Cut-Set Bound Is Loose for Gaussian Relay Networks

Xiugang Wu and Ayfer Özgür
Department of Electrical Engineering
Stanford University, Stanford, CA 94305
Email: x23wu@stanford.edu; aozgur@stanford.edu

***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 is based on typicality arguments and concentration of Gaussian measure. Combined with a simple 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 leads to a lower bound on the pre-constant.**

I. Introduction

The relay channel, and its Gaussian version in particular, 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, characterizing the capacity of the Gaussian relay channel has been one of the central problems in information theory over the past couple of decades.

The single relay channel has been introduced by van der Meulen in [1] and the seminal work of Cover and El Gamal in 1979 [2] has developed two basic achievability schemes for this setup, namely decode-and-forward and compress-and-forward, as well as an upper bound on its capacity, now known as the cut-set bound. Over the following 35 years, many new relaying strategies have been discovered such as amplify-and-forward, hash-and-forward, quantize-map-and-forward, compute-and-forward [3], [4], [5], [6], [7] etc., however the cut-set bound has remained as the only upper bound on the capacity of the Gaussian relay channel. To our knowledge, it is not even known if the cut-set bound is tight or not for this 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.[1] 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 [5]. See Figure 1. Here, the relay can be thought of as communicating to the destination over a Gaussian channel in a separate frequency band, or equivalently the destination can be thought of as equipped with two receive antennas, one directed to the source and one directed to the relay with no interference in between.[2] Our upper bound is tighter than the cut-set bound for this channel for all channel parameters. While this result is developed in the single-relay setting, it has implications also for Gaussian networks with multiple relays. In particular, combined with a simple tensorization argument recently proposed in [11], 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 fundamental. Indeed, the true 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.

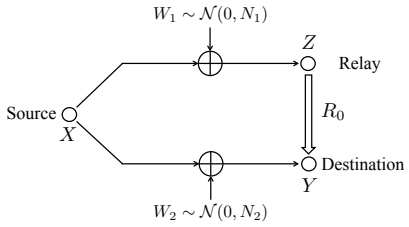

Fig. 1. Gaussian primitive relay channel.

Our upper bound builds on the approach we developed in our recent work [14] for bounding the capacity of the discrete memoryless primitive relay channel. Similar to earlier bounds on the capacity of the discrete memoryless primitive relay channel [12], [13], the bound we developed in [14] builds on the (generalized) blowing-up lemma, however unlike these earlier bounds does not critically rely on the finiteness of the alphabet size, which allows us to extend it to the Gaussian case in the current paper. Analogous to the results for the discrete memoryless case [12], [13], [14], a key ingredient of our upper bound for the Gaussian case is a Gaussian measure concentration result.

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

[1]For the sake of simplicity, in this paper we only focus on the symmetric case where the channels from the source to the relay and the destination have the same SNRs. Our arguments can be extended to the asymmetric case via channel simulation arguments.

[2]Note that due to network equivalence, the rate limited channel from the relay to the destination in Figure 1 can be equivalently thought of as a Gaussian channel of the same capacity [10].

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,

$$\mathcal{C}_{(n,R)} = \{x^n(m), m \in \{1, 2, \ldots, 2^{nR}\}\}$$

where

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(m) \le P, \ \forall m \in \{1, 2, \ldots, 2^{nR}\};$$

2) An encoding function at the relay Z,

$$f_n : \mathbb{R}^n \to \{1, 2, \ldots, 2^{nR_0}\};$$

3) A decoding function at the destination Y,

$$g_n : \mathbb{R}^n \times \{1, 2, \ldots, 2^{nR_0}\} \to \{1, 2, \ldots, 2^{nR}\}.$$

The average probability of error of the code is defined as

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

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

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

such that the average probability of error $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity of the primitive relay channel is the supremum of all achievable rates, denoted by $C(R_0)$.

B. The Cut-Set Bound

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

Proposition 2.1 (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

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

It can be easily shown that both $I(X; Y, Z)$ and $I(X; Y)$ in Proposition 2.1 are maximized when $X \sim \mathcal{N}(0, P)$, leading us to the following corollary.

Corollary 2.1: 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

To simplify the exposition, in this paper we only concentrate on the symmetric case of the Gaussian primitive relay channel, that is, when $N_1 = N_2 =: N$. Our results can be extended to the asymmetric case by using channel simulation arguments. We defer this extension to the longer version of the paper. The following theorem states the main result of this paper.

Theorem 3.1: For the symmetric 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

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

As in the case of the cut-set bound, since both $I(X; Y, Z)$ and $I(X; Y)$ in Theorem 3.1 are maximized when $X \sim \mathcal{N}(0, P)$, we have the following corollary.

Corollary 3.1: For the symmetric Gaussian primitive relay channel, if a rate R is achievable, then there exists some $a \in [0, R_0]$ such that

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

Note that in the symmetric case, by Corollary 2.1, the cut-set bound says that if a rate R is achievable, then

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

Clearly the bound on R in Corollary 3.1 is tighter than the cut-set bound since (9) will only reduce to (12) if $a = 0$. However, if $a = 0$ then (10) will constrain the rate R by the capacity of the source-destination link. The constraint on R, jointly imposed by (9) and (10) can be found by equating them to yield

$$R_0 = 2a^* + \sqrt{2a^* \ln 2}\log e. \quad (13)$$

Corollary 3.1 can be restated in terms of a^* as follows: if a rate R is achievable, then

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

Note that both the cut-set bound and our new bound depend on the channel parameters through $\frac{P}{N}$ and R_0. It is interesting to evaluate the largest gap between these two bounds over all parameter values for the symmetric Gaussian primitive relay channel. For this, it can be shown that when $\frac{P}{N} \to \infty$ and $R_0 = 0.5$, the gap takes its largest value and is given by the solution of equation (13), which is $a^* = 0.0535$. We formally summarize this observation in the following proposition.

Proposition 3.1: Let $\Delta\left(\frac{P}{N}, R_0\right)$ denote the gap between the two bounds and Δ^* its largest possible value over all symmetric Gaussian primitive relay channels, i.e.,

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

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

A. Gaussian Relay Networks

While the setup we consider in this paper can be regarded as a special case of a Gaussian relay network, the upper bound we develop for this special case can be used to infer how tightly the capacity of general Gaussian relay networks can be approximated by the cut-set bound. Initiated by the work of Avestimehr, Diggavi and Tse [6], there has been significant recent interest [8], [9] 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. The gap in these approximation results is linear in the number of nodes in the network but independent of the channel SNRs and network topology. In particular, the best currently known approximation result [19] has a gap of $0.5N$ where N is the total number of nodes. While some recent works [15], [16], [17], [18] demonstrate sublinear in the number of nodes (or in the total number of antennas in the case of multiple antenna nodes) gap to the cut-set bound for specific topologies, a recent tensorization argument proposed in [11] shows that the gap between the capacity and the cut-set bound can be bounded by a sublinear function of the number of nodes, independent of network topology and channel configurations, if, and only if, capacity is equal to the cut-set bound for *all* Gaussian relay networks. Moreover, Theorem 3 of [11] implies that an explicit gap to the cut-set bound for any specific network with specific channel parameters and topology would imply a lower bound on the preconstant in these approximation results. In particular, the gap 0.0535 in Proposition 3.1 for the Gaussian primitive relay channel (which can be thought of as a Gaussian network with two receive antennas at the destination, so four antennas in total) implies that the capacity of Gaussian relay networks can not be approximated by the cutset bound, independent of the topology and channel coefficients, with a gap that is smaller than $(0.0535/4)N \approx 0.01N$.

IV. Proof of Theorem 3.1

In this section, we prove bounds (5)–(7) sequentially with the focus on showing (7).

Suppose a rate R is achievable. Then there exists a sequence of codes

$$\{(\mathcal{C}_{(n,R)}, f_n, g_n)\}_{n=1}^{\infty} \tag{14}$$

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

For this sequence of codes, we have

$$\begin{aligned}
nR &= H(M) \\
&= I(M; Y^n, Z^n) + H(M|Y^n, Z^n) \\
&\le I(X^n; Y^n, Z^n) + H(M|Y^n, f_n(Z^n)) \\
&\le I(X^n; Y^n, Z^n) + n\mu \qquad (15) \\
&= h(Y^n, Z^n) - h(Y^n, Z^n|X^n) + n\mu \\
&= \sum_i^n [h(Y_i, Z_i|Y^{i-1}, Z^{i-1}) - h(Y_i, Z_i|X_i)] + n\mu \\
&\le \sum_i^n [h(Y_i, Z_i) - h(Y_i, Z_i|X_i)] + n\mu \\
&= \sum_i^n I(X_i; Y_i, Z_i) + n\mu \\
&= n(I(X_Q; Y_Q, Z_Q|Q) + \mu) \qquad (16) \\
&= n(h(Y_Q, Z_Q|Q) - h(Y_Q, Z_Q|Q, X_Q) + \mu) \\
&\le n(h(Y_Q, Z_Q) - h(Y_Q, Z_Q|X_Q) + \mu) \\
&= n(I(X_Q; Y_Q, Z_Q) + \mu)
\end{aligned}$$

i.e.,

$$R \le I(X_Q; Y_Q, Z_Q) + \mu \tag{17}$$

for any $\mu > 0$ and sufficiently large n, where (15) follows from Fano's inequality, (16) follows by defining the time sharing random variable Q to be uniformly distributed over $[1 : n]$, and

$$E[X_Q^2] = \frac{1}{n}\sum_i^n E[X_i^2] = \frac{1}{n} E\left[\sum_i^n X_i^2\right] \le P. \tag{18}$$

Moreover, for any $\mu > 0$ and sufficiently large n,

$$\begin{aligned}
nR &= H(M) \\
&= I(M; Y^n, f_n(Z^n)) + H(M|Y^n, f_n(Z^n)) \\
&\le I(X^n; Y^n, f_n(Z^n)) + n\mu \qquad (19) \\
&= I(X^n; Y^n) + I(X^n; f_n(Z^n)|Y^n) + n\mu \\
&= I(X^n; Y^n) + H(f_n(Z^n)|Y^n) - H(f_n(Z^n)|X^n) + n\mu \\
&\le n(I(X_Q; Y_Q) + R_0 - a_n + \mu), \qquad (20)
\end{aligned}$$

i.e.,

$$R \le I(X_Q; Y_Q) + R_0 - a_n + \mu, \tag{21}$$

where $a_n = \frac{1}{n} H(I_n|X^n)$ with $I_n := f_n(Z^n)$, and a_n satisfies

$$0 \le a_n \le R_0. \tag{22}$$

So far we have made only standard information theoretic arguments and in particular recovered the cut-set bound; note that the fact that $a_n \ge 0$ together with (17), (21) and (18)

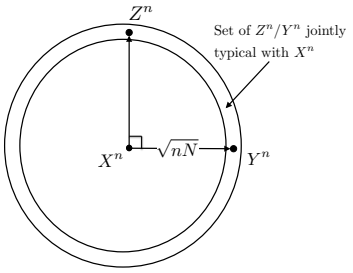

Fig. 2. Jointly typical set with X^n.

yields the cut-set bound given in Proposition 2.1. However, instead of simply lower bounding a_n by 0 in (21), in the sequel we will prove a third inequality involving a_n, which will force 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, or equivalently all Z^n sequences jointly typical with X^n are mapped to the same I_n. See Figure 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 the transmitted codeword X^n can be decoded based solely on Y^n. This will force the rate to be smaller than $I(X_Q;Y_Q)$. 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.

From (19),

$$\begin{aligned} nR &\le I(X^n;Y^n,I_n)+n\mu \\ &= I(X^n;I_n)+I(X^n;Y^n|I_n)+n\mu \\ &= H(X^n)-nb_n+h(Y^n|I_n)-\frac{n}{2}\log 2\pi eN+n\mu, \end{aligned} \tag{23}$$

where we define $b_n := \frac{1}{n}H(X^n|I_n)$ and use the fact that $h(Y^n|X^n,I_n)=h(Y^n|X^n)=\frac{n}{2}\log 2\pi eN$. In Section IV-A, we prove the following key upper bound on the conditional entropy of Y^n given the relay's transmission I_n,

$$h(Y^n|I_n) \le n(b_n - c_n + \frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n \ln 2}\log e) \tag{24}$$

where $c_n := \frac{1}{n}H(X^n|Z^n)$. Combined with (23), this yields

$$\begin{aligned} nR &\le I(X^n;Z^n)+n(a_n+\sqrt{2a_n\ln 2}\log e)+n\mu, \\ &= I(X^n;Y^n)+n(a_n+\sqrt{2a_n\ln 2}\log e)+n\mu, \\ &\le n(I(X_Q;Y_Q)+a_n+\sqrt{2a_n\ln 2}\log e+\mu) \end{aligned}$$

where the equality follows from the fact that Z^n and Y^n are statistically equivalent given X^n. Equivalently,

$$R \le I(X_Q;Y_Q)+a_n+\sqrt{2a_n\ln 2}\log e+\mu. \tag{25}$$

Combining (17), (21) and (25) , we have that if a rate R is achievable, then for any $\mu>0$ and sufficiently large n,

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

where $E[X_Q^2]\le P$ and $a_n\in[0,R_0]$. Since μ can be arbitrarily small, this proves Theorem 3.1.

A. *Proving Inequality* (24)

The remaining step then is to prove the relation in (24). To prove this inequality we will look at B-length i.i.d. sequences of the random vectors X^n, Y^n, Z^n, and I_n, and derive some typicality properties for these sequences which hold with high probability when B is large.[3]

Specifically, consider the following B-length i.i.d. sequence

$$\{(X^n(b),Y^n(b),Z^n(b),I_n(b))\}_{b=1}^B, \tag{26}$$

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 $[X^n(1),X^n(2),\ldots,X^n(B)]$ as $\mathbf{X}$ and similarly define $\mathbf{Y},\mathbf{Z}$ and $\mathbf{I}$; note here we have $\mathbf{I} = [f_n(Z^n(1)),f_n(Z^n(2)),\ldots,f_n(Z^n(B))] =: f(\mathbf{Z})$.

We now present a key lemma in our proof, which gives a lower bound on the conditional probability density $f(\mathbf{y}|\mathbf{i})$ for a set of "typical" $(\mathbf{y},\mathbf{i})$ pairs. The formal proof of this lemma is delegated to Appendix B. In the next subsection we provide a proof sketch for this lemma that summarizes the main ideas.

Lemma 4.1: 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 $\mathbf{y}$ such that

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

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

$$f(\mathbf{y}|\mathbf{i}) \ge 2^{-nB(b_n-c_n+\frac{1}{2}\log 2\pi eN+a_n+\sqrt{2a_n\ln 2}\log e+\delta_1)},$$

where $\delta_1\to 0$ as $\delta\to 0$.

Equipped with this lemma, it is not difficult to prove (24). For this, first consider $h(\mathbf{Y}|\mathbf{i})$ for any $\mathbf{i}\in\mathcal{I}$. We have

$$\begin{aligned} h(\mathbf{Y}|\mathbf{i}) &\le h(\mathbf{Y}|\mathbf{i})+1-I(\mathbf{Y};\mathbb{I}(\mathbf{Y}\in\mathcal{Y}_\mathbf{i})|\mathbf{i}) && (27)\\ &= 1+h(\mathbf{Y}|\mathbb{I}(\mathbf{Y}\in\mathcal{Y}_\mathbf{i}),\mathbf{i}) \\ &= 1+\Pr(\mathbf{Y}\in\mathcal{Y}_\mathbf{i}|\mathbf{i})h(\mathbf{Y}|\mathbf{i},\mathbf{Y}\in\mathcal{Y}_\mathbf{i}) \\ &\quad +\Pr(\mathbf{Y}\notin\mathcal{Y}_\mathbf{i}|\mathbf{i})h(\mathbf{Y}|\mathbf{i},\mathbf{Y}\notin\mathcal{Y}_\mathbf{i}), && (28) \end{aligned}$$

where $\mathbb{I}(A)$ is the indicator function defined as 1 if A holds and 0 otherwise, and (27) follows since

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

[3]Note that X^n here 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.

To bound $h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i})$, we have by Lemma 4.1 that,

$$\begin{aligned}
&h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) \\
&= -\int_{\mathbf{y}\in\mathcal{Y}_\mathbf{i}} f(\mathbf{y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) \log f(\mathbf{y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) d\mathbf{y} \\
&\leq -\int_{\mathbf{y}\in\mathcal{Y}_\mathbf{i}} f(\mathbf{y}|\mathbf{i}, \mathbf{Y} \in \mathcal{Y}_\mathbf{i}) \log f(\mathbf{y}|\mathbf{i}) d\mathbf{y} \\
&\leq nB(b_n - c_n + \frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n \ln 2}\log e + \delta_1).
\end{aligned} \tag{29}$$

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

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

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

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

and

$$\begin{aligned}
&\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{i}) h(\mathbf{Y}|\mathbf{i}, \mathbf{Y} \notin \mathcal{Y}_\mathbf{i}) \\
&\leq \frac{nB}{2}\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{i}) \log 2\pi e \frac{P+N}{\Pr(\mathbf{Y} \notin \mathcal{Y}_\mathbf{i}|\mathbf{i})} \\
&\leq nB\delta_2,
\end{aligned} \tag{30}$$

for some $\delta_2 \to 0$ as $\delta \to 0$.

Plugging (29) and (30) into (28), we have for any $\mathbf{i} \in \mathcal{I}$,

$$\begin{aligned}
&h(\mathbf{Y}|\mathbf{i}) \\
&\leq 1 + \Pr(\mathbf{Y} \in \mathcal{Y}_\mathbf{i}|\mathbf{i}) nB[b_n - c_n + \frac{1}{2}\log 2\pi eN \\
&\qquad + a_n + \sqrt{2a_n \ln 2}\log e + \delta_1] + nB\delta_2 \\
&= nB(b_n - c_n + \frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n \ln 2}\log e + \delta_3)
\end{aligned}$$

where $\delta_3 \to 0$ as $\delta \to 0$ and $B \to \infty$. Therefore, for sufficiently large B,

$$\begin{aligned}
&h(\mathbf{Y}|\mathbf{I}) \\
&= \sum_i p(\mathbf{i}) h(\mathbf{Y}|\mathbf{i}) \\
&= \sum_{\mathbf{i}\in\mathcal{I}} p(\mathbf{i}) h(\mathbf{Y}|\mathbf{i}) + \sum_{\mathbf{i}\notin\mathcal{I}} p(\mathbf{i}) h(\mathbf{Y}|\mathbf{i}) \\
&\leq \sum_{\mathbf{i}\in\mathcal{I}} p(\mathbf{i}) nB(b_n - c_n + \frac{1}{2}\log 2\pi eN + a_n \\
&\quad + \sqrt{2a_n \ln 2}\log e + \delta_3) + \sum_{\mathbf{i}\notin\mathcal{I}} p(\mathbf{i}) \frac{nB}{2} \log 2\pi e(P+N) \\
&= nB(b_n - c_n + \frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n \ln 2}\log e + \delta_4)
\end{aligned} \tag{31}$$

where $\delta_4 \to 0$ as $\delta \to 0$ and $B \to \infty$. Finally observing that

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

and taking $B \to \infty$ complete the proof of inequality (24).

B. Proof Idea for Lemma 4.1

We now provide a proof sketch for Lemma 4.1. The formal proof can be found in Appendix B.

By the law of large numbers, if $H(I_n|X^n) = na_n$, then given a typical $(\mathbf{x}, \mathbf{i})$ pair, it can be shown that

$$\Pr(\mathbf{Z} \in \mathcal{Z}_{(\mathbf{x},\mathbf{i})}|\mathbf{x}) \doteq 2^{-B\, I(Z^n; I_n|X^n)} = 2^{-nBa_n},$$

where $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ can be roughly viewed as the set of $\mathbf{z}$ that are jointly typical with $(\mathbf{x}, \mathbf{i})$.

Now we apply the following lemma, whose proof relies on a Gaussian measure concentration result and is included in Appendix A.

Lemma 4.2: Let $U_1, U_2, \ldots, U_n$ be n i.i.d. Gaussian random variables with $U_i \sim \mathcal{N}(0, N), \forall i \in \{1, 2, \ldots, n\}$. Then, for any $A \subseteq \mathbb{R}^n$ with $\Pr(U^n \in A) \geq 2^{-na_n}$,

$$\Pr(U^n \in \Gamma_{\sqrt{n}(\sqrt{2Na_n \ln 2}+r)}(A)) \geq 1 - 2^{-\frac{nr^2}{2N}}, \forall r > 0,$$

where

$$\begin{aligned}
&\Gamma_{\sqrt{n}(\sqrt{2Na_n \ln 2}+r)}(A) \\
&:= \{\underline{\omega} \in \mathbb{R}^n : \exists\ \underline{\omega}' \in A \text{ s.t. } d(\underline{\omega}, \underline{\omega}') \leq \sqrt{n}(\sqrt{2Na_n \ln 2} + r)\},
\end{aligned}$$

with $d(\underline{\omega}, \underline{\omega}') := ||\underline{\omega} - \underline{\omega}'||$ denoting the Euclidean distance between $\underline{\omega}$ and $\underline{\omega}'$.

With Lemma 4.2, it can be shown that if one blows up $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ with a radius $\sqrt{nB}\sqrt{2Na_n \ln 2}$, the resultant set, denoted by $\Gamma_{\sqrt{nB}\sqrt{2Na_n \ln 2}}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})$, has probability nearly 1, i.e.,

$$\Pr(\mathbf{Z} \in \Gamma_{\sqrt{nB}\sqrt{2Na_n \ln 2}}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})|\mathbf{x}) \approx 1. \tag{32}$$

Due to the symmetry of the channel, (32) still holds with $\mathbf{Z}$ replaced by $\mathbf{Y}$.

Now given a typical $(\mathbf{x}, \mathbf{i})$ pair, we lower bound the conditional density $f(\mathbf{y}|\mathbf{i})$ for all $\mathbf{y} \in \Gamma_{\sqrt{nB}\sqrt{2Na_n \ln 2}}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})$. Given such $\mathbf{y}$, there exists some $\mathbf{z} \in \mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ such that $d(\mathbf{y}, \mathbf{z}) \leq \sqrt{nB}\sqrt{2Na_n \ln 2}$. Consider the set of all $\mathbf{x}$ that are jointly typical with this $\mathbf{z}$. It can be shown that the $\mathbf{x}$'s that are jointly typical with a given $\mathbf{z} \in \mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ are such that

$$d(\mathbf{x}, \mathbf{z}) \leq \sqrt{nBN},$$

and

$$p(\mathbf{x}|\mathbf{i}) \doteq 2^{-nBb_n}.$$

Therefore for each $\mathbf{x}$ in this set

$$\begin{aligned}
d(\mathbf{x}, \mathbf{y}) &\leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}) \\
&\leq \sqrt{nB}(\sqrt{N} + \sqrt{2Na_n \ln 2}),
\end{aligned}$$

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

$$f(\mathbf{y}|\mathbf{x}) \dot{\geq} 2^{-nB\left(\frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n \ln 2}\log e\right)},$$

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^{nBc_n}. Combining this with the above, we have

$$\begin{aligned}
f(\mathbf{y}|\mathbf{i}) &= \sum_{\mathbf{x}} f(\mathbf{y}|\mathbf{x}) p(\mathbf{x}|\mathbf{i}) \\
&\dot{\geq} 2^{nBc_n} 2^{-nBb_n} 2^{-nB\left(\frac{1}{2}\log 2\pi eN + a_n + \sqrt{2a_n \ln 2}\log e\right)}.
\end{aligned}$$

Using the fact $(\mathbf{x}, \mathbf{i})$ are jointly typical with high probability and given a typical $(\mathbf{x}, \mathbf{i})$ the above lower bound holds for all $\mathbf{y}$ with high probability completes the proof of Lemma 4.1. A rigorous proof is given in the sequel.

APPENDIX A
PROOF OF LEMMA 4.2

Given $A \subseteq \mathbb{R}^n$, let $B := \{\underline{\omega} \in \mathbb{R}^n : \sqrt{N}\underline{\omega} \in A\}$ 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

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

We next invoke Gaussian measure concentration as stated in (1.6) of [20]: for any $B \subseteq \mathbb{R}^n$ and

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

we have

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

Thus, for any $r > 0$,

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

Noting that

$$\Gamma_{\sqrt{n}(\sqrt{2Na_n \ln 2} + r)}(A) = \left\{\sqrt{N}\underline{\omega} : \underline{\omega} \in \Gamma_{\sqrt{n}(\sqrt{2a_n \ln 2} + \frac{r}{\sqrt{N}})}(B)\right\},$$

we have

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

APPENDIX B
PROOF OF LEMMA 4.1

A. Definitions of High Probability Sets

By considering the B-length i.i.d. extensions of the n-letter random variables involved, law of large numbers allows us to concentrate on a series of "high probability" sets defined in the following.[4]

Definition of $\tilde{S}(X^n, Z^n)$

Lemma B.1: Assume $H(I_n|X^n) = na_n, H(X^n|I_n) = nb_n, H(X^n|Z^n) = nc_n$ for the n-channel use code. Given any $\epsilon > 0$ and sufficiently large B, we have

$$\Pr((\mathbf{X}, \mathbf{Z}) \in \tilde{S}(X^n, Z^n)) \geq 1 - \epsilon$$

where

$$\begin{aligned}
\tilde{S}(X^n, Z^n) := \Big\{(\mathbf{x}, \mathbf{z}) : d(\mathbf{x}, \mathbf{z}) \in [\sqrt{nB}(\sqrt{N} - \epsilon), \sqrt{nB}(\sqrt{N} + \epsilon)] \\
2^{-nB(a_n + \epsilon)} \leq p(f(\mathbf{z})|\mathbf{x}) \leq 2^{-nB(a_n - \epsilon)} \\
2^{-nB(b_n + \epsilon)} \leq p(\mathbf{x}|f(\mathbf{z})) \leq 2^{-nB(b_n - \epsilon)} \\
2^{-nB(c_n + \epsilon)} \leq p(\mathbf{x}|\mathbf{z}) \leq 2^{-nB(c_n - \epsilon)}\Big\}
\end{aligned}$$

The lemma is a simple consequence of the law of large numbers.

Definition of $S(X^n, Z^n)$

To define $S(X^n, Z^n)$, we first consider the following lemma, which has been proved in [12].

Lemma B.2: 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

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

Now, define

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

Clearly $S(X^n, Z^n)$ is a subset of $\tilde{S}(X^n, Z^n)$. The following lemma says that it is also a high probability set.

Lemma B.3: $\Pr(S(X^n, Z^n)) \geq 1 - 2\sqrt{\epsilon}$ for B sufficiently large.

Proof: Consider B sufficiently large. Due to Lemma B.2 and the fact that $\Pr(\tilde{S}(X^n, Z^n)) \geq 1 - \epsilon$, we have

$$\Pr\{(\mathbf{x}, \mathbf{z}) : \Pr(\tilde{S}(X^n, Z^n)|_{\mathbf{z}}|\mathbf{z}) \geq 1 - \sqrt{\epsilon}\} \geq 1 - \sqrt{\epsilon}.$$

Then by the definition of $S(X^n, Z^n)$,

$$\begin{aligned}
&\Pr(S^c(X^n, Z^n)) \\
&\leq \Pr(\tilde{S}^c(X^n, Z^n)) + \Pr\{(\mathbf{x}, \mathbf{z}) : \Pr(\tilde{S}(X^n, Z^n)|_{\mathbf{z}}|\mathbf{z}) < 1 - \sqrt{\epsilon}\} \\
&\leq \epsilon + \sqrt{\epsilon} \\
&\leq 2\sqrt{\epsilon},
\end{aligned}$$

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

Definitions of $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ and $S(X^n, I_n)$

Define

$$\mathcal{Z}_{(\mathbf{x},\mathbf{i})} = \{\mathbf{z} : f(\mathbf{z}) = \mathbf{i}, (\mathbf{x}, \mathbf{z}) \in S(X^n, Z^n)\}$$

and

$$S(X^n, I_n) = \{(\mathbf{x}, \mathbf{i}) : \Pr(\mathcal{Z}_{(\mathbf{x},\mathbf{i})}|\mathbf{x}, \mathbf{i}) \geq 1 - \sqrt[4]{\epsilon}\}.$$

Lemma B.4: $\Pr(S(X^n, I_n)) \geq 1 - 2\sqrt[4]{\epsilon}$ for B sufficiently large.

Proof: For B sufficiently large, consider $\Pr(\mathbf{Z} \notin \mathcal{Z}_{(\mathbf{X},\mathbf{I})})$. We have

$$\Pr(\mathbf{Z} \notin \mathcal{Z}_{(\mathbf{X},\mathbf{I})}) = \Pr(f(\mathbf{Z}) = \mathbf{I}, (\mathbf{X}, \mathbf{Z}) \notin S(X^n, Z^n)) \leq 2\sqrt{\epsilon}.$$

[4]The high probability sets defined here are analogous to strongly typical sets that are widely used in information theory. In the Gaussian case, the notion of strong typicality doesn't apply and thus we need to develop our own customized high probability sets. In the discrete memoryless case [14], one can simply resort to strong typicality.

On the other hand,

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

Therefore, $\Pr(S^c(X^n,I_n)) \leq 2\sqrt{\epsilon}/\sqrt[4]{\epsilon} = 2\sqrt[4]{\epsilon}$, and $\Pr(S(X^n,I_n))\geq 1-2\sqrt[4]{\epsilon}$. ∎

Lemma B.5: For any $(\mathbf{x},\mathbf{i})\in S(X^n,I_n)$, we have

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

and for sufficiently large B,

$$\Pr(\mathcal{Z}_{(\mathbf{x},\mathbf{i})}|\mathbf{x})\geq 2^{-nB(a_n+2\epsilon)}.$$

Proof: Consider any $(\mathbf{x},\mathbf{i})\in S(X^n,I_n)$. From the definition of $S(X^n,I_n)$, $\Pr(\mathcal{Z}_{(\mathbf{x},\mathbf{i})}|\mathbf{x},\mathbf{i})\geq 1-\sqrt[4]{\epsilon}$. Therefore, $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ must be nonempty, i.e., there exists at least one $\mathbf{z}\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$.

Consider any $\mathbf{z}\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$. By the definition of $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$, we have $f(\mathbf{z})=\mathbf{i}$ and $(\mathbf{x},\mathbf{z})\in S(X^n,Z^n)\subseteq\tilde{S}(X^n,Z^n)$. Then, it follows from the definition of $\tilde{S}(X^n,Z^n)$ that

$$2^{-nB(a_n+\epsilon)}\leq p(f(\mathbf{z})|\mathbf{x})\leq 2^{-nB(a_n-\epsilon)},$$

i.e.,

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

Furthermore,

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

for sufficiently large B. This finishes the proof of the lemma. ∎

B. Blowing Up $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$

Lemma B.6: For any $(\mathbf{x},\mathbf{i})\in S(X^n,I_n)$, consider the following blown-up set of $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$:

$$\begin{aligned}\Gamma_{\sqrt{nB}(\sqrt{2Na_n}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})}) = \{\underline{\omega}\in\mathbb{R}^{nB}:\exists\ \underline{\omega}'\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}\\ \text{s.t. }\ d(\underline{\omega},\underline{\omega}')\leq\sqrt{nB}(\sqrt{2Na_n}+3\sqrt{N\epsilon})\}.\end{aligned}$$

We have

1) $\Pr(\mathbf{Y}\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})|\mathbf{x})\geq 1-\epsilon$ for sufficiently large B;
2) For any $\mathbf{y}\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})$,

$$f(\mathbf{y}|\mathbf{i})\geq 2^{-nB(b_n-c_n+\frac{1}{2}\log 2\pi eN+(a_n+\sqrt{2a_n})\log e+\epsilon')}$$

where $\epsilon'\to 0$ as $\epsilon\to 0$ and $B\to\infty$.

Proof: From Lemma B.5, for any $(\mathbf{x},\mathbf{i})\in S(X^n,I_n)$ and sufficiently large B,

$$\Pr(\mathbf{Z}\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}|\mathbf{x})\geq 2^{-nB(a_n+2\epsilon)},$$

i.e.,

$$\begin{aligned}\Pr(\mathbf{x}+\mathbf{W}_1\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}|\mathbf{x}) &= \Pr(\mathbf{W}_1\in\{\underline{\omega}-\mathbf{x}:\underline{\omega}\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}\})\\ &\geq 2^{-nB(a_n+2\epsilon)}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\Pr(\mathbf{Y}\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})|\mathbf{x})\\ &=\Pr(\mathbf{x}+\mathbf{W}_2\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})|\mathbf{x})\\ &=\Pr(\mathbf{W}_2\in\{\underline{\omega}-\mathbf{x}:\underline{\omega}\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})\})\\ &=\Pr(\mathbf{W}_1\in\{\underline{\omega}-\mathbf{x}:\underline{\omega}\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})\})\\ &=\Pr(\mathbf{W}_1\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(\{\underline{\omega}-\mathbf{x}:\underline{\omega}\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}\}))\\ &\geq\Pr(\mathbf{W}_1\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2+4N\epsilon\ln 2}+\sqrt{N\epsilon})}(\{\underline{\omega}-\mathbf{x}:\underline{\omega}\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}\}))\\ &\geq 1-2^{-\frac{nB\epsilon}{2}} \qquad (33)\\ &\geq 1-\epsilon\end{aligned}$$

for sufficiently large B, where (33) follows from Lemma 4.2.

To prove Part 2), consider any $\mathbf{y}\in\Gamma_{\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})$. We can find one $\mathbf{z}\in\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ such that $d(\mathbf{y},\mathbf{z})\leq\sqrt{nB}(\sqrt{2Na_n\ln 2}+3\sqrt{N\epsilon})$, and for this $\mathbf{z}$, we have from the definition of $\mathcal{Z}_{(\mathbf{x},\mathbf{i})}$ that: i) $f(\mathbf{z})=\mathbf{i}$ and ii) $\Pr(\tilde{S}(X^n,Z^n)|_{\mathbf{z}}|\mathbf{z})\geq 1-\sqrt{\epsilon}$, where

$$\begin{aligned}\tilde{S}(X^n,Z^n)|_{\mathbf{z}} = \Big\{\mathbf{x}:\ & d(\mathbf{x},\mathbf{z})\in[\sqrt{nB}(\sqrt{N}-\epsilon),\sqrt{nB}(\sqrt{N}+\epsilon)]\\ & 2^{-nB(a_n+\epsilon)}\leq p(f(\mathbf{z})|\mathbf{x})\leq 2^{-nB(a_n-\epsilon)}\\ & 2^{-nB(b_n+\epsilon)}\leq p(\mathbf{x}|f(\mathbf{z}))\leq 2^{-nB(b_n-\epsilon)}\\ & 2^{-nB(c_n+\epsilon)}\leq p(\mathbf{x}|\mathbf{z})\leq 2^{-nB(c_n-\epsilon)}\Big\}.\end{aligned}$$

The size of $\tilde{S}(X^n,Z^n)|_{\mathbf{z}}$ can be lower bounded by considering the following

$$\begin{aligned}1-\sqrt{\epsilon}&\leq\Pr(\tilde{S}(X^n,Z^n)|_{\mathbf{z}}|\mathbf{z})\\ &=\sum_{\mathbf{x}\in\tilde{S}(X^n,Z^n)|_{\mathbf{z}}}p(\mathbf{x}|\mathbf{z})\\ &\leq 2^{-nB(c_n-\epsilon)}\big|\tilde{S}(X^n,Z^n)|_{\mathbf{z}}\big|,\end{aligned}$$

i.e.,

$$\big|\tilde{S}(X^n,Z^n)|_{\mathbf{z}}\big|\geq(1-\sqrt{\epsilon})2^{nB(c_n-\epsilon)}.$$

Then,

$$\begin{aligned}f(\mathbf{y}|\mathbf{i}) &= \sum_{\mathbf{x}}f(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i})\\ &\geq\sum_{\mathbf{x}\in\tilde{S}(X^n,Z^n)|_{\mathbf{z}}}f(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i})\\ &\geq 2^{-nB(b_n+\epsilon)}\sum_{\mathbf{x}\in\tilde{S}(X^n,Z^n)|_{\mathbf{z}}}f(\mathbf{y}|\mathbf{x})\\ &\geq 2^{-nB(b_n+\epsilon)}\big|\tilde{S}(X^n,Z^n)|_{\mathbf{z}}\big|\min_{\mathbf{x}\in\tilde{S}(X^n,Z^n)|_{\mathbf{z}}}f(\mathbf{y}|\mathbf{x})\\ &\geq(1-\sqrt{\epsilon})2^{-nB(b_n+\epsilon)}2^{nB(c_n-\epsilon)}\min_{\mathbf{x}\in\tilde{S}(X^n,Z^n)|_{\mathbf{z}}}f(\mathbf{y}|\mathbf{x}).\end{aligned}\qquad(34)$$

For any $\mathbf{x} \in \tilde{S}(X^n, Z^n)|_{\mathbf{z}}$, we have

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

and thus,

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

where $\epsilon_1, \epsilon_2 \to 0$ as $\epsilon \to 0$. Plugging this into (34) yields that

$$\begin{aligned} f(\mathbf{y}|\mathbf{i}) &\ge (1-\sqrt{\epsilon}) 2^{-nB(b_n+\epsilon)} 2^{nB(c_n-\epsilon)} \\ &\quad \times 2^{-nB\left(\frac{1}{2} \log 2\pi e N + a_n + \sqrt{2a_n \ln 2} \log e + \epsilon_2\right)} \\ &\ge 2^{-nB(b_n - c_n + \frac{1}{2} \log 2\pi e N + a_n + \sqrt{2a_n \ln 2} \log e + \epsilon_3)} \end{aligned}$$

for some $\epsilon_3 \to 0$ as $\epsilon \to 0$. ∎

C. Constructions of $\mathcal{I}$ and $\mathcal{Y}_{\mathbf{i}}$

Let $\mathcal{I} = \{\mathbf{i} : \Pr(S(X^n, I_n)|_{\mathbf{i}}|\mathbf{i}) \ge 1 - 2\sqrt[8]{\epsilon}\}$. For sufficiently large B, $\Pr(S(X^n, I_n)) \ge 1 - 2\sqrt[4]{\epsilon}$ from Lemma B.4, and thus by Lemma B.2 again,

$$\begin{aligned} \Pr(\mathcal{I}) &\ge \Pr\left\{\mathbf{i} : \Pr(S(X^n, I_n)|_{\mathbf{i}}|\mathbf{i}) \ge 1 - \sqrt{2\sqrt[4]{\epsilon}}\right\} \\ &\ge 1 - \sqrt{2\sqrt[4]{\epsilon}} \\ &\ge 1 - 2\sqrt[8]{\epsilon}. \end{aligned}$$

Lemma B.7: For any $\mathbf{i} \in \mathcal{I}$, let

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

Then for sufficiently large B,

$$\Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{i}) \ge 1 - 3\sqrt[8]{\epsilon},$$

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

$$f(\mathbf{y}|\mathbf{i}) \ge 2^{-nB(b_n - c_n + \frac{1}{2} \log 2\pi e N + a_n + \sqrt{2a_n \ln 2} \log e + \epsilon_3)}.$$

Proof: For any $\mathbf{i} \in \mathcal{I}$ and sufficiently large B, we have

$$\begin{aligned} &\Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{i}) \\ &= \sum_{\mathbf{x}} \Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{x}) p(\mathbf{x}|\mathbf{i}) \\ &\ge \sum_{\mathbf{x} \in S(X^n, I_n)|_{\mathbf{i}}} \Pr(\mathbf{Y} \in \mathcal{Y}_{\mathbf{i}}|\mathbf{x}) p(\mathbf{x}|\mathbf{i}) \\ &\ge \sum_{\mathbf{x} \in S(X^n, I_n)|_{\mathbf{i}}} \Pr(\mathbf{Y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n \ln 2} + 3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})|\mathbf{x}) p(\mathbf{x}|\mathbf{i}) \\ &\ge (1-\epsilon) \Pr(S(X^n, I_n)|_{\mathbf{i}}|\mathbf{i}) \\ &\ge (1-\epsilon)(1 - 2\sqrt[8]{\epsilon}) \\ &\ge 1 - 3\sqrt[8]{\epsilon}. \end{aligned}$$

Now consider any $\mathbf{y} \in \mathcal{Y}_{\mathbf{i}}$. There exists some $\mathbf{x} \in S(X^n, I_n)|_{\mathbf{i}}$ such that $\mathbf{y} \in \Gamma_{\sqrt{nB}(\sqrt{2Na_n \ln 2} + 3\sqrt{N\epsilon})}(\mathcal{Z}_{(\mathbf{x},\mathbf{i})})$. It then follows immediately from Part 2) of Lemma B.6 that

$$f(\mathbf{y}|\mathbf{i}) \ge 2^{-nB(b_n - c_n + \frac{1}{2} \log 2\pi e N + a_n + \sqrt{2a_n \ln 2} \log e + \epsilon_3)}.$$

∎

Finally, choosing δ to be $3\sqrt[8]{\epsilon}$ completes the proof of Lemma 4.1.

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] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *IEEE Trans. Info. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[4] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. of IEEE International Symposium on Information Theory*, pp. 22, June 2000.

[5] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2007.

[6] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless Network Information Flow: A Deterministic Approach," *IEEE Trans. Info. Theory*, vol. 57, no. 4, pp. 1872–1905, 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, 2011.

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

[9] S. H. Lim, Y.-H. Kim, A. El Gamal, S.-Y. Chung, "Noisy network coding," *IEEE Trans. Info. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

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

[11] T. Courtade and A Ozgur, "Approximate capacity of Gaussian relay networks: Is a sublinear gap to the cutset bound plausible?" in *Proc. of IEEE International Symposium on Information Theory*, Hong Kong, June 2015.

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

[13] 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.

[14] X. Wu, L.-L. Xie, A. Ozgur, "Upper bounds on the capacity of symmetric primitive relay channels," in *Proc. of IEEE International Symposium on Information Theory*, Hong Kong, June 2015.

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

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

[17] B. Chern and A. Ozgur, "Achieving the capacity of the n-relay Gaussian diamond network within log n bits," in *Proc. of IEEE Information Theory Workshop*, 2012.

[18] R. Kolte, and A. Ozgur, "Improved capacity approximations for Gaussian relay networks," in *Proc. of IEEE Information Theory Workshop*, 2013.

[19] S. H. Lim, K. T. Kim, and Y.-H. Kim, "Distributed decode-forward for multicast," in *Proc. of IEEE International Symposium on Information Theory*, pp. 636–640, July 2014.

[20] M. Talagrand, "Transportation cost for Gaussian and other product measures," *Geometric & Functional Analysis*, pp. 587–600.