Cut-Set Bound Is Loose for Gaussian Relay Networks

Xiugang Wu, *Member, IEEE*, and Ayfer Özgür, *Member, IEEE*

***Abstract*— The cut-set bound developed by Cover and El Gamal in 1979 has since remained the best known upper bound on the capacity of the Gaussian relay channel. We develop a new upper bound on the capacity of the Gaussian primitive relay channel, which is tighter than the cut-set bound. Our proof uses Gaussian measure concentration to establish geometric relations, satisfied with high probability, between the *n*-letter random variables associated with a reliable code for communicating over this channel. We then translate these geometric relations into new information inequalities that cannot be obtained with classical methods. Combined with a tensorization argument proposed by Courtade and Ozgur in 2015, our result also implies that the current capacity approximations for Gaussian relay networks, which have linear gap to the cut-set bound in the number of nodes, are order-optimal and lead to a lower bound on the pre-constant.**

***Index Terms*— Gaussian relay channel, cut-set bound, converse, capacity approximation, information inequality, geometry.**

I. Introduction

The single-relay channel is one of the simplest examples of a network information theory problem, which defies our complete understanding despite decades of research. The Gaussian version of this problem models the communication scenario where a wireless link is assisted by a single relay. Motivated by the need to increase the spectral efficiency of wireless systems and the increasing importance of relaying for small cells, it has been studied extensively since its formulation by van der Meulen in 1971 [1]. However, the characterization of its capacity still remains an open problem. Perhaps more interestingly, the existing literature almost exclusively focuses on developing achievable strategies for this channel as well as larger relay networks. This has led to a plethora of relaying schemes over the last decade, such as decode-and-forward, compress-and-forward, amplify-and-forward, compute-and-forward, quantize-map-and-forward, noisy network coding, etc [2]–[8]. In sharp contrast, the only available upper bound on the capacity of the Gaussian relay channel is the so called cut-set bound developed by Cover and El Gamal in 1979 [2]. In the 40-year long literature on the problem, the cut-set bound has been consistently used as a benchmark for performance – for example the recent approximation approach [6], [8], [11] in wireless information theory focuses on bounding the gap of the achievable strategies to the cut-set bound of the network– however to our knowledge, whether the cut-set bound is indeed achievable or not in a Gaussian relay channel (except in trivial cases) remains unknown to date.

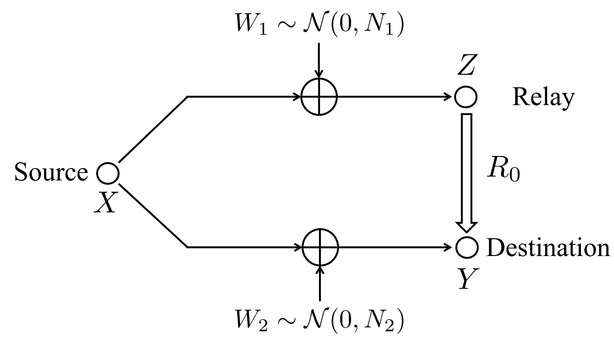

Fig. 1. Gaussian primitive relay channel.

In this paper, we make progress on this problem by developing a new upper bound on the capacity of the Gaussian primitive relay channel. This is a special case of the Gaussian single relay channel where the multiple-access channel from the source and the relay to the destination has orthogonal components [9], [10]. See Fig. 1. Here, the relay can be thought of as communicating to the destination over a Gaussian channel in a separate frequency band. Our upper bound is tighter than the cut-set bound for this channel for all (non-trivial) channel parameters. While this result is developed for the single-relay setting, it has implications also for networks with multiple relays. In particular, combined with a recently proposed tensorization argument [13], it implies that the linear (in the number of nodes) gap to the cut-set bound in current capacity approximations for Gaussian relay networks is indeed fundamental. The capacity of Gaussian relay networks can have linear gap to the cut-set bound and our result can be used to obtain a lower bound on the pre-constant.

Proving that the cut-set bound is not tight requires to capture the following phenomenon: if a relay is not able to decode the transmitted message and therefore remove the noise in its received signal by decoding, then the signal it forwards

Manuscript received June 4, 2016; revised March 5, 2017; accepted July 17, 2017. Date of publication July 26, 2017; date of current version January 18, 2018. This work was supported in part by NSF under Award CIF-1704624 and 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 Allerton Conference on Communication, Control, and Computing [36] and the 2016 IEEE International Symposium on Information Theory [37].

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: x23wu@stanford.edu; aozgur@stanford.edu).

Communicated by A. Khisti, 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.2017.2731963

necessarily contains noise along with information. The injected noise then decreases the end-to-end achievable rate with respect to the cut-set bound, where the latter simply upper bounds the end-to-end capacity by the maximal information flow over cuts of the network assuming all nodes on the source side of the cut have noiseless access to the message and all nodes on the destination side can freely cooperate to decode the transmitted message. As basic as it sounds, existing approaches for developing infeasibility results in information theory seem insufficient to quantitatively capture this phenomenon.[1]

In this and our concurrent work [15]–[17] on the discrete memoryless version of this problem, we build a novel geometric approach to capture this phenomenon. The main idea is to study the geometric relations that are satisfied by typical realizations of the n-letter random variables associated with a reliable code for communicating over the relay channel. (Equivalently, these are the geometric relations that are satisfied with high probability by these n-letter random variables.) We then translate these geometric relations into new and surprising relations between the entropies of the corresponding random variables. A key ingredient in this approach is a measure concentration result, namely the blowing-up lemma due to Marton [29], which says that under a product measure slightly blowing up any set with a small but exponentially significant probability suffices to increase its probability to nearly 1.[2] This lemma allows us to obtain distance relations between typical sets, which we then translate to entropy relations.

While our bounds for the discrete memoryless relay channel in [15]–[17] and the Gaussian case treated in the current paper have similar flavor, these two cases also comprise some significant differences. In particular, the discrete memoryless case seems easier to deal with as one can make explicit counting arguments and rely on the standard notion of strong typicality. Indeed, the Gaussian case has proven to be associated with some inherent difficulty historically—for example, the earlier results by Zhang [18], and Aleksic *et al.* [20] that demonstrate the looseness of the cut-set bound in the discrete memoryless case do not have counterparts in the Gaussian case. Also, the recent upper bound developed by Xue [19] for the discrete memoryless relay channel cannot be extended to the Gaussian case, as it relies on a counting argument that is valid only when the output alphabet is finite.[3] To develop an upper bound on the capacity of the primitive relay channel in the Gaussian case, this paper develops a new argument for translating geometric relations between typical sets of random variables into relations between their entropies. We also construct a series of typical sets for a mixed set of discrete and continuous random variables that enjoy some properties of strong typical sets.

A. Organization of the Paper

The remainder of the paper is organized as follows. First Section II introduces the channel model and reviews the classical cut-set bound on the capacity of the Gaussian primitive relay channel. Then Section III presents our new upper bound and discusses its implication on the capacity approximation problem for Gaussian relay networks, followed by the proof of our bound in Sections IV and V. Finally in Section VI, we provide another bound which sharpens our main result for certain regimes of the channel parameters. We include this result as to illustrate that there may be significant potential for improving our results by refining our method and arguments.

II. Preliminaries

A. Channel Model

Consider a Gaussian primitive relay channel as depicted in Fig. 1, where $X \in \mathbb{R}$ denotes the source signal which is constrained to average power P, and $Z \in \mathbb{R}$ and $Y \in \mathbb{R}$ denote the received signals of the relay and the destination. We have

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

where 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. The relay can communicate to the destination via an error-free digital link of rate R_0.

For this channel, a code of rate R and blocklength n, 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,
$$\mathcal{C}_{(n,R)} = \{x^n\ (m), m \in \{1, 2, \dots, 2^{nR}\}\}$$
where
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(m) \le P,\ \forall m \in \{1, 2, \dots, 2^{nR}\};$$
2) An encoding function at the relay Z,
$$f_n : \mathbb{R}^n \to \{1, 2, \dots, 2^{nR_0}\};$$
3) A decoding function at the destination Y,
$$g_n : \mathbb{R}^n \times \{1, 2, \dots, 2^{nR_0}\} \to \{1, 2, \dots, 2^{nR}\}.$$

The average probability of error of the code is defined as

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

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

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

[1]A similar observation was pointed out in [14].

[2]For a more detailed discussions regarding concentration of measure, and the blowing-up lemma along with its earlier applications in network information theory, see the comprehensive monograph by Raginsky and Sason [34]. In this context, it is also worth mentioning that tools related to Gaussian concentration and Marton's transportation-cost inequalities have also been invoked in a recent work by Polyanskiy and Wu [33] to solve the "missing corner point" problem for the two-user Gaussian interference channel.

[3]Note that this issue cannot be resolved by the standard discretization procedure that is typically used for extending an achievability theorem for a discrete memoryless channel to a continuous channel, because as the quantization interval goes to zero the upper bound in [19] obtained by a counting argument becomes arbitrarily large.

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 $C(R_0)$.

B. The Cut-Set Bound

For the Gaussian primitive relay channel, the cut-set bound can be stated as follows.

Proposition 1 (Cut-Set Bound): For the Gaussian primitive relay channel, if a rate R is achievable, then there exists a random variable X satisfying $E[X^2] \le P$ such that

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

Note that constraints (1) and (2) correspond to the broadcast channel X-YZ and multiple-access channel XZ-Y, and hence are generally known as the broadcast and multiple-access constraints, respectively. Also it can be easily shown that both $I(X; Y, Z)$ and $I(X; Y)$ in Proposition 1 are maximized when $X \sim \mathcal{N}(0, P)$, which leads us to the following corollary.

Corollary 2: For the Gaussian primitive relay channel, if a rate R is achievable, then

$$\begin{cases} R \le \frac{1}{2}\log\left(1 + \frac{P}{N_1} + \frac{P}{N_2}\right) & (3) \\ R \le \frac{1}{2}\log\left(1 + \frac{P}{N_2}\right) + R_0. & (4) \end{cases}$$

III. Main Result

Our main result in this paper is the following theorem, which provides a new upper bound on the capacity of the Gaussian primitive relay channel that is tighter than the cut-set bound. The proof of this theorem is given in Section IV.

Theorem 3: For the Gaussian primitive relay channel, if a rate R is achievable, then there exists a random variable X satisfying $E[X^2] \le P$ and some $a \in [0, R_0]$ such that

$$\begin{cases} R \le I(X; Y, Z) & (5) \\ R \le I(X; Y) + R_0 - a & (6) \\ R \le \max\{I(X; Y), \quad I(X; Z)\} + a + \sqrt{2a \ln 2}\log e. & (7) \end{cases}$$

As in the cut-set bound, all the mutual information terms $I(X; Y, Z)$, $I(X; Y)$ and $I(X; Z)$ in the above theorem are maximized when $X \sim \mathcal{N}(0, P)$, and therefore our bound can be re-stated more explicitly in terms of the logarithmic function as follows.

Corollary 4: For the Gaussian primitive relay channel, if a rate R is achievable, then there exists some $a \in [0, R_0]$ such that

$$\begin{cases} R \le \frac{1}{2}\log\left(1 + \frac{P}{N_1} + \frac{P}{N_2}\right) & (8) \\ R \le \frac{1}{2}\log\left(1 + \frac{P}{N_2}\right) + R_0 - a & (9) \\ R \le \frac{1}{2}\log\left(1 + \max\left\{\frac{P}{N_1}, \frac{P}{N_2}\right\}\right) \\ \qquad + a + \sqrt{2a \ln 2}\log e. & (10) \end{cases}$$

Since $a \ge 0$ in the above, our bound is in general tighter than the cut-set bound in Corollary 2. In fact, our bound can be *strictly* tighter than the cut-set bound when the multiple-access constraint (4) is active in the cut-set bound. To see this, first consider the symmetric case when $N_1 = N_2 =: N$. For this case, the cut-set bound in Corollary 2 says that if a rate R is achievable, then

$$\begin{cases} R \le \frac{1}{2}\log\left(1 + \frac{2P}{N}\right) & (11) \\ R \le \frac{1}{2}\log\left(1 + \frac{P}{N}\right) + R_0, & (12) \end{cases}$$

while our bound in Corollary 4 asserts that any achievable rate R must satisfy

$$\begin{cases} R \le \frac{1}{2}\log\left(1 + \frac{2P}{N}\right) & (13) \\ R \le \frac{1}{2}\log\left(1 + \frac{P}{N}\right) + R_0 - a^*, & (14) \end{cases}$$

where a^* is the solution to the following equation:

$$R_0 = 2a^* + \sqrt{2a^* \ln 2}\log e, \qquad (15)$$

which is obtained by equating the R.H.S. of constraints (9) and (10). Obviously, if $R_0 > 0$, then $a^* > 0$ and (14) is tighter than (12). Therefore, when constraint (12) is more stringent between (11) and (12), our bound is strictly tighter than the cut-set bound. The same argument and conclusion also apply when $N_1 \ge N_2$, in which case our bound reduces to

$$\begin{cases} R \le \frac{1}{2}\log\left(1 + \frac{P}{N_1} + \frac{P}{N_2}\right) & (16) \\ R \le \frac{1}{2}\log\left(1 + \frac{P}{N_2}\right) + R_0 - a^*, & (17) \end{cases}$$

where a^* is similarly defined as in (15). Finally it can be easily checked that when $N_1 \le N_2$, our bound is also strictly tighter than the cut-set bound as long as

$$\frac{1}{2}\log\left(1 + \frac{P}{N_1}\right) \le \frac{1}{2}\log\left(1 + \frac{P}{N_2}\right) + R_0.$$

Note that both the cut-set bound and our bound depend on the channel parameters through $\frac{P}{N_1}$, $\frac{P}{N_2}$ and R_0. It is interesting to evaluate the largest gap between these two bounds over all parameter values $(\frac{P}{N_1}, \frac{P}{N_2}, R_0)$. For this we show in Appendix A the following proposition, which says that the largest gap occurs in the symmetric case when $\frac{P}{N_1} = \frac{P}{N_2} \to \infty$ and $R_0 = 0.5$.

Proposition 5: Let $\Delta\left(\frac{P}{N_1}, \frac{P}{N_2}, R_0\right)$ denote the gap between our bound and the cut-set bound, and Δ^* be its largest possible value over all Gaussian primitive relay channels, i.e.,

$$\Delta^* := \sup_{\frac{P}{N_1}, \frac{P}{N_2}, R_0} \Delta\left(\frac{P}{N_1}, \frac{P}{N_2}, R_0\right).$$

Then, $\Delta^* = \Delta(\infty, \infty, 0.5) = 0.0535$.

A. Gaussian Relay Networks

The primitive single-relay channel we consider in this paper can be regarded as a special case of a Gaussian relay network. However, the upper bound we develop for this special case has also implications for larger Gaussian relay networks with multiple relays. In particular, it can be used to infer how tightly the capacity of general Gaussian relay networks can be approximated by the cut-set bound. Consider a discrete memoryless Gaussian relay network of N nodes, in which a source node s aims to reliably communicate a message to a destination node d.[4] For each node $i \in \{1, 2, \ldots, N\} =: \mathcal{N}$, we let R_i and T_i denote the numbers of receive-antennas and transmit-antennas of node i, respectively. We adopt the usual Gaussian relay network setting, where if $\mathbf{x}_j[t] \in \mathbb{R}^{T_j}$ is the signal transmitted by node j at time instant t, the signal received at node i is given by

$$\mathbf{y}_i[t] = \sum_{j \in \mathcal{N}} G_{ij} \mathbf{x}_j[t] + \mathbf{z}_i[t], \tag{18}$$

where $G_{ij} \in \mathbb{R}^{R_i \times T_j}$ is a known $R_i \times T_j$ matrix describing the channel gain from node j to i, $\mathbf{z}_i[t] \sim \mathcal{N}(0, I_{R_i \times R_i})$ is additive Gaussian noise with $\{\mathbf{z}_1[t], \mathbf{z}_2[t], \ldots, \mathbf{z}_N[t]\}_{t=1,2,\ldots}$ being mutually independent. In this manner, a Gaussian relay network is completely characterized by the triple (G, s, d), where G denotes the collection of channel gain matrices $\{G_{ij} : i, j \in \mathcal{N}\}$. For a network (G, s, d), it will be convenient to define the quantity

$$\kappa(G, s, d) := \sum_{i \in \mathcal{N}} \max\{T_i, R_i\} \tag{19}$$

since it will be referred to frequently. When the network (G, s, d) under consideration is clear from context, we will abbreviate $\kappa \equiv \kappa(G, s, d)$. A code and an achievable rate for a Gaussian relay network (G, s, d) and the *capacity* $C(G, s, d)$ are defined in the standard way (see for example [13]).

For a network (G, s, d), the cut-set bound [39] is given by:

$$\bar{C}(G, s, d) = \sup_{f(\mathbf{x}_1, \ldots, \mathbf{x}_N)} \min_{S: s \in S, d \in S^c} I(\mathbf{X}_S; \mathbf{Y}_{S^c} | \mathbf{X}_{S^c}), \tag{20}$$

where the supremum is over all joint distributions $f(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ on $\prod_{i=1}^N \mathbb{R}^{T_i}$ satisfying the power constraints $E\left[\|\mathbf{X}_i\|^2\right] \leq T_i P$ for $i \in \mathcal{N}$, the minimum is over all subsets $S \subset \mathcal{N}$ that separate s from d, and the conditional distribution of $\mathbf{y}_1, \ldots, \mathbf{y}_N$ given $\mathbf{x}_1, \ldots, \mathbf{x}_N$ is induced by the channel model (18).[5] Initiated by the work of Avestimehr *et al.* [6], there has been significant recent interest in approximating the capacity of general Gaussian relay networks with the cut-set bound, i.e. bounding the gap between the rates achieved by specific schemes and the cut-set bound on capacity. In particular, following a series of other works (e.g., [6], [8], [11]), Lim *et al.* [27] have proved the following approximation result:

Proposition 6 [27]: For any Gaussian relay network (G, s, d),

$$C(G, s, d) \geq \bar{C}(G, s, d) - 0.5\,\kappa(G, s, d). \tag{21}$$

Since $C(G, s, d) \leq \bar{C}(G, s, d)$ always, Proposition 6 establishes that the cut-set bound approximates the capacity $C(G, s, d)$ within a factor that is linear in the parameter κ but independent of the channel gains G. An interesting question is whether (21) can be substantially improved. For example, is it possible to replace the slack term 0.5κ with 0.1κ, or with a sublinear term such as $\frac{\kappa}{\log \log \kappa}$? This was posed as an open question by Niesen and Diggavi in [23] and by Avestimehr *et al.* in [21]. Some recent results [22]–[25] encourage this possibility by demonstrating that a sublinear in κ gap to the cut-set bound can be achieved when additional constraints are imposed on the topology of the network. As a specific example, it has been shown by Chern and Ozgur [24] that, for the diamond network with $N-2$ relays,

$$C(G, s, d) \geq \bar{C}(G, s, d) - 2\log(\kappa - 2) \tag{22}$$

when all nodes have one antenna (i.e., $\kappa = N$).

However, more recently Courtade and Ozgur observe in [13] that such an improvement is impossible, unless the cut-set bound is tight for *all* Gaussian relay networks. Toward doing so, they define a general template for approximating capacity via the cut-set bound. In the spirit of the approximation results proved in [6], [8], [11], and [22]–[27], a *Gaussian Relay Network Approximation Theorem* with parameter γ (abbreviated as γ-GRNAT) is defined to be a claim of the following form:

Claim 7: There exists a constant $\gamma \geq 0$ and a function $f(n) = o(n)$ such that, for any Gaussian relay network (G, s, d),

$$C(G, s, d) \geq \bar{C}(G, s, d) - (\gamma\kappa + f(\kappa)). \tag{23}$$

It should be emphasized that a γ-GRNAT makes an assertion that is independent of network topology, channel SNRs, and so forth. In particular, Proposition 6 provides a concrete example of a 0.5-GRNAT, with the $f(\kappa)$ term being zero.

The main result of [13] is to show that improving the linear term 0.5κ in (21) to a sublinear term $o(\kappa)$ is equivalent to proving the cut-set bound is tight for *all* Gaussian relay networks. This is formally stated as follows:

Proposition 8 [13]: A 0-GRNAT exists if and only if $C(G, s, d) = \bar{C}(G, s, d)$ for *all* Gaussian relay networks (G, s, d).

However, Courtade and Ozgur [13] also point out that they are not aware of any results which show that the cut-set bound is not tight for a Gaussian relay network. Combined with the result of the current paper, which shows that the cut-set bound is not tight for one specific Gaussian network, the above theorem asserts that the $\Theta(\kappa)$ term in approximations of the form (23) is fundamental. Note that the rate limited channel from the relay to the destination in Fig. 1 can be equivalently thought of as a Gaussian channel of the same capacity (c.f. [12]) and therefore the primitive relay setting we consider here can be thought of as one instance of a Gaussian relay network where the destination is equipped with two receive antennas, one directed to the source and one directed to the relay with no interference in between.

Since Proposition 8 asserts that the $\Theta(\kappa)$ term in approximations of the form (23) is fundamental, the following definition

[4]We adopt the notation and formulation in [13].

[5]Note that to be more precise P should be included in the characterization (G, s, d) of a network, however we avoid that to keep our notation consistent with [13].

is well-motivated:

$$\gamma^\star = \inf\{\gamma : \text{a } \gamma\text{-GRNAT holds}\}. \tag{24}$$

In words, $\gamma^\star$ characterizes the best possible linear factor in (23). Clearly, Propositions 6 and 8 imply that

$$0 < \gamma^\star \leq 0.5. \tag{25}$$

To this end, the following observation, noted in [13], implies that an explicit gap to the cut-set bound for any specific network with specific channel parameters and topology can be used to obtain a lower bound on $\gamma^\star$:

Proposition 9 [13]: If (G, s, d) is a Gaussian relay network and $C(G, s, d) \leq \bar{C}(G, s, d) - \beta$, then

$$\gamma^\star \geq \frac{\beta}{\kappa(G, s, d)}. \tag{26}$$

Therefore, the gap 0.0535 in Proposition 5 for the Gaussian primitive relay channel combined with the fact that $\kappa(G, s, d) = 4$ for this network implies that

$$\gamma^\star \geq 0.01.$$

In other words, the capacity of Gaussian relay networks can not be approximated by the cut-set bound within a gap that is smaller than $(0.0535/4)\kappa \approx 0.01\kappa$. A more recent result we prove in [28] demonstrates a gap of 0.2075 for the Gaussian primitive relay channel and implies an improved lower bound on $\gamma^\star$,

$$\gamma^\star \geq 0.05.$$

IV. Proof of Theorem 3

In this section we prove Theorem 3 for both the symmetric ($N_1 = N_2$) and asymmetric ($N_1 \neq N_2$) cases. The proofs for both cases rely on the below lemma, which is the main technical focus of this paper and whose proof is provided in Section V. We now state this lemma and show how it leads to the bound in Theorem 3.

Lemma 10: Consider any discrete random vector $X^n \in \mathbb{R}^n$. Let $Z^n = X^n + W_1^n$ 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. Also let $I_n = f_n(Z^n)$ be a function of Z^n which takes value on a finite set. Then, if $H(I_n|X^n) = na_n$, we have

$$I(X^n; I_n) - I(Y^n; I_n) \leq n(a_n + \sqrt{2a_n \ln 2} \log e). \tag{27}$$

Note that $I_n - Z^n - X^n - Y^n$ in the above lemma form a Markov chain and the result of the lemma can be equivalently regarded as fixing $I(X^n; I_n) = H(I_n) - na_n$ and controlling the second mutual information $I(Y^n; I_n)$. In this sense, there is some similarity in flavor between our result (27) and the strong data processing inequality [35]. However, when deriving strong data processing inequalities one is typically interested in upper bounding $I(Y^n; I_n)$ while we are interested in lower bounding it. Moreover, here we assume more specific structure for the Markov chain $I_n - Z^n - X^n - Y^n$.

Equipped with the above lemma, we are now ready to prove Theorem 3.

A. Symmetric Case ($N_1 = N_2$)

First consider the symmetric case when $N_1 = N_2 := N$. Suppose a rate R is achievable. Then there exists a sequence of codes

$$\{(\mathcal{C}_{(n,R)}, f_n, g_n)\}_{n=1}^{\infty} \tag{28}$$

such that the average probability of error $P_e^{(n)} \to 0$ as $n \to \infty$. For this sequence of codes, we have

$$\begin{aligned}
nR &= H(M) \\
&= I(M; Y^n, Z^n) + H(M|Y^n, Z^n) \\
&\leq I(X^n; Y^n, Z^n) + H(M|Y^n, f_n(Z^n)) \\
&\leq I(X^n; Y^n, Z^n) + n\mu \qquad (29)\\
&= h(Y^n, Z^n) - h(Y^n, Z^n|X^n) + n\mu \\
&= \sum_{i=1}^{n} [h(Y_i, Z_i|Y^{i-1}, Z^{i-1}) - h(Y_i, Z_i|X_i)] + n\mu \\
&\leq \sum_{i=1}^{n} [h(Y_i, Z_i) - h(Y_i, Z_i|X_i)] + n\mu \\
&= \sum_{i=1}^{n} I(X_i; Y_i, Z_i) + n\mu \\
&= n(I(X_Q; Y_Q, Z_Q|Q) + \mu) \qquad (30)\\
&= n(h(Y_Q, Z_Q|Q) - h(Y_Q, Z_Q|Q, X_Q) + \mu) \\
&\leq n(h(Y_Q, Z_Q) - h(Y_Q, Z_Q|X_Q) + \mu) \\
&= n(I(X_Q; Y_Q, Z_Q) + \mu)
\end{aligned}$$

i.e.,

$$R \leq I(X_Q; Y_Q, Z_Q) + \mu \tag{31}$$

for any $\mu > 0$ and sufficiently large n, where (29) follows from Fano's inequality, (30) follows by defining the time sharing random variable Q to be uniformly distributed over $[1 : n]$, and

$$E[X_Q^2] = \frac{1}{n} \sum_{i=1}^{n} E[X_i^2] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i^2\right] \leq P. \tag{32}$$

Moreover, letting $I_n := f_n(Z^n)$, we have for any $\mu > 0$ and sufficiently large n,

$$\begin{aligned}
nR &= H(M) \\
&= I(M; Y^n, I_n) + H(M|Y^n, I_n) \\
&\leq I(X^n; Y^n, I_n) + n\mu \qquad (33)\\
&= I(X^n; Y^n) + I(X^n; I_n|Y^n) + n\mu \\
&= I(X^n; Y^n) + H(I_n|Y^n) - H(I_n|X^n) + n\mu \qquad (34)\\
&\leq n(I(X_Q; Y_Q) + R_0 - a_n + \mu),
\end{aligned}$$

i.e.,

$$R \leq I(X_Q; Y_Q) + R_0 - a_n + \mu, \tag{35}$$

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

$$0 \leq a_n \leq R_0. \tag{36}$$

Note that in (34) we use the fact that $H(I_n|Y^n, X^n) = H(I_n|X^n)$ due to the Markov chain $I_n - X^n - Y^n$.

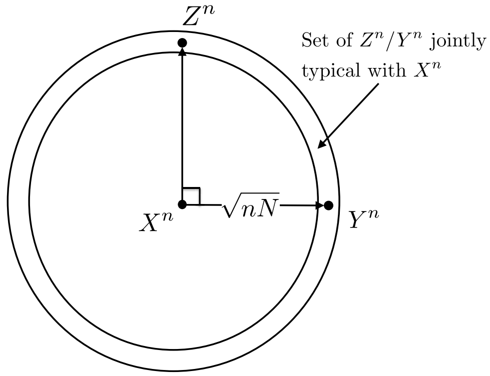

Fig. 2. Jointly typical set with X^n.

So far we have made only standard information theoretic arguments and in particular recovered the cut-set bound; note that the fact that $a_n \geq 0$ together with (31), (32) and (35) yields the cut-set bound given in Proposition 1. However, instead of simply lower bounding a_n by 0 in (35), in the sequel we will apply Lemma 10 and prove a third inequality involving a_n that forces a_n to be strictly larger than 0. Indeed, it is intuitively easy to see that a_n can not be arbitrarily small. Assume $a_n = \frac{1}{n} H(I_n|X^n) \approx 0$. Roughly speaking, this implies that given the transmitted codeword X^n, there is no ambiguity about I_n, i.e. I_n is a deterministic function of X^n. Recalling that I_n is mapping of Z^n to a set of integers, this means that all the Z^n sequences except on a set of zero measure, i.e. all Z^n sequences jointly typical with X^n are mapped to a single I_n. See Fig. 2. However, since Y^n and Z^n are statistically equivalent given X^n (they share the same typical set given X^n) this would imply that I_n can be determined based on Y^n and therefore $H(I_n|Y^n) \approx 0$, which forces the rate to be even smaller than $I(X_Q; Y_Q)$ in view of (34). In general, there is a trade-off between how close the rate can get to the multiple-access bound $I(X_Q; Y_Q) + R_0$ and how much it can exceed the point-to-point capacity $I(X_Q; Y_Q)$ of the X-Y link. We capture this trade-off as follows.

Adding and subtracting $H(I_n)$ to the R.H.S. of (34), we have

$$nR \leq I(X^n; Y^n) + I(X^n; I_n) - I(Y^n; I_n) + n\mu. \tag{37}$$

We now apply Lemma 10 to upper bound $I(X^n; I_n) - I(Y^n; I_n)$ in the above inequality. First note that the random variables (I_n, Z^n, X^n, Y^n) associated with our relay channel trivially satisfy the conditions of Lemma 10. In particular, X^n in our case is a discrete random vector whose distribution is dictated by the uniform distribution on the set of possible messages and the source codebook, Y^n and Z^n are continuous random vectors and I_n is an integer valued random variable. In light of this, Lemma 10 combined with (37) immediately yields that

$$nR \leq n(I(X_Q; Y_Q) + a_n + \sqrt{2a_n \ln 2} \log e + \mu),$$

i.e.,

$$R \leq I(X_Q; Y_Q) + a_n + \sqrt{2a_n \ln 2} \log e + \mu. \tag{38}$$

Combining (31), (35) and (38), we conclude that if a rate R is achievable, then for any $\mu > 0$ and sufficiently large n,

$$\begin{cases} R \leq I(X_Q; Y_Q, Z_Q) + \mu \\ R \leq I(X_Q; Y_Q) + R_0 - a_n + \mu \\ R \leq I(X_Q; Y_Q) + a_n + \sqrt{2a_n \ln 2} \log e + \mu \end{cases}$$

where $E[X_Q^2] \leq P$ and $a_n \in [0, R_0]$. Since μ can be made arbitrarily small, this proves Theorem 3 for the symmetric case.

B. Asymmetric Case ($N_1 \neq N_2$)

We now prove Theorem 3 for the asymmetric case when $N_1 \neq N_2$. Note that the proofs of (5)–(6) in this case follow exactly the same lines as their proofs in the symmetric case, i.e., by applying Fano's inequality and letting $H(I_n|X^n) = na_n$, so in the sequel we only prove (7) for $N_1 \neq N_2$.

First assume $N_1 < N_2$. In this case we 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} and W_{22} are zero-mean Gaussian random variables with variances N_1, N_1 and $N_2 - N_1$ respectively, and they are independent of each other and X. Based on this, we write

$$\begin{cases} Z^n = X^n + W_1^n & (39) \\ Y^n = \tilde{Z}^n + W_{22}^n & (40) \end{cases}$$

where

$$\tilde{Z}^n := X^n + W_{21}^n. \tag{41}$$

To prove (7) for $N_1 < N_2$, we continue with (33) and modify the proof for the symmetric case to be:

$$\begin{aligned} nR &\leq I(X^n; Y^n, I_n) + n\mu \\ &\leq I(X^n; \tilde{Z}^n, I_n) + n\mu \\ &= I(X^n; \tilde{Z}^n) + I(X^n; I_n) - I(\tilde{Z}^n; I_n) + n\mu, \end{aligned}$$

where the second inequality follows from the data processing inequality applied to the Markov chain $X^n - (\tilde{Z}^n, I_n) - (Y^n, I_n)$. Now observe that $(I_n, Z^n, X^n, \tilde{Z}^n)$ satisfy the conditions of Lemma 10 and therefore we have

$$\begin{aligned} nR &\leq nI(X_Q; \tilde{Z}_Q) + n(a_n + \sqrt{2a_n \ln 2} \log e) + n\mu \\ &= n(I(X_Q; Z_Q) + a_n + \sqrt{2a_n \ln 2} \log e + \mu), \end{aligned}$$

where $a_n = \frac{1}{n} H(I_n|X^n)$. This proves constraint (7) for the $N_1 < N_2$ case.

Now consider the case when $N_1 > N_2$. 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 $N_1 - N_2$, and is independent of the other random variables in the problem. Applying Lemma 10 to $(I_n, Z^n, X^n, \tilde{Z}^n)$ we have

$$I(X^n; I_n) - I(\tilde{Z}^n; I_n) \le n(a_n + \sqrt{2a_n \ln 2} \log e),$$

which combined with the Markov relation $I_n - X^n - Y^n - \tilde{Z}^n$ further implies that

$$I(X^n; I_n) - I(Y^n; I_n) \le n(a_n + \sqrt{2a_n \ln 2} \log e). \quad (42)$$

Combining this with inequality (37) then proves constraint (7) for the $N_1 > N_2$ case and concludes the proof of Theorem 3.

V. Proof of Lemma 10

We now prove Lemma 10. For this, we first introduce some auxiliary results that will be used in the proof.

A. Auxiliary Results

We begin with the following measure concentration result which is a blowing-up lemma for general Gaussian random variables.

Lemma 11: For any subset $A \subseteq \mathbb{R}^n$, let $\Gamma_r(A)$ be its blown-up set of radius r defined as

$$\Gamma_r(A) := \left\{b^n \in \mathbb{R}^n : \exists\ a^n \in A \ \text{ s.t. } \ d(a^n, b^n) \le r\right\},$$

where

$$d(a^n, b^n) := \|a^n - b^n\| \quad (43)$$

denotes the Euclidean distance between the two sequences a^n and b^n. Let $U_1, U_2, \dots, U_n$ be n i.i.d. Gaussian random variables with $U_i \sim \mathcal{N}(0, N), \forall i \in \{1, 2, \dots, n\}$. Then, for any $A \subseteq \mathbb{R}^n$ with $\Pr(U^n \in A) \ge 2^{-na_n}$,

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

Lemma 11 is essentially due to Marton [29] (see also [30]– [31]). In Appendix B, we provide a simple proof of Lemma 11 that extends from [32, eq. (1.6)], which is a version of the lemma stated by Talagrand for standard Gaussian random variables. For more discussions on the blowing-up lemma and its applications, see the recent comprehensive monograph by Raginsky and Sason [34].

The next lemma constructs a series of typical sets in a nested manner (for a mixed set of continuous and discrete random variables) which satisfy certain properties that will be used in the proof of Lemma 10.

Lemma 12: Let $X - Z - I$ form a Markov chain where X and I are discrete random variables (or vectors), Z is a continuous random variable (or vector) and I is a deterministic function of Z. Let $(\mathbf{X}, \mathbf{Z}, \mathbf{I})$ be a B-length sequence i.i.d. generated from the joint distribution of (X, Z, I), with $\mathbf{I}$ being a function of $\mathbf{Z}$ denoted by $\mathbf{I} = f(\mathbf{Z})$. Then one can construct a series of typical sets satisfying certain properties as follows:

1) Let $S_\epsilon^{(B)}(X, Z, I)$ be the set of $(\mathbf{x}, \mathbf{z}, \mathbf{i})$ sequences defined as

$$\begin{aligned} S_\epsilon^{(B)}(X, Z, I) = \{(\mathbf{x}, \mathbf{z}, \mathbf{i}) :& \\ 2^{-B(h(Z|X)+\epsilon)} &\le f(\mathbf{z}|\mathbf{x}) \le 2^{-B(h(Z|X)-\epsilon)}, \\ 2^{-B(H(X|Z)+\epsilon)} &\le p(\mathbf{x}|\mathbf{z}) \le 2^{-B(H(X|Z)-\epsilon)}, \\ 2^{-B(H(I|X)+\epsilon)} &\le p(\mathbf{i}|\mathbf{x}) \le 2^{-B(H(I|X)-\epsilon)}, \\ 2^{-B(H(X|I)+\epsilon)} &\le p(\mathbf{x}|\mathbf{i}) \le 2^{-B(H(X|I)-\epsilon)}\}. \end{aligned}$$

Then for any $\epsilon > 0$ and B sufficiently large,

$$\Pr((\mathbf{X}, \mathbf{Z}, \mathbf{I}) \in S_\epsilon^{(B)}(X, Z, I)) \ge 1 - \epsilon.$$

2) For any $\mathbf{z}$, let $S_\epsilon^{(B)}(X|\mathbf{z})$ be the set of $\mathbf{x}$ sequences defined as

$$S_\epsilon^{(B)}(X|\mathbf{z}) = \{\mathbf{x} : (\mathbf{x}, \mathbf{z}, f(\mathbf{z})) \in S_\epsilon^{(B)}(X, Z, I)\},$$

and let $S_\epsilon^{(B)}(X, Z)$ be the set of $(\mathbf{x}, \mathbf{z})$ sequences defined as

$$\begin{aligned} &S_\epsilon^{(B)}(X, Z) = \{(\mathbf{x}, \mathbf{z}) : \\ &\quad \mathbf{x} \in S_\epsilon^{(B)}(X|\mathbf{z}), \Pr(\mathbf{X} \in S_\epsilon^{(B)}(X|\mathbf{z})|\mathbf{Z} = \mathbf{z}) \ge 1 - \sqrt{\epsilon}\}. \end{aligned}$$

Then for any $\epsilon > 0$ and B sufficiently large,

$$\Pr((\mathbf{X}, \mathbf{Z}) \in S_\epsilon^{(B)}(X, Z)) \ge 1 - 2\sqrt{\epsilon}.$$

3) For any $(\mathbf{x}, \mathbf{i})$, let $S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i})$ be the set of $\mathbf{z}$ sequences defined as

$$S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i}) = \{\mathbf{z} : f(\mathbf{z}) = \mathbf{i}, (\mathbf{x}, \mathbf{z}) \in S_\epsilon^{(B)}(X, Z)\},$$

and let $S_\epsilon^{(B)}(X, I)$ be the set of $(\mathbf{x}, \mathbf{i})$ sequences defined as

$$\begin{aligned} &S_\epsilon^{(B)}(X, I) = \{(\mathbf{x}, \mathbf{i}) : \\ &\quad \Pr(\mathbf{Z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i})|\mathbf{X} = \mathbf{x}, \mathbf{I} = \mathbf{i}) \ge 1 - \sqrt[4]{\epsilon}\}. \end{aligned}$$

Then for any $\epsilon > 0$ and B sufficiently large,

$$\Pr((\mathbf{X}, \mathbf{I}) \in S_\epsilon^{(B)}(X, I)) \ge 1 - 2\sqrt[4]{\epsilon}.$$

Furthermore, for any $(\mathbf{x}, \mathbf{i}) \in S_\epsilon^{(B)}(X, I)$,

$$2^{-B(H(I|X)+\epsilon)} \le p(\mathbf{i}|\mathbf{x}) \le 2^{-B(H(I|X)-\epsilon)},$$

and for B sufficiently large,

$$\Pr(\mathbf{Z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i})|\mathbf{X} = \mathbf{x}) \ge 2^{-B(H(I|X)+2\epsilon)}.$$

4) For any $\mathbf{i}$, let $S_\epsilon^{(B)}(X|\mathbf{i})$ be the set of $\mathbf{x}$ sequences defined as

$$S_\epsilon^{(B)}(X|\mathbf{i}) = \{\mathbf{x} : (\mathbf{x}, \mathbf{i}) \in S_\epsilon^{(B)}(X, I)\},$$

and let $S_\epsilon^{(B)}(I)$ be the set of $\mathbf{i}$ sequences defined as

$$S_\epsilon^{(B)}(I) = \{\mathbf{i} : \Pr(\mathbf{X} \in S_\epsilon^{(B)}(X|\mathbf{i})|\mathbf{I} = \mathbf{i}) \ge 1 - 2\sqrt[8]{\epsilon}\}.$$

Then for any $\epsilon > 0$ and B sufficiently large,

$$\Pr(\mathbf{I} \in S_\epsilon^{(B)}(I)) \ge 1 - 2\sqrt[8]{\epsilon}.$$

Note that if X, Z, I were all discrete, we could directly use the strongly typical sets as defined in [38, Ch. 2] and

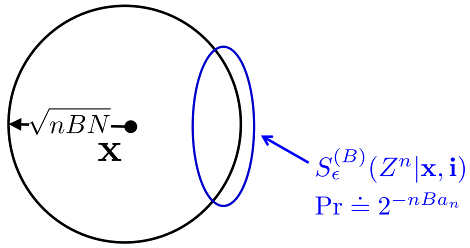

Fig. 3. The set $S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$.

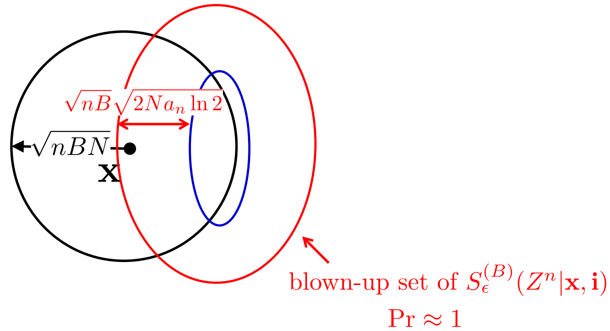

Fig. 4. Blow up the set $S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$.

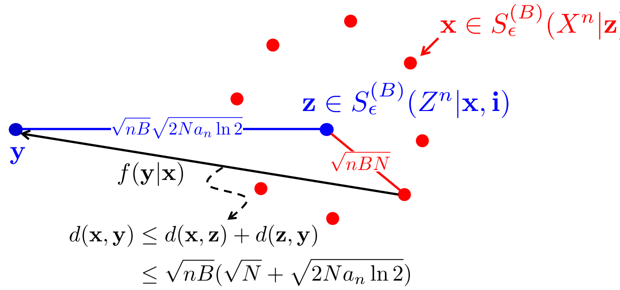

Fig. 5. Lower bound the conditional density $f(\mathbf{y}|\mathbf{i})$.

all the properties above will naturally follow.[6] Unfortunately, there is no counterpart of strong typicality for continuous random variables and if one uses weak typicality [39, Ch. 3] instead then the above properties can not be all guaranteed. Under this circumstance, Lemma 12 says that by its discussed nested construction one can still have typical sets satisfying all these properties, however the proof of this lemma is more complex than simply invoking weak typicality and is included in Appendix C.

B. Proof of Lemma 10

We are now ready to prove Lemma 10. For this we will lift the random variables X^n, Y^n, Z^n and I_n to a higher dimensional, say nB dimensional space, and invoke the typical sets as constructed in Lemma 12. Specifically, consider the following B-length i.i.d. sequence

$$\{(X^n(b), Y^n(b), Z^n(b), I_n(b))\}_{b=1}^{B}, \tag{44}$$

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 sequence $[X^n(1), X^n(2), \dots, X^n(B)]$ as $\mathbf{X}$ and similarly define $\mathbf{Y}, \mathbf{Z}$ and $\mathbf{I}$; note here we have $\mathbf{I} = [f_n(Z^n(1)), f_n(Z^n(2)), \dots, f_n(Z^n(B))] =: f(\mathbf{Z})$. Since the random variables (X^n, Z^n, I_n) under consideration satisfy the condition of Lemma 12, i.e. $X^n - Z^n - I_n$ form a Markov chain where X^n and I_n are discrete, Z^n is continuous and I_n is a deterministic function of Z^n, we have the series of typical sets for $(\mathbf{X}, \mathbf{Z}, \mathbf{I})$ as described in Lemma 12. Our proof in the sequel will build on these typical sets and their properties, as well as the blowing-up lemma for Gaussian measure stated in Lemma 11.

In particular, from Lemma 12-3), for any $(\mathbf{x}, \mathbf{i}) \in S^{(B)}(X^n, I_n)$ and B sufficiently large,

$$\begin{aligned}\Pr(\mathbf{Z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})|\mathbf{X}=\mathbf{x}) &\ge 2^{-B(H(I_n|X^n)+2\epsilon)} \\ &= 2^{-nB(a_n+2\epsilon)},\end{aligned} \tag{45}$$

where $a_n := \frac{1}{n}H(I_n|X^n)$. See Fig. 3. We now blow up the set $S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$ and use Lemma 11 to show that if the blowing-up radius is about $\sqrt{nB}\sqrt{2Na_n \ln 2}$, then the resultant blown-up set has probability nearly 1. In particular, noting that $\mathbf{Z}$ is Gaussian given $\mathbf{X}$ and using Lemma 11, we have for B sufficiently large,

$$\begin{aligned}&\Pr(\mathbf{Z} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n \ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))|\mathbf{X}=\mathbf{x}) \\ &\ge \Pr(\mathbf{Z} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n \ln 2+4N\epsilon \ln 2}+\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))|\mathbf{X}=\mathbf{x}) \\ &\ge 1 - 2^{-\frac{nB\epsilon}{2}} \\ &\ge 1-\epsilon.\end{aligned}$$

See Fig. 4. Since $\mathbf{Y}$ and $\mathbf{Z}$ are identically distributed given $\mathbf{X}$, we also have

$$\begin{aligned}\Pr(\mathbf{Y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n \ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))|\mathbf{X}=\mathbf{x}) \\ \ge 1-\epsilon.\end{aligned} \tag{46}$$

We next lower bound the conditional density $f(\mathbf{y}|\mathbf{i})$ for each $\mathbf{y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n \ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))$. The approach is geometric and the readers may facilitate their understanding by referring to Fig. 5. First, note that for each $\mathbf{y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n \ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))$, there exists (at least) one $\mathbf{z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$ such that $d(\mathbf{y},\mathbf{z}) \le \sqrt{nB}(\sqrt{2Na_n \ln 2}+3\sqrt{N\epsilon})$. By the construction of $S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$ as in Lemma 12-3), for this $\mathbf{z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$, we have $f(\mathbf{z}) = \mathbf{i}$ and $(\mathbf{x},\mathbf{z}) \in S_\epsilon^{(B)}(X^n, Z^n)$. Using the definition of $S_\epsilon^{(B)}(X^n, Z^n)$ in Lemma 12-2), we further have

$$\Pr(\mathbf{X} \in S_\epsilon^{(B)}(X^n|\mathbf{z})|\mathbf{Z}=\mathbf{z}) \ge 1-\sqrt{\epsilon},$$

where, by Lemma 12-1) and 2), $S_\epsilon^{(B)}(X^n|\mathbf{z})$ consists of all $\mathbf{x}$ satisfying the following properties:

$$2^{-B(h(Z^n|X^n)+\epsilon)} \le f(\mathbf{z}|\mathbf{x}) \le 2^{-B(h(Z^n|X^n)-\epsilon)}, \tag{47}$$

$$2^{-B(H(X^n|Z^n)+\epsilon)} \le p(\mathbf{x}|\mathbf{z}) \le 2^{-B(H(X^n|Z^n)-\epsilon)}, \tag{48}$$

$$2^{-B(H(I_n|X^n)+\epsilon)} \le p(\mathbf{i}|\mathbf{x}) \le 2^{-B(H(I_n|X^n)-\epsilon)}, \tag{49}$$

$$2^{-B(H(X^n|I_n)+\epsilon)} \le p(\mathbf{x}|\mathbf{i}) \le 2^{-B(H(X^n|I_n)-\epsilon)}. \tag{50}$$

[6] Indeed, in our parallel paper [17] which considers improving on the cut-set for discrete memoryless relay channels, we directly resort to the notion of strong typicality instead of using the typicality sets as discussed in Lemma 12.

Note that since by assumption Z^n is Gaussian given X^n, we have $h(Z^n|X^n) = \frac{n}{2}\log 2\pi eN$ and

$$f(\mathbf{z}|\mathbf{x}) = \frac{1}{(2\pi N)^{\frac{nB}{2}}} e^{-\frac{\|\mathbf{y}-\mathbf{x}\|^2}{2N}},$$

and therefore property (47) can be shown to imply that

$$d(\mathbf{x},\mathbf{z}) \in [\sqrt{nBN(1-2\epsilon)},\quad \sqrt{nBN(1+2\epsilon)}]. \tag{51}$$

Moreover, with property (48) we can lower bound the size of $S_\epsilon^{(B)}(X^n|\mathbf{z})$ by considering the following:

$$\begin{aligned} 1-\sqrt{\epsilon} &\le \Pr(\mathbf{X} \in S_\epsilon^{(B)}(X^n|\mathbf{z})|\mathbf{Z}=\mathbf{z}) \\ &= \sum_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{z})} p(\mathbf{x}|\mathbf{z}) \\ &\le 2^{-B(H(X|Z)-\epsilon)} \left|S_\epsilon^{(B)}(X^n|\mathbf{z})\right|, \end{aligned}$$

i.e.,

$$\left|S_\epsilon^{(B)}(X^n|\mathbf{z})\right| \ge (1-\sqrt{\epsilon})2^{B(H(X^n|Z^n)-\epsilon)}. \tag{52}$$

Based on (50), (51) and (52), we now lower bound $f(\mathbf{y}|\mathbf{i})$ for each $\mathbf{y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))$. In particular, we have for B sufficiently large,

$$\begin{aligned} f(\mathbf{y}|\mathbf{i}) &= \sum_{\mathbf{x}} f(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i}) \\ &\ge \sum_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{z})} f(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i}) \\ &\ge 2^{-B(H(X^n|I_n)+\epsilon)} \sum_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{z})} f(\mathbf{y}|\mathbf{x}) \qquad (53) \\ &\ge 2^{-B(H(X^n|I_n)+\epsilon)} \left|S_\epsilon^{(B)}(X^n|\mathbf{z})\right| \min_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{z})} f(\mathbf{y}|\mathbf{x}) \\ &\ge (1-\sqrt{\epsilon})2^{-B(H(X^n|I_n)+\epsilon)} \\ &\quad \times 2^{B(H(X^n|Z^n)-\epsilon)} \min_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{z})} f(\mathbf{y}|\mathbf{x}), \qquad (54) \end{aligned}$$

where the $\mathbf{z}$ throughout the above is the $\mathbf{z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$ such that $d(\mathbf{y},\mathbf{z}) \le \sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})$, (53) follows from (50), and (54) follows from (52). To lower bound the last term in (54), note that for any $\mathbf{x} \in S_\epsilon^{(B)}(X^n|\mathbf{z})$, we have due to (51) that

$$\begin{aligned} d(\mathbf{x},\mathbf{y}) &\le d(\mathbf{x},\mathbf{z}) + d(\mathbf{z},\mathbf{y}) \\ &\le \sqrt{nB}(\sqrt{N(1+2\epsilon)} + \sqrt{2Na_n\ln 2} + 3\sqrt{N\epsilon}) \\ &=: \sqrt{nB}(\sqrt{N} + \sqrt{2Na_n\ln 2} + \epsilon_1) \end{aligned}$$

and thus,

$$\begin{aligned} f(\mathbf{y}|\mathbf{x}) &= \frac{1}{(2\pi N)^{\frac{nB}{2}}} e^{-\frac{\|\mathbf{y}-\mathbf{x}\|^2}{2N}} \\ &\ge 2^{-\frac{nB(\sqrt{N}+\sqrt{2Na_n\ln 2}+\epsilon_1)^2}{2N}\log e - \frac{nB}{2}\log 2\pi N} \\ &= 2^{-nB\left(\frac{(\sqrt{N}+\sqrt{2Na_n\ln 2}+\epsilon_1)^2}{2N}\log e + \frac{1}{2}\log 2\pi N\right)} \\ &=: 2^{-nB\left(\frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n\ln 2}\log e + \epsilon_2\right)} \end{aligned}$$

where $\epsilon_1, \epsilon_2 \to 0$ as $\epsilon \to 0$. Plugging this into (54) yields that

$$\begin{aligned} &f(\mathbf{y}|\mathbf{i}) \\ &\ge (1-\sqrt{\epsilon})2^{-B(H(X^n|I_n)+\epsilon)}2^{B(H(X^n|Z^n)-\epsilon)} \\ &\quad \times 2^{-nB\left(\frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n\ln 2}\log e + \epsilon_2\right)} \\ &\ge 2^{-B\left[H(X^n|I_n) - H(X^n|Z^n) + n(\frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n\ln 2}\log e + \epsilon_3)\right]} \end{aligned} \tag{55}$$

for any $\mathbf{y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))$ and B sufficiently large, where $\epsilon_3 \to 0$ as $\epsilon \to 0$.

For any $\mathbf{i} \in S_\epsilon^{(B)}(I_n)$, let $\mathcal{Y}_{\mathbf{i}}$ be a set of $\mathbf{y}$ sequences defined as

$$\mathcal{Y}_{\mathbf{i}} := \bigcup_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{i})} \Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})).$$

Then for each $\mathbf{y} \in \mathcal{Y}_{\mathbf{i}}$, there exists some $\mathbf{x} \in S_\epsilon^{(B)}(X^n|\mathbf{i})$ such that $\mathbf{y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))$, and by (55) it follows that for B sufficiently large,

$$\begin{aligned} &f(\mathbf{y}|\mathbf{i}) \\ &\quad \ge 2^{-B\left[H(X^n|I_n) - H(X^n|Z^n) + n(\frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n\ln 2}\log e + \epsilon_3)\right]}. \end{aligned}$$

Moreover, for any $\mathbf{i} \in S_\epsilon^{(B)}(I_n)$, we have for B sufficiently large,

$$\begin{aligned} &\Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{I}=\mathbf{i}) \\ &\quad = \sum_{\mathbf{x}} \Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{X}=\mathbf{x})p(\mathbf{x}|\mathbf{i}) \\ &\quad \ge \sum_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{i})} \Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{X}=\mathbf{x})p(\mathbf{x}|\mathbf{i}) \\ &\quad \ge \sum_{\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{i})} \Pr(\mathbf{Y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})} \\ &\qquad \times (S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))|\mathbf{X}=\mathbf{x})p(\mathbf{x}|\mathbf{i}) \\ &\quad \ge (1-\epsilon)\Pr(\mathbf{X} \in S_\epsilon^{(B)}(X^n|\mathbf{i})|\mathbf{I}=\mathbf{i}) \qquad (56) \\ &\quad \ge (1-\epsilon)(1-2\sqrt[8]{\epsilon}) \qquad (57) \\ &\quad \ge 1-3\sqrt[8]{\epsilon}, \end{aligned}$$

where (56) follows from (46) and (57) follows from Lemma 12-4). Finally, recalling from Lemma 12-4) that

$$\Pr(\mathbf{I} \in S_\epsilon^{(B)}(I)) \ge 1-2\sqrt[8]{\epsilon}$$

and choosing δ to be $\max\{3\sqrt[8]{\epsilon}, \epsilon_3\}$, we arrive at the following proposition.

Proposition 13: For any $\delta > 0$ and sufficiently large B, there exists a set $\mathcal{I}$ of $\mathbf{i}$ such that

$$\Pr(\mathbf{I} \in \mathcal{I}) \ge 1-\delta,$$

and for any $\mathbf{i} \in \mathcal{I}$, there exists a set $\mathcal{Y}_{\mathbf{i}}$ of $\mathbf{y}$ satisfying

$$\Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{I}=\mathbf{i}) \ge 1-\delta,$$

and for any $\mathbf{y} \in \mathcal{Y}_{\mathbf{i}}$

$$\begin{aligned} &f(\mathbf{y}|\mathbf{i}) \\ &\quad \ge 2^{-B(H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN + na_n + n\sqrt{2a_n\ln 2}\log e + n\delta)}. \end{aligned}$$

We now use Proposition 13 to prove Lemma 10. For this, first consider $h(\mathbf{Y}|\mathbf{i})$ for any $\mathbf{i} \in \mathcal{I}$. We have

$$\begin{aligned} h(\mathbf{Y}|\mathbf{i}) &\le h(\mathbf{Y}|\mathbf{i}) + 1 - I(\mathbf{Y}; \mathbb{I}(\mathbf{Y} \in \mathcal{Y}_\mathbf{i})|\mathbf{i}) \quad (58)\\ &= 1 + h(\mathbf{Y}|\mathbb{I}(\mathbf{Y} \in \mathcal{Y}_\mathbf{i}), \mathbf{i})\\ &= 1 + \Pr(\mathbf{Y} \in \mathcal{Y}_\mathbf{i}|\mathbf{I} = \mathbf{i}) h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i})\\ &\quad + \Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{I} = \mathbf{i}) h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \notin \mathcal{Y}_\mathbf{i}), \quad (59) \end{aligned}$$

where $\mathbb{I}(A)$ is the indicator function defined as 1 if A holds and 0 otherwise, and (58) follows since

$$I(\mathbf{Y}; \mathbb{I}(\mathbf{Y} \in \mathcal{Y}_\mathbf{i})|\mathbf{i}) \le H(\mathbb{I}(\mathbf{Y} \in \mathcal{Y}_\mathbf{i})|\mathbf{i}) \le 1.$$

To bound $h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i})$, we have by Proposition 13 that,

$$\begin{aligned} &h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i})\\ &= -\int_{\mathbf{y}\in\mathcal{Y}_\mathbf{i}} f(\mathbf{y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) \log f(\mathbf{y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) d\mathbf{y}\\ &\le -\int_{\mathbf{y}\in\mathcal{Y}_\mathbf{i}} f(\mathbf{y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) \log f(\mathbf{y}|\mathbf{i}) d\mathbf{y}\\ &\le B\Big(H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN + na_n\\ &\quad + n\sqrt{2a_n \ln 2}\log e + n\delta\Big) \cdot \int_{\mathbf{y}\in\mathcal{Y}_\mathbf{i}} f(\mathbf{y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) d\mathbf{y}\\ &= B\Big(H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN + na_n\\ &\quad + n\sqrt{2a_n \ln 2}\log e + n\delta\Big). \quad (60) \end{aligned}$$

Now consider $E[\|\mathbf{Y}\|^2|\mathbf{i}]$ for any $\mathbf{i}$. We have

$$E[\|\mathbf{Y}\|^2|\mathbf{i}] = E[\|\mathbf{X}\|^2|\mathbf{i}] + E[\|\mathbf{W_2}\|^2|\mathbf{i}] \le nB(P + N),$$

where the equality follows from the independence between $\mathbf{X}$ and $\mathbf{W_2}$ even conditioned on $\mathbf{i}$. Therefore,

$$E[\|\mathbf{Y}\|^2|\mathbf{i}, \mathbf{Y} \notin \mathcal{Y}_\mathbf{i}] \le \frac{E[\|\mathbf{Y}\|^2|\mathbf{i}]}{\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{i})} \le \frac{nB(P + N)}{\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{i})},$$

and

$$\begin{aligned} &\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{I} = \mathbf{i}) h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \notin \mathcal{Y}_\mathbf{i})\\ &\quad \le \frac{nB}{2}\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{I} = \mathbf{i}) \log 2\pi e \frac{P + N}{\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{I} = \mathbf{i})}\\ &\quad \le nB\delta_1, \quad (61) \end{aligned}$$

for some $\delta_1 \to 0$ as $\delta \to 0$.

Plugging (60) and (61) into (59), we have for any $\mathbf{i} \in \mathcal{I}$ and sufficiently large B,

$$\begin{aligned} h(\mathbf{Y}|\mathbf{i}) &\le \Pr(\mathbf{Y} \in \mathcal{Y}_\mathbf{i}|\mathbf{I} = \mathbf{i}) B\Big(H(X^n|I_n) - H(X^n|Z^n)\\ &\quad + \frac{n}{2}\log 2\pi eN + na_n + n\sqrt{2a_n \ln 2}\log e + n\delta\Big)\\ &\quad + 1 + nB\delta_1\\ &\le B\Big(H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN\\ &\quad + na_n + n\sqrt{2a_n \ln 2}\log e + n\delta_2 + 1/B\Big) \quad (62) \end{aligned}$$

for some $\delta_2 \to 0$ as $\delta \to 0$. Therefore, for sufficiently large B,

$$\begin{aligned} h(\mathbf{Y}|\mathbf{I}) &= \sum_i p(\mathbf{i}) h(\mathbf{Y}|\mathbf{i})\\ &= \sum_{\mathbf{i}\in\mathcal{I}} p(\mathbf{i}) h(\mathbf{Y}|\mathbf{i}) + \sum_{\mathbf{i}\notin\mathcal{I}} p(\mathbf{i}) h(\mathbf{Y}|\mathbf{i})\\ &\le \sum_{\mathbf{i}\in\mathcal{I}} p(\mathbf{i}) B\Big(H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN\\ &\quad + na_n + n\sqrt{2a_n \ln 2}\log e + n\delta_2 + 1/B\Big)\\ &\quad + \sum_{\mathbf{i}\notin\mathcal{I}} p(\mathbf{i}) \frac{nB}{2}\log 2\pi e(P + N)\\ &\le B\Big(H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN\\ &\quad + na_n + n\sqrt{2a_n \ln 2}\log e + n\delta_3 + 1/B\Big) \quad (63) \end{aligned}$$

for some $\delta_3 \to 0$ as $\delta \to 0$. Observing that

$$h(\mathbf{Y}|\mathbf{I}) = \sum_{b=1}^{B} h(Y^n(b)|I_n(b)) = Bh(Y^n|I_n)$$

and noting that both δ_3 and $1/B$ in (63) can be made arbitrarily small by choosing B sufficiently large, we obtain

$$\begin{aligned} h(Y^n|I_n) &\le H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN\\ &\quad + na_n + n\sqrt{2a_n \ln 2}\log e. \quad (64) \end{aligned}$$

Finally, using the relation (64), we have

$$\begin{aligned} &I(X^n; I_n) - I(Y^n; I_n)\\ &= H(X^n) - H(X^n|I_n) - h(Y^n) + h(Y^n|I_n)\\ &\le H(X^n) - H(X^n|I_n) - h(Y^n) + H(X^n|I_n)\\ &\quad - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN + na_n + n\sqrt{2a_n \ln 2}\log e\\ &= I(X^n; Z^n) - [h(Y^n) - \frac{n}{2}\log 2\pi eN]\\ &\quad + na_n + n\sqrt{2a_n \ln 2}\log e\\ &= I(X^n; Z^n) - I(X^n; Y^n) + na_n + n\sqrt{2a_n \ln 2}\log e\\ &= na_n + n\sqrt{2a_n \ln 2}\log \quad (65) \end{aligned}$$

where the last step follows because Z^n and Y^n are conditionally i.i.d. given X^n, i.e. $I(X^n; Z^n) = I(X^n; Y^n)$. This finishes the proof of Lemma 10.

VI. Further Improvement

In this section we show that in the case of $N_1 \le N_2$, our bound in Theorem 3 can be further sharpened for certain regimes of channel parameters. In particular, we will prove the following proposition.

Proposition 14: For a Gaussian primitive relay channel with $N_1 \le N_2$, if a rate R is achievable, then there exists some $a \in [0, R_0]$ such that (8), (9) and the following two constraints

$$\begin{cases} R \le \frac{1}{2}\log\left(1+\frac{P}{N_1}\right)+a+\sqrt{2a\ln 2}\log e & (66)\\ R \le \frac{1}{2}\log\left(1+\frac{P}{N_1}\right)+\frac{N_1}{N_2}a \\ \quad +\sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a\ln 2+1-\frac{N_1}{N_2}\right)}\log e & (67)\end{cases}$$

are simultaneously satisfied.

Proposition 14 improves upon Theorem 3 for the $N_1 \le N_2$ case by introducing a new constraint (67) that is structurally similar to (66). Note that neither constraint (66) nor (67) is dominating the other and which one is tighter depends on the channel parameter. This makes the bound in Proposition 14 in general tighter than that in Theorem 3 for the $N_1 \le N_2$ case. Nevertheless, in Appendix D we show that the largest gap between the bound in Proposition 14 and the cut-set bound remains to be 0.0535, which is still attained when $\frac{P}{N_1}=\frac{P}{N_2}\to\infty$ and $R_0=0.5$.

To show Proposition 14, we only need to show the new constraint (67), which follows immediately from the following lemma.

Lemma 15: Consider any discrete random vector $X^n \in \mathbb{R}^n$. Let $Z^n = X^n + W_1^n$ and $Y^n = X^n + W_2^n$, where W_1^n and W_2^n are i.i.d. sequences of Gaussian random variables with zero mean and variance N_1 and N_2 respectively, and they are independent of each other and X^n. Also let $I_n = f_n(Z^n)$ be a function of Z^n which takes value on a finite set. Then, if $N_1 \le N_2$ and $H(I_n|X^n) = na_n$, we have

$$I(X^n;I_n)-I(Y^n;I_n) \le I(X^n;Z^n)-I(X^n;Y^n) + n\left(\frac{N_1}{N_2}a_n+\sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a_n\ln 2+1-\frac{N_1}{N_2}\right)}\log e\right). \quad (68)$$

Note that Lemma 15 provides a new bound (68) on the difference $I(X^n;I_n)-I(Y^n;I_n)$ under the assumption of $N_1 \le N_2$. Since the random variables (X^n, Y^n, Z^n, I_n) associated with the relay channel for the $N_1 \le N_2$ case trivially satisfies the condition of Lemma 15, one can combine (68) with (37), which immediately yields the new constraint (67). In the sequel we focus on proving Lemma 15.

A. Proof of Lemma 15

Without loss of generality, write Z^n and Y^n as

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

with

$$\tilde{Z}^n := X^n + W_{21}^n,$$

where W_1^n, W_{21}^n and W_{22}^n are i.i.d. sequences of Gaussian random variables with zero mean and variance N_1, N_1 and $N_2 - N_1$ respectively, and they are independent of each other and X^n.

Consider the B-length i.i.d. extensions of the above random variables. By applying the typicality argument and blowing-up lemma along the same lines as in the proof of Lemma 10, we have for any $(\mathbf{x},\mathbf{i}) \in S^{(B)}(X^n, I_n)$ and B sufficiently large,

$$\Pr(\mathbf{Z}\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N_1\epsilon})}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))|\mathbf{X}=\mathbf{x}) \ge 1-\epsilon. \quad (69)$$

Since $\tilde{\mathbf{Z}}$ and $\mathbf{Z}$ are identically distributed given $\mathbf{X}$, (69) also holds with $\mathbf{Z}$ replaced by $\tilde{\mathbf{Z}}$. In other words, we have for any $(\mathbf{x},\mathbf{i}) \in S^{(B)}(X^n, I_n)$ and B sufficiently large,

$$\Pr(\exists\ \mathbf{z}\in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}) \text{ s.t. } d\ (\tilde{\mathbf{Z}},\mathbf{z}) \le \sqrt{nB}(\sqrt{2N_1a_n\ln 2}+3\sqrt{N_1\epsilon})|\mathbf{X}=\mathbf{x}) \ge 1-\epsilon. \quad (70)$$

Consider any specific pair of $(\tilde{\mathbf{z}},\mathbf{z})$ with $d\ (\tilde{\mathbf{z}},\mathbf{z}) \le \sqrt{nB}(\sqrt{2N_1a_n\ln 2}+3\sqrt{N_1\epsilon})$ and recall $\mathbf{Y}=\tilde{\mathbf{z}}+\mathbf{W}_{22}$. We have

$$\begin{aligned} d^2(\mathbf{Y},\mathbf{z}) &= \|\mathbf{Y}-\mathbf{z}\|^2 \\ &= \|\mathbf{W}_{22}+\tilde{\mathbf{z}}-\mathbf{z}\|^2 \\ &= [\mathbf{W}_{22}+(\tilde{\mathbf{z}}-\mathbf{z})]^T[\mathbf{W}_{22}+(\tilde{\mathbf{z}}-\mathbf{z})] \\ &= \|\mathbf{W}_{22}\|^2+2\mathbf{W}_{22}^T(\tilde{\mathbf{z}}-\mathbf{z})+\|(\tilde{\mathbf{z}}-\mathbf{z})\|^2 \\ &= \|\mathbf{W}_{22}\|^2+2\mathbf{W}_{22}^T(\tilde{\mathbf{z}}-\mathbf{z})+d^2(\tilde{\mathbf{z}},\mathbf{z}). \end{aligned}$$

From the weak law of large numbers, for any $\epsilon > 0$ and sufficiently large B, we have

$$\Pr(\|\mathbf{W}_{22}\|^2 \in [nB(N_2-N_1-\epsilon/2),\quad nB(N_2-N_1+\epsilon/2)]) \ge 1-\epsilon/2$$

and

$$\Pr(2\mathbf{W}_{22}^T(\tilde{\mathbf{z}}-\mathbf{z}) \in [-nB\epsilon/2, nB\epsilon/2]) \ge 1-\epsilon/2.$$

Therefore, by the union bound, for any $\epsilon > 0$ and sufficiently large B,

$$\begin{aligned} 1-\epsilon &\le \Pr(d^2(\mathbf{Y},\mathbf{z}) \le nB(N_2-N_1+\epsilon)+d^2(\tilde{\mathbf{z}},\mathbf{z})) \\ &\le \Pr(d^2(\mathbf{Y},\mathbf{z}) \le nB(N_2-N_1+\epsilon) \\ &\quad + nB(\sqrt{2N_1a_n\ln 2}+3\sqrt{N_1\epsilon})^2) \\ &= \Pr(d(\mathbf{Y},\mathbf{z}) \le \sqrt{nB}\sqrt{(N_2+N_1(2a_n\ln 2-1)+\epsilon_1)}), \end{aligned} \quad (71)$$

where ϵ_1 is defined such that

$$(N_2-N_1+\epsilon)+(\sqrt{2N_1a_n\ln 2}+3\sqrt{N_1\epsilon})^2 = N_2+N_1(2a_n\ln 2-1)+\epsilon_1$$

and $\epsilon_1 \to 0$ as $\epsilon \to 0$. In light of (70) and (71), we have (72), as shown at the top of the next page, for any $(\mathbf{x},\mathbf{i}) \in S^{(B)}(X^n, I_n)$ and B sufficiently large.

Now along the similar lines as in the proof of Lemma 10, we can lower bound the conditional density $f(\mathbf{y}|\mathbf{i})$ for any $\mathbf{y}\in\Gamma_{\sqrt{nB}\sqrt{N_2+N_1(2a_n\ln 2-1)+\epsilon_1}}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))$, $(\mathbf{x},\mathbf{i}) \in S^{(B)}(X^n, I_n)$. In particular, consider a specific $\mathbf{z}\in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i})$ such that

$$d\ (\mathbf{y},\mathbf{z}) \le \sqrt{nB}\sqrt{N_2+N_1(2a_n\ln 2-1)+\epsilon_1}. \quad (73)$$

For any $\mathbf{x}\in S_\epsilon^{(B)}(X^n|\mathbf{z})$ where $\mathbf{z}$ is the one as described in (73), we have by (51) that

$$d(\mathbf{x},\mathbf{z}) \le \sqrt{nBN_1(1+2\epsilon)},$$

$$\begin{aligned}
&\Pr(\mathbf{Y} \in \Gamma_{\sqrt{nB}\sqrt{N_2+N_1(2a_n\ln 2-1)+\epsilon_1}}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))|\mathbf{X}=\mathbf{x})\\
&\quad = \Pr(\exists\ \mathbf{z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}) \text{ s.t. } d\ (\mathbf{Y},\mathbf{z}) \le \sqrt{nB}\sqrt{N_2+N_1(2a_n\ln 2-1)+\epsilon_1}\big|\mathbf{X}=\mathbf{x})\\
&\quad \ge \Pr(\exists\ \mathbf{z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}) \text{ s.t. } d\ (\tilde{\mathbf{Z}},\mathbf{z}) \le \sqrt{nB}(\sqrt{2N_1a_n\ln 2}+3\sqrt{N_1\epsilon})\big|\mathbf{X}=\mathbf{x})\\
&\qquad \times\Pr(\exists\ \mathbf{z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}) \text{ s.t. } d\ (\mathbf{Y},\mathbf{z}) \le \sqrt{nB}\sqrt{N_2+N_1(2a_n\ln 2-1)+\epsilon_1}\\
&\qquad\qquad \big|\mathbf{X}=\mathbf{x}, \exists\ \mathbf{z} \in S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}) \text{ s.t. } d\ (\tilde{\mathbf{Z}},\mathbf{z}) \le \sqrt{nB}(\sqrt{2N_1a_n\ln 2}+3\sqrt{N_1\epsilon}))\\
&\quad \ge (1-\epsilon)^2
\end{aligned} \tag{72}$$

and therefore by the triangle inequality,

$$d(\mathbf{x},\mathbf{y}) \le \sqrt{nBN_1(1+2\epsilon)} + \sqrt{nB}\sqrt{N_2+N_1(2a_n\ln 2-1)+\epsilon_1},$$

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

$$f(\mathbf{y}|\mathbf{x}) \ge 2^{-nB\left(\frac{1}{2}\log 2\pi eN_2+\frac{N_1}{N_2}a_n+\sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a_n\ln 2+1-\frac{N_1}{N_2}\right)}\log e+\epsilon_2\right)}$$

where $\epsilon_2 \to 0$ as $\epsilon \to 0$. Plugging this into (54) yields that

$$f(\mathbf{y}|\mathbf{i}) \ge 2^{-B\left(H(X^n|I_n)-H(X^n|Z^n)+\frac{n}{2}\log 2\pi eN_2\right)} \cdot 2^{-B\left(n\left(\frac{N_1}{N_2}a_n+\sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a_n\ln 2+1-\frac{N_1}{N_2}\right)}\log e\right)+n\epsilon_3\right)} \tag{74}$$

for any $\mathbf{y} \in \Gamma_{\sqrt{nB}\sqrt{N_2+N_1(2a_n\ln 2-1)+\epsilon_1}}(S_\epsilon^{(B)}(Z^n|\mathbf{x},\mathbf{i}))$, $(\mathbf{x},\mathbf{i}) \in S^{(B)}(X^n, I_n)$, where $\epsilon_3 \to 0$ as $\epsilon \to 0$. Finally, following the same procedure as in the proof of Lemma 10 to translate (74) to the upper bound on $h(Y^n|I_n)$, we have

$$h(Y^n|I_n) \le H(X^n|I_n) - H(X^n|Z^n) + \frac{n}{2}\log 2\pi eN_2 + n\left(\frac{N_1}{N_2}a_n + \sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a_n\ln 2+1-\frac{N_1}{N_2}\right)}\log e\right),$$

which combined with (65) immediately yields the new bound (68) on the difference $I(X^n; I_n) - I(Y^n; I_n)$. This concludes the proof of Lemma 15.

VII. Conclusion

We consider the Gaussian primitive relay channel, and establish a new upper bound on its capacity that is tighter than the cut-set bound. Combined with a tensorization argument [13], this result also implies that the current capacity approximations for Gaussian relay networks, which have linear gap to the cut-set bound in the number of nodes, are order-optimal and leads to a lower bound on the pre-constant.

The proof of our bound involves quantitively characterizing the tensions between the n-dimensional information measures involved in the problem. The main idea is to use measure concentration to study the geometric relations that are satisfied by typical realizations of the n-letter random variables associated with the problem, and then translate these geometric relations into new and surprising relations between the entropies of the corresponding n-letter random variables. In our forthcoming work [28], we further strengthen this geometric approach and develop a tighter upper bound on the capacity of the Gaussian primitive relay channel, which directly leads to a solution of a long-standing open question posed by Cover [9] in the Gaussian case.

Appendix A
Proof of Proposition 5

First rewrite our new bound in Corollary 4 as

$$\begin{cases} R \le \frac{1}{2}\log\left(1+\frac{P}{N_1}+\frac{P}{N_2}\right) & (75)\\ R \le \frac{1}{2}\log\left(1+\frac{P}{N_2}\right)+R_0-a^* & (76)\end{cases}$$

where a^* is the solution to the following equation:

$$\frac{1}{2}\log\left(1+\frac{P}{N_2}\right)+R_0-\frac{1}{2}\log\left(1+\max\left\{\frac{P}{N_1},\frac{P}{N_2}\right\}\right) = 2a^*+\sqrt{2a^*\ln 2}\log e. \tag{77}$$

Observe that the gap $\Delta(\frac{P}{N_1}, \frac{P}{N_2}, R_0)$ between our new bound and the cut-set bound is positive only if the channel parameters $(\frac{P}{N_1}, \frac{P}{N_2}, R_0)$ are such that between (75) and (76) of our bound, constraint (76) is active. This is because if in our bound constraint (75) is active, then for the cut-set bound also (3) is active and these two bounds become the same.

Thus to find the largest gap, one can without loss of generality assume constraint (76) is active for our bound. We now argue that the largest gap happens only when (4) is active for the cut-set bound. Suppose this is not true, i.e., when the largest gap happens constraint (3) instead of (4) is active. Then this implies that the R.H.S. of (3) is strictly less than that of (4) and thus one can reduce R_0 to further increase the gap, which contradicts with the largest gap assumption. Therefore, only when (76) and (4) are active, the gap attains the largest value that is given by the solution a^* to equation (77). The largest value that the L.H.S. of (77) can take while still maintaining (76) and (4) are active is 0.5, in which case the channel parameter $(\frac{P}{N_1}, \frac{P}{N_2}, R_0)$ has to be $(\infty, \infty, 0.5)$. Solving equation (77) with L.H.S. $=$ 0.5, we obtain $\Delta^* = \Delta(\infty,\infty,0.5) = 0.0535$.

Appendix B
Proof of Lemma 11

Given $A \subseteq \mathbb{R}^n$, let $B := \{b^n \in \mathbb{R}^n : \sqrt{N}b^n \in A\}$ and $V_i = \frac{U_i}{\sqrt{N}}, \forall i \in \{1,2,\ldots,n\}$. Then $V_1, V_2, \ldots, V_n$ are n i.i.d. standard Gaussian random variables with $V_i \sim \mathcal{N}(0,1)$, $\forall i \in \{1,2,\ldots,n\}$, and

$$\Pr(V^n \in B) = \Pr(\sqrt{N}V^n \in A) = \Pr(U^n \in A) \ge 2^{-na_n}.$$

We next invoke Gaussian measure concentration as stated in (1.6) of [32]: for any $B \subseteq \mathbb{R}^n$ and

$$r \geq \sqrt{-2 \ln \Pr(V^n \in B)},$$

we have

$$\Pr(V^n \in \Gamma_r(B)) \geq 1 - e^{-\frac{1}{2}\left(r - \sqrt{-2\ln \Pr(V^n \in B)}\right)^2}.$$

Thus, for any $t > 0$,

$$\begin{aligned}
&\Pr(V^n \in \Gamma_{\sqrt{n}(\sqrt{2a_n \ln 2} + \frac{t}{\sqrt{N}})}(B)) \\
&\quad \geq \Pr(V^n \in \Gamma_{\sqrt{-2\ln \Pr(V^n \in B)} + \sqrt{\frac{n}{N}}t}(B)) \\
&\quad \geq 1 - 2^{-\frac{nt^2}{2N}}.
\end{aligned}$$

Noting that

$$\begin{aligned}
&\Gamma_{\sqrt{n}(\sqrt{2Na_n \ln 2} + t)}(A) \\
&\qquad = \left\{\sqrt{N} b^n : b^n \in \Gamma_{\sqrt{n}(\sqrt{2a_n \ln 2} + \frac{r}{\sqrt{N}})}(B)\right\},
\end{aligned}$$

we have

$$\begin{aligned}
&\Pr(U^n \in \Gamma_{\sqrt{n}(\sqrt{2Na_n \ln 2} + t)}(A)) \\
&\quad = \Pr(\sqrt{N} V^n \in \Gamma_{\sqrt{n}(\sqrt{2Na_n \ln 2} + t)}(A)) \\
&\quad = \Pr(V^n \in \Gamma_{\sqrt{n}(\sqrt{2a_n \ln 2} + \frac{t}{\sqrt{N}})}(B)) \\
&\quad \geq 1 - 2^{-\frac{nt^2}{2N}}.
\end{aligned}$$

Appendix C
Proof of Lemma 12

Lemma 12-1) is a simple consequence of the law of large numbers. To prove Lemma 12-2)–4), we will repeatedly use the following lemma, which has been proved in [18].

Lemma 16: Let $A \subseteq C \times D$. For $x \in C$, use $A|_x$ to denote the set

$$A|_x = \{y \in D : (x, y) \in A\}.$$

If $\Pr(A) \geq 1 - \epsilon$, then $\Pr(B) \geq 1 - \sqrt{\epsilon}$, where

$$B := \{x \in C : \Pr(A|_x | x) \geq 1 - \sqrt{\epsilon}\}.$$

A. Proof of Lemma 12-2)

Consider B sufficiently large. Due to Lemma 16 and the fact that

$$\Pr((\mathbf{X}, \mathbf{Z}, \mathbf{I}) \in S_\epsilon^{(B)}(X, Z, I)) \geq 1 - \epsilon,$$

we have

$$\begin{aligned}
&\Pr\{(\mathbf{x}, \mathbf{z}) : \Pr((\mathbf{X}, \mathbf{z}, f(\mathbf{z})) \in S_\epsilon^{(B)}(X, Z, I) | \mathbf{Z} = \mathbf{z}) \\
&\qquad\qquad \geq 1 - \sqrt{\epsilon}\} \geq 1 - \sqrt{\epsilon},
\end{aligned}$$

i.e.,

$$\Pr\{(\mathbf{x}, \mathbf{z}) : \Pr(\mathbf{X} \in S_\epsilon^{(B)}(X|\mathbf{z}) | \mathbf{Z} = \mathbf{z}) \geq 1 - \sqrt{\epsilon}\} \geq 1 - \sqrt{\epsilon}.$$

Then by the definition of $S_\epsilon^{(B)}(X, Z)$,

$$\begin{aligned}
&\Pr((\mathbf{X}, \mathbf{Z}) \notin S_\epsilon^{(B)}(X, Z)) \\
&\quad \leq \Pr(\mathbf{X} \notin S_\epsilon^{(B)}(X|\mathbf{Z})) \\
&\qquad + \Pr\{(\mathbf{x}, \mathbf{z}) : \Pr(\mathbf{X} \in S_\epsilon^{(B)}(X|\mathbf{z}) | \mathbf{Z} = \mathbf{z}) < 1 - \sqrt{\epsilon}\} \\
&\quad \leq \epsilon + \sqrt{\epsilon} \\
&\quad \leq 2\sqrt{\epsilon},
\end{aligned}$$

and thus $\Pr(S_\epsilon^{(B)}(X, Z)) \geq 1 - 2\sqrt{\epsilon}$.

B. Proof of Lemma 12-3)

Consider B sufficiently large. We have

$$\begin{aligned}
&\Pr(\mathbf{Z} \notin S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{I})) \\
&\quad = \Pr(f(\mathbf{Z}) = \mathbf{I}, (\mathbf{X}, \mathbf{Z}) \notin S_\epsilon^{(B)}(X, Z)) \\
&\quad \leq 2\sqrt{\epsilon}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\Pr(\mathbf{Z} \notin S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{I})) \\
&= \sum_{(\mathbf{x}, \mathbf{i}) \in S_\epsilon^{(B)}(X, I)} \Pr(\mathbf{Z} \notin S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{I}) | \mathbf{X} = \mathbf{x}, \mathbf{I} = \mathbf{i}) p(\mathbf{x}, \mathbf{i}) \\
&\quad + \sum_{(\mathbf{x}, \mathbf{i}) \notin S_\epsilon^{(B)}(X, I)} \Pr(\mathbf{Z} \notin S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{I}) | \mathbf{X} = \mathbf{x}, \mathbf{I} = \mathbf{i}) p(\mathbf{x}, \mathbf{i}) \\
&\geq \sqrt[4]{\epsilon} \cdot \Pr((\mathbf{X}, \mathbf{I}) \notin S_\epsilon^{(B)}(X, I)).
\end{aligned}$$

Therefore, $\Pr((\mathbf{X}, \mathbf{I}) \notin S_\epsilon^{(B)}(X, I)) \leq 2\sqrt{\epsilon}/\sqrt[4]{\epsilon} = 2\sqrt[4]{\epsilon}$, and $\Pr((\mathbf{X}, \mathbf{I}) \in S_\epsilon^{(B)}(X, I)) \geq 1 - 2\sqrt[4]{\epsilon}$.

Consider any $(\mathbf{x}, \mathbf{i}) \in S_\epsilon^{(B)}(X, I)$. From the definition of $S_\epsilon^{(B)}(X, I)$,

$$\Pr(\mathbf{Z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i}) | \mathbf{X} = \mathbf{x}, \mathbf{I} = \mathbf{i}) \geq 1 - \sqrt[4]{\epsilon}.$$

Therefore, $S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i})$ must be nonempty, i.e., there exists at least one $\mathbf{z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i})$. Pick up any $\mathbf{z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i})$. By the definition of $S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i})$, we have i) $f(\mathbf{z}) = \mathbf{i}$ and ii) $(\mathbf{x}, \mathbf{z}) \in S_\epsilon^{(B)}(X, Z)$ such that $(\mathbf{x}, \mathbf{z}, \mathbf{i}) \in S_\epsilon^{(B)}(X, Z, I)$. Then, it follows from the definition of $S_\epsilon^{(B)}(X, Z, I)$ that

$$2^{-nB(a_n + \epsilon)} \leq p(\mathbf{i}|\mathbf{x}) \leq 2^{-nB(a_n - \epsilon)}.$$

Furthermore,

$$\begin{aligned}
&\Pr(\mathbf{Z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i}) | \mathbf{X} = \mathbf{x}) \\
&= \frac{\Pr(f(\mathbf{Z}) = \mathbf{i} | \mathbf{X} = \mathbf{x}) \Pr(\mathbf{Z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i}) | \mathbf{X} = \mathbf{x}, f(\mathbf{Z}) = \mathbf{i})}{\Pr(f(\mathbf{Z}) = \mathbf{i} | \mathbf{Z} \in S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i}), \mathbf{X} = \mathbf{x})} \\
&= p(\mathbf{i}|\mathbf{x}) \Pr(S_\epsilon^{(B)}(Z|\mathbf{x}, \mathbf{i}) | \mathbf{X} = \mathbf{x}, \mathbf{I} = \mathbf{i}) \\
&\geq 2^{-nB(a_n + \epsilon)}(1 - \sqrt[4]{\epsilon}) \\
&\geq 2^{-nB(a_n + 2\epsilon)}
\end{aligned}$$

for sufficiently large B.

C. Proof of Lemma 12-4)

For B sufficiently large,

$$\begin{aligned}
&\Pr(\mathbf{I} \in S_\epsilon^{(B)}(I)) \\
&\quad \geq \Pr\left\{\mathbf{i} : \Pr(\mathbf{X} \in S_\epsilon^{(B)}(X|\mathbf{i}) | \mathbf{I} = \mathbf{i}) \geq 1 - \sqrt{2\sqrt[4]{\epsilon}}\right\} \\
&\quad \geq 1 - \sqrt{2\sqrt[4]{\epsilon}} \\
&\quad \geq 1 - 2\sqrt[8]{\epsilon}.
\end{aligned}$$

$$a^* = \frac{(\frac{x_2}{x_1}+1)\ln(1+\frac{x_2}{1+x_1}) + 2\frac{x_2^2}{x_1^2}}{2(\frac{x_2}{x_1}+1)^2\ln 2} - \frac{\sqrt{((\frac{x_2}{x_1}+1)\ln(1+\frac{x_2}{1+x_1}) + 2\frac{x_2^2}{x_1^2})^2 - (\frac{x_2}{x_1}+1)^2[\ln^2(1+\frac{x_2}{1+x_1}) + \frac{x_2}{x_1}(\frac{x_2}{x_1}-1)]}}{2(\frac{x_2}{x_1}+1)^2\ln 2} \tag{81}$$

Appendix D

Consider the following upper bound jointly imposed by (8)–(9) and (67),

$$\begin{cases} R \le \frac{1}{2}\log\left(1+\frac{P}{N_1}+\frac{P}{N_2}\right) & (78)\\ R \le \frac{1}{2}\log\left(1+\frac{P}{N_2}\right) + R_0 - a^* & (79)\end{cases}$$

where a^* is the solution to the equation

$$\frac{1}{2}\log\left(1+\frac{P}{N_2}\right) + R_0 - \frac{1}{2}\log\left(1+\frac{P}{N_1}\right) = \left(\frac{N_1}{N_2}+1\right)a^* + \sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a^*\ln 2 + 1 - \frac{N_1}{N_2}\right)}\log e. \tag{80}$$

To show that the largest gap between our bound in Proposition 14 and the cut-set bound in (3)–(4) remains to be $\Delta^* = 0.0535$, it suffices to show that the above bound and the cut-set bound differ from each other at most 0.0535.

Similarly as in Appendix A, one can argue that the largest gap between the above bound and the cut-set bound happens only when (79) and (4) are active respectively, in which case the gap is given by the a^* satisfying (80). Note that for (4) to be active in the cut-set bound, one must have

$$\frac{1}{2}\log\left(1+\frac{P}{N_2}\right) + R_0 \le \frac{1}{2}\log\left(1+\frac{P}{N_1}+\frac{P}{N_2}\right).$$

Then to find the largest a^* we impose the following relation:

$$\frac{1}{2}\log\left(1+\frac{P}{N_1}+\frac{P}{N_2}\right) - \frac{1}{2}\log\left(1+\frac{P}{N_1}\right) = \left(\frac{N_1}{N_2}+1\right)a^* + \sqrt{\frac{N_1}{N_2}\left(\frac{N_1}{N_2}2a^*\ln 2 + 1 - \frac{N_1}{N_2}\right)}\log e.$$

Letting $x_i = \frac{P}{N_i}$ for $i \in \{1,2\}$ and solving the above equation, we have (81), as shown at the top of this page, where the maximum value $a^* = 0.0535$ is attained when $x_1 = x_2 = \infty$. This shows that the largest gap between our bound in Proposition 14 and the cut-set bound remains to be 0.0535.

Acknowledgment

The authors would like to thank the anonymous reviewers and the Associate Editor for many valuable comments that helped improve the presentation of this paper.

References

[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, no. 1, pp. 120–154, 1971.

[2] 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.

[3] 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.

[4] 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.

[5] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2000, p. 22.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. 45th Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Sep. 2007, p. 2007.

[11] A. Ozgur and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice-based QMF codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8275–8294, Dec. 2013.

[12] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence—Part I: Point-to-point Channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, Feb. 2011.

[13] T. Courtade and A. Özgür, "Approximate capacity of Gaussian relay networks: Is a sublinear gap to the cutset bound plausible?" in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 2251–2255.

[14] N. Fawaz and M. Médard, "A converse for the wideband relay channel with physically degraded broadcast," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2011, pp. 425–429.

[15] 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. 1282–1286.

[16] X. Wu and A. Özgür, "Improving on the cut-set bound via geometric analysis of typical sets," in *Proc. Int. Zurich Seminar Commun.*, pp. 130–134, 2016.

[17] X. Wu, A. Özgür, and L.-L. Xie, "Improving on the cut-set bound via geometric analysis of typical sets," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2254–2277, Apr. 2017.

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

[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] 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.

[21] A. Avestimehr, S. Diggavi, C. Tian, and D. Tse, "An approximation approach to network information theory," *Found. Trends Commun. Inf. Theory*, vol. 12, nos. 1–2, pp. 1–183, 2015.

[22] U. Niesen, B. Nazer, and P. Whiting, "Computation alignment: Capacity approximation without noise accumulation," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3811–3832, Jun. 2013.

[23] U. Niesen and S. N. Diggavi, "The approximate capacity of the Gaussian n-relay diamond network," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 845–859, Feb. 2013.

[24] B. Chern and A. Özgür, "Achieving the capacity of the N-relay Gaussian diamond network within logn bits," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2012, pp. 377–380.
[25] R. Kolte and A. Özgür, "Improved capacity approximations for Gaussian relay networks," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2013, pp. 1–5.
[26] S. Kannan, A. Raja, and P. Viswanath, "Approximately optimal wireless broadcasting," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7154–7167, Dec. 2012.
[27] S. H. Lim, K. T. Kim, and Y.-H. Kim, "Distributed decode-forward for multicast," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 636–640.
[28] X. Wu, L. Barnes, and A. Özgür. (Jan. 2017). "Cover's open problem: 'The capacity of the relay channel.'" [Online]. Available: https://arxiv.org/abs/1701.02043
[29] K. Marton, "Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration," *Ann. Probab.*, vol. 24, no. 2, pp. 857–866, 1996.
[30] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, no. 3, pp. 445–446, May 1986.
[31] K. Marton, "Distance-divergence inequalities," *IEEE Inf. Theory Soc. Newslett.*, vol. 64, no. 1, pp. 9–13, Mar. 2014.
[32] M. Talagrand, "Transportation cost for Gaussian and other product measures," *Geometr. Funct. Anal.*, vol. 6, no. 3, pp. 587–600, 1996.
[33] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3992–4002, Jul. 2016.
[34] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications and coding," *Found. Trends Commun. Inf. Theory*, vol. 10, nos. 1–2, pp. 1–250, Oct. 2014.
[35] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the Markov operator," *Ann. Probab.*, vol. 4, no. 6, pp. 925–939, 1976.
[36] 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./Oct. 2015, pp. 1135–1142.
[37] X. Wu and A. Özgür, "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.
[38] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[39] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

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.