on a brief overview of the proof of Proposition 3.1 which we provide next.

### A. Proof of Proposition 3.1

Let the bin index forwarded by the relay, which is a function of $Z^n$, be denoted by $I_n \in [1 : 2^{nR_0}]$. The cut-set bound in (7)–(8) can be derived by first using Fano's inequality to conclude that $nR \leq I(X^n; Y^n, I_n) + n\epsilon$, and then single-letterizing this mutual information in two different ways. To obtain (7) we proceed as

$$I(X^n; Y^n, I_n) \leq I(X^n; Y^n, Z^n) \leq nI(X; Y, Z);$$

and to obtain (8) we can observe that

$$I(X^n; Y^n, I_n) = I(X^n; Y^n) + H(I_n|Y^n) - H(I_n|X^n), \quad (12)$$

and upper bound $I(X^n; Y^n)$ and $H(I_n|Y^n)$ with $nI(X; Y)$ and $nR_0$ respectively while lower bounding $H(I_n|X^n)$ by 0.

To obtain Proposition 3.1, instead of simply lower bounding $H(I_n|X^n)$ by 0 in the last step, we let $H(I_n|X^n) = na_n$, yielding $R \leq I(X; Y) + R_0 - a_n + \epsilon$, and prove a third constraint that forces $a_n$ to be strictly non-zero. This new constraint is obtained by bounding $H(I_n|Y^n)$ in (12) as

$$H(I_n|Y^n) \leq n\left[H\left(\sqrt{\frac{a_n \ln 2}{2}}\right)+\sqrt{\frac{a_n \ln 2}{2}}\log(|\Omega|-1)\right],$$

for which the proof can be found in [7] and omitted here. Interestingly, such a proof is based on considering the i.i.d. extensions of the random variables $(X^n, Y^n, Z^n, I_n)$ and studying the geometric relations between their typical sets using the generalized blowing-up lemma. It holds for any set of random variables $(X^n, Y^n, Z^n, I_n)$ satisfying the following properties:

(a) $X^n, Y^n, Z^n, I_n$ are discrete random variables, where $I_n$ is a function of $Z^n$ such that $H(I_n|X^n) = na_n$;
(b) $I_n \leftrightarrow Z^n \leftrightarrow X^n \leftrightarrow Y^n$ form a Markov chain;
(c) $\Omega_Y = \Omega_Z =: \Omega$ and

$$p_{Y^n|X^n}(\omega^n|x^n) = p_{Z^n|X^n}(\omega^n|x^n) = \prod_{i=1}^{n} p(\omega_i|x_i).$$

### B. Reversely Stochastically Degraded Case

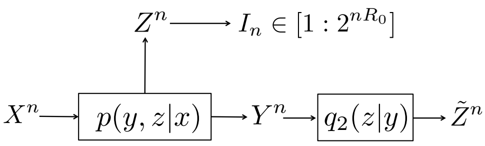


Fig. 2. Random variables in the reversely stochastically degraded case.

We now prove Theorem 3.1 for the case when $Z$ is a stochastically degraded version of $Y$, i.e., there exists some transition probability distribution $q_2(z|y)$ such that $p_{Z|X}(z|x) = \sum_y p_{Y|X}(y|x)q_2(z|y)$. For this, we introduce an auxiliary random variable $\tilde{Z}^n$, which is obtained by passing $Y^n$ through the channel $q_2(z|y)$. See Fig. 2. Since $Z^n$ and $\tilde{Z}^n$ are conditionally i.i.d. given $X^n$, conditions (a)–(c) at the end of the previous subsection are satisfied for random variables $(X^n, \tilde{Z}^n, Z^n, I_n)$ and we have

$$H(I_n|\tilde{Z}^n) \leq n\left[H\left(\sqrt{\frac{a_n \ln 2}{2}}\right)+\sqrt{\frac{a_n \ln 2}{2}}\log(|\Omega_Z|-1)\right]. \quad (13)$$

Noting the Markov chain $I_n \leftrightarrow Y^n \leftrightarrow \tilde{Z}^n$, we further have

$$H(I_n|Y^n) \leq H(I_n|\tilde{Z}^n) \leq \text{R.H.S. of (13)},$$

which combined with (12) yields the following constraint

$$R \leq I(X;Y)+H\left(\sqrt{\frac{a_n \ln 2}{2}}\right)+\sqrt{\frac{a_n \ln 2}{2}}\log(|\Omega_Z|-1)-a_n+\epsilon$$

This proves Theorem 3.1 for the reversely stochastically degraded case.

### C. Stochastically Degraded Case

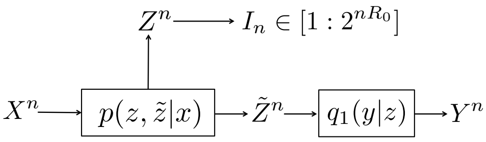


Fig. 3. Random variables in the stochastically degraded case.

When $Y$ is a stochastically degraded version of $Z$, i.e., there exists some transition probability distribution $q_1(y|z)$ such that $p_{Y|X}(y|x) = \sum_z p_{Z|X}(z|x)q_1(y|z)$, we can equivalently think of the random variables $(X^n, Y^n, Z^n, I_n)$ as generated according to the structure depicted in Fig. 3, where we introduce an intermediate random variable $\tilde{Z}^n$ so that $\tilde{Z}^n$ and $Z^n$ are conditionally i.i.d. given $X^n$, and $Y^n$ is obtained by passing $\tilde{Z}^n$ through the channel $q_1(y|z)$. Again, conditions (a)–(c) in Section IV-A are satisfied for $(X^n, \tilde{Z}^n, Z^n, I_n)$ and inequality (13) also holds here. Noting the Markov chain $X^n \leftrightarrow (\tilde{Z}^n, I_n) \leftrightarrow (Y^n, I_n)$, we have the following:

$$\begin{aligned} nR &\le I(X^n; Y^n, I_n) + n\epsilon \\ &\le I(X^n; \tilde{Z}^n, I_n) + n\epsilon \\ &= I(X^n; \tilde{Z}^n) + H(I_n|\tilde{Z}^n) - H(I_n|X^n) + n\epsilon \\ &\le n(I(X;Z) + H\left(\sqrt{\frac{a_n \ln 2}{2}}\right) \\ &\qquad + \sqrt{\frac{a_n \ln 2}{2}} \log(|\Omega_Z| - 1) - a_n + \epsilon), \quad (14) \end{aligned}$$

where in the last step we have used inequality (13) and the fact that $\tilde{Z}^n$ and $Z^n$ are conditionally i.i.d. given $X^n$. This finishes the proof of Theorem 3.1 for the stochastically degraded case.

### D. General Case

In the general case, the construction of the auxiliary random variable $\tilde{Z}^n$ (so that $\tilde{Z}^n$ and $Z^n$ are conditionally i.i.d. given $X^n$) may not be readily doable as in the previous two subsections, and we will resort to channel simulation theory [9]–[10]. Specifically, consider the channel simulation setup as depicted in Fig. 4, where we want to simulate some channel $p_{\tilde{Z}|X}(z|x)$ such that $p_{\tilde{Z}|X}(z|x) = p_{Z|X}(z|x)$, i.e., $\tilde{Z}$ has the same conditional distribution as $Z$. The simulation encoder sees the source $X^n$, side information $Y^n$, and a common random variable $K_n$ which is uniformly distributed on $[1 : 2^{nR_2}]$ and independent of those random variables $(X^n, Y^n, Z^n, I_n)$ associated with the original channel, and it generates a simulation codeword $J_n \in [1 : 2^{nR_1}]$ based on a randomized encoding function $E_n(X^n, Y^n, K_n)$. The simulation decoder also observes $Y^n$ and $K_n$, and upon receiving $J_n$ it outputs a random variable $\tilde{Z}^n$ based on a randomized decoding function $D_n(J_n, Y^n, K_n)$.

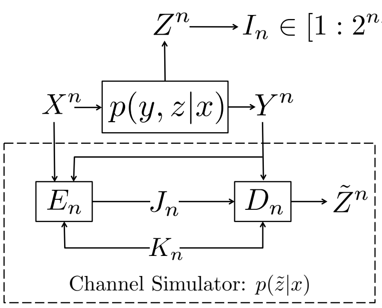


Fig. 4. Channel simulation for the general case.

Following the similar lines as in [9]–[10] and [5], it can be shown that the channel $p_{\tilde{Z}|X}(\tilde{z}|x)$ can be simulated in the above setup, i.e., $\tilde{Z}^n$ is (essentially) the same as if it is generated by passing $X^n$ through the channel $p_{\tilde{Z}|X}(\tilde{z}|x)$, if $R_1 = I(X; \tilde{Z}|Y) + \epsilon$ and $R_2$ is sufficiently large. In this case, $\tilde{Z}^n$ and $Z^n$ are conditionally identically distributed given $X^n$, and due to the Markov chain $Z^n \leftrightarrow X^n \leftrightarrow (Y^n, J_n, K_n) \leftrightarrow \tilde{Z}^n$ they are also conditionally independent given $X^n$. Therefore inequality (13) also holds here. Now consider expanding $H(I_n, J_n, K_n|Y^n)$ in two different ways as follows:

$$\begin{aligned} &H(I_n, J_n, K_n|Y^n) \\ &= H(I_n|Y^n) + H(K_n|Y^n, I_n) + H(J_n|Y^n, I_n, K_n) \\ &= H(K_n|Y^n) + H(J_n|Y^n, K_n) + H(I_n|Y^n, K_n, J_n). \end{aligned}$$

Therefore,

$$\begin{aligned} &H(I_n|Y^n) \\ &= H(K_n|Y^n) + H(J_n|Y^n, K_n) + H(I_n|Y^n, K_n, J_n) \\ &\qquad - H(K_n|Y^n, I_n) - H(J_n|Y^n, I_n, K_n) \\ &\le H(K_n|Y^n) + H(J_n|Y^n, K_n) + H(I_n|Y^n, K_n, J_n) \\ &\qquad - H(K_n|Y^n, I_n) \\ &= H(J_n|Y^n, K_n) + H(I_n|Y^n, K_n, J_n) \quad (15) \\ &= H(J_n|Y^n, K_n) + H(I_n|Y^n, K_n, J_n, \tilde{Z}^n) \quad (16) \\ &\le n(I(X; \tilde{Z}|Y) + \epsilon) + H(I_n|\tilde{Z}^n) \quad (17) \\ &\le n(I(X; \tilde{Z}|Y) + \epsilon) + \text{R.H.S. of (13)} \end{aligned}$$

where (15) follows because $K_n$ is independent of $(Y^n, I_n)$, (16) follows from the Markov chain $I_n \leftrightarrow (Y^n, K_n, J_n) \leftrightarrow \tilde{Z}^n$, and (17) follows because $H(J_n|Y^n, K_n) \le nR_1 = n(I(X; \tilde{Z}|Y) + \epsilon)$ and removing condition does not reduce entropy. Finally, it is straightforward to check that the above upper bound on $H(I_n|Y^n)$ combined with (12) yields constraint (6), completing the proof of Theorem 3.1.

## V. Gaussian Primitive Relay Channels

We now sketch the proof of Theorem 3.2 for Gaussian primitive relay channels. Similarly, we first briefly overview the proof for the symmetric case, and then prove the theorem for the general case.

### A. Symmetric Case

The proofs of (9)–(10) follow the same lines as the proofs of (4)–(5), i.e., by applying Fano's inequality and letting $H(I_n|X^n) = na_n$. In the symmetric case, (11) simplifies to:

$$R \le \frac{1}{2}\log\left(1 + \frac{P}{N}\right) + a + \sqrt{2a \ln 2}\log e, \quad (18)$$

which can be proved by expanding $I(X^n; Y^n, I_n)$ as

$$I(X^n; Y^n, I_n) = I(X^n; I_n) + h(Y^n|I_n) - h(Y^n|X^n) \quad (19)$$

and upper bounding $h(Y^n|I_n)$ by

$$h(Y^n|I_n) \le H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi e N + n(a_n + \sqrt{2a_n \ln 2}\log e). \quad (20)$$

To derive (20), we consider the $B$-letter i.i.d. extensions of the random variables $(X^n, Y^n, Z^n, I_n)$, denoted by $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{I})$, and study the geometric relations between their typical sets via Gaussian measure concentration. The outcome of our geometric analysis is an interesting result that can be roughly stated as follows. Suppose:

(a) $Z^n = X^n + W_1^n$ and $Y^n = X^n + W_2^n$, where both $W_1^n$ and $W_2^n$ are i.i.d. sequences of Gaussian random variables with zero mean and variance $N$ and they are independent of each other and $X^n$;
(b) $I_n$ is a function of $Z^n$ which takes value on a finite set and satisfies $H(I_n|X^n) = na_n$.

Then we have for a typical $(\mathbf{y}, \mathbf{i})$ pair, there exists some $\mathbf{z}$ belonging to the $\mathbf{i}$th bin such that

$$d(\mathbf{y}, \mathbf{z}) \le \sqrt{nB}\sqrt{2Na_n \ln 2}, \quad (21)$$

with $d(\mathbf{y}, \mathbf{z}) := ||\mathbf{y} - \mathbf{z}||$ denoting the Euclidean distance. Based on the geometric relation (21), one can then derive (20). We will not expose the argument here but instead we will extend it to the general case in the next subsection and explain it there in detail.

### B. General Case

To prove Theorem 3.2, we focus on the case when $N_1 \le N_2$ to illustrate the main argument while only pointing out the difference for the $N_1 > N_2$ case. When $N_1 \le N_2$, one can equivalently think of $Z$ and $Y$ as given by

$$\begin{cases} Z = X + W_1 \\ Y = X + W_{21} + W_{22} \end{cases}$$

where $W_1, W_{21}, W_{22}$ are zero-mean Gaussian random variables with variances $N_1, N_1, N_2 - N_1$ respectively, and they are independent of each other and $X$. Based on this, we write

$$\begin{cases} Z^n = X^n + W_1^n \\ Y^n = \tilde{Z}^n + W_{22}^n \end{cases}$$

with $\tilde{Z}^n := X^n + W_{21}^n$. Note that now $(X^n, \tilde{Z}^n, Z^n, I_n)$ satisfy conditions (a)–(b) in the previous subsection, and thus for a typical $(\tilde{\mathbf{z}}, \mathbf{i})$ pair, there exists some $\mathbf{z}$ belonging to the $\mathbf{i}$th bin such that $d(\tilde{\mathbf{z}}, \mathbf{z}) \le \sqrt{nB}\sqrt{2N_1 a_n \ln 2}$. Moreover, since $Y^n = \tilde{Z}^n + W_{22}^n$ with $\tilde{Z}^n$ and $W_{22}^n$ being independent, it can be shown that for a fixed pair of $(\tilde{\mathbf{z}}, \mathbf{z})$, the following pythagorean relation holds with high probability:

$$d^2(\mathbf{Y}, \mathbf{z}) \approx d^2(\mathbf{Y}, \tilde{\mathbf{z}}) + d^2(\tilde{\mathbf{z}}, \mathbf{z}) \quad (22)$$

$$\approx nB(N_2 - N_1) + d^2(\tilde{\mathbf{z}}, \mathbf{z}), \quad (23)$$

and thus for a typical $(\mathbf{y}, \mathbf{i})$ pair, there exists some $\mathbf{z}$ belonging to the $\mathbf{i}$th bin such that

$$d(\mathbf{y}, \mathbf{z}) \le \sqrt{nB}\sqrt{N_2 + N_1(2a_n \ln 2 - 1)}. \quad (24)$$

We now lower bound the conditional density $f(\mathbf{y}|\mathbf{i})$ for a typical $(\mathbf{y}, \mathbf{i})$ pair based on the geometric relation (24). In particular, consider the set of $\mathbf{x}$'s that are jointly typical with the $\mathbf{z}$ satisfying (24). It can be shown that the $\mathbf{x}$'s that are jointly typical with this $\mathbf{z}$ satisfy $p(\mathbf{x}|\mathbf{i}) \doteq 2^{-BH(X^n|I_n)}$ and $d(\mathbf{x}, \mathbf{z}) \le \sqrt{nBN_1}$. Therefore, by the triangle inequality for each $\mathbf{x}$ in this set

$$d(\mathbf{x}, \mathbf{y}) \le \sqrt{nB}(\sqrt{N_1} + \sqrt{N_2 + N_1(2a_n \ln 2 - 1)}),$$

which leads to the following lower bound on $f(\mathbf{y}|\mathbf{x})$,

$$f(\mathbf{y}|\mathbf{x}) \mathrel{\dot{\ge}} 2^{-nB\left(\frac{1}{2}\log 2\pi e N_2 + \frac{N_1}{N_2}a + \sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a_n \ln 2 + 1 - \frac{N_1}{N_2}\right)}\log e\right)} \quad (25)$$

by using the fact that $\mathbf{Y}$ is Gaussian given $\mathbf{X}$. The set of such $\mathbf{x}$'s can be shown to have cardinality approximately given by $2^{BH(X^n|Z^n)}$. Combining this with the above, we have

$$f(\mathbf{y}|\mathbf{i}) = \sum_{\mathbf{x}} f(\mathbf{y}|\mathbf{x}) p(\mathbf{x}|\mathbf{i})$$

$$\mathrel{\dot{\ge}} 2^{B(H(X^n|Z^n) - H(X^n|I_n))} \times \text{R.H.S. of (25)}.$$

Finally, translating the above lower bound on $f(\mathbf{y}|\mathbf{i})$ for typical $(\mathbf{y}, \mathbf{i})$ pairs to the upper bound on $h(Y^n|I_n)$, we have

$$h(Y^n|I_n) \le H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi e N_2 + n\left(\frac{N_1}{N_2}a + \sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a_n \ln 2 + 1 - \frac{N_1}{N_2}\right)}\log e\right)$$

Plugging the above inequality into (19), we obtain constraint (11) with $N_1 \le N_2$, which proves the theorem for this case.

To prove the theorem for the case of $N_1 > N_2$, construct an auxiliary random variable $\tilde{Z}^n$ as $\tilde{Z}^n := Y^n + \tilde{W}^n$, where $\tilde{W}^n$ is an i.i.d. sequence of Gaussian random variables with zero mean and variance $N_1 - N_2$. Again $(X^n, \tilde{Z}^n, Z^n, I_n)$ satisfy conditions (a)–(b) in the previous subsection, and one can then proceed as in the $N_1 \le N_2$ case.

## REFERENCES


[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[2] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, 1979.

[3] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1106–1110, Sept. 1988.

[4] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 3, pp. 921–930, 2009.

[5] F. Xue, "A new upper bound on the capacity of a primitive relay channel based on channel simulation," *IEEE Trans. Inform. Theory*, vol. 60, pp. 4786–4798, Aug. 2014.

[6] X. Wu and A. Ozgur, "Improving on the cut-set bound via geometric analysis of typical sets," *International Zurich Seminar on Communications*, 2016.

[7] X. Wu, A. Ozgur, L.-L. Xie, "Improving on the cut-set bound via geometric analysis of typical sets," submitted to *IEEE Trans. Inform. Theory*. Available: http://arxiv.org/abs/1602.08540

[8] X. Wu and A. Ozgur, "Cut-set bound is loose for Gaussian relay networks," *53rd Annual Allerton Conf. Commun., Contr., Comput.*, 2015.

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

[10] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symposium on Information Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1393–1397.