Bounds on Capacity and Minimum Energy-Per-Bit for AWGN Relay Channels

Abbas El Gamal, Mehdi Mohseni and Sina Zahedi
Information Systems Lab
Department of Electrical Engineering
Stanford University, Stanford, CA 94305-9510
{abbas, mmohseni, szahedi}@stanford.edu

Abstract

Upper and lower bounds on the capacity and minimum energy-per-bit for general additive white Gaussian noise (AWGN) and frequency division additive white Gaussian noise (FD-AWGN) relay channel models are established. First the max-flow min-cut bound and the generalized block Markov coding scheme are used to derive upper and lower bounds on capacity. These bounds are never tight for the general AWGN model and are tight only under certain conditions for the FD-AWGN model. The gap between the upper and lower bounds is the largest when the gain of the channel to the relay is comparable or worse than that of the direct channel. To obtain tighter lower bounds, two coding schemes that do not require the relay to decode any part of the message are investigated. First the side information coding scheme is shown to outperform the block Markov coding scheme. It is shown that the achievable rate of the side-information coding scheme can be improved via time-sharing and a general expression for the achievable rate with side information is found for relay channels in general. In the second scheme, the relaying functions are restricted to be linear. A simple sub-optimal linear relaying scheme is shown to significantly outperform the generalized block Markov and the side-information coding schemes for the general AWGN model in some cases. It is shown that the optimal rate using linear relaying can be obtained by solving a sequence of non-convex optimization problems. The problem is reduces to a "single-letter" non-convex optimization problem for the FD-AWGN model. The paper establishes a relationship between the minimum energy-per-bit and capacity of the AWGN relay channel. This relationship together with the lower and upper bounds on capacity are used to establish corresponding lower and upper bounds on the minimum energy-per-bit for the general and FD-AWGN relay channels. The bounds are very close and do not differ by more than a factor of 1.45 for the FD-AWGN relay channel model and 1.7 for the general AWGN model.

Index Terms–Relay channel, channel capacity, minimum energy-per-bit, additive white Gaussian noise channels.

1 Introduction

The relay channel, first introduced by van der Meulen [1] in 1971, consists of a sender-receiver pair whose communication is aided by a relay node. In [1] and [2], simple lower and upper bounds on the capacity of the discrete-memoryless relay channel were established. In [3] and [4], capacity theorems were established for: (i) physically degraded and reversely degraded discrete memoryless relay channels, (ii) physically degraded and reversely degraded additive white Gaussian noise (AWGN) relay channels with average power constraints, (iii) deterministic relay channels, and (iv) relay channels with feedback. A max-flow min-cut upper bound and a general lower bound based on combining the generalized block Markov and side information coding schemes were also established in [4]. In [5], the capacity of the relay channel with one deterministic component was established. It is interesting to note that in all special cases where the relay channel capacity is known, it is equal to the max-flow min-cut bound. Generalizations of some of the single relay channel results to channels with many relays were given in [6]. In [7], Aref established the capacity for a cascade of degraded relay channels. The relay channel did not receive much attention and no further progress was made toward establishing its capacity for a long time after this early work.

Recent interest in multi-hop and ad hoc wireless networks has spurred the interest in studying additive white Gaussian noise (AWGN) relay channels. In [8] achievable rates for AWGN channels with two relays were investigated. In [9], the capacity of a class of orthogonal relay channels was established. In [10] upper and lower bounds on the capacity of AWGN channels were established. The capacity of AWGN relay networks as the number of nodes becomes very large were investigated in (e.g., [11]- [14]). Motivated by energy constraints in sensor and mobile networks, recent work has also investigated the saving in transmission energy using relaying [16]. In [18], upper and lower bounds on the capacity of AWGN relay channels were used to establish bounds on the minimum energy-per-bit that do not differ by more than a factor of 2. The capacity and minimum energy-per-bit for AWGN relay channels, however, are not known in general.

In this paper, we provide detailed discussion and several extensions of the bounds on capacity and minimum energy-per-bit for AWGN relay channels presented in [10] and [18], including correcting an error in the capacity with linear relaying result reported in [10]. We consider the two discrete-time AWGN relay channel models depicted in Figure 1. In the general model, Figure 1(a), the received signal at the relay and at the receiver at time $i \geq 1$ are given by

$$Y_{1i} = aX_i + Z_{1i}, \text{ and } Y_i = X_i + bX_{1i} + Z_i,$$

where X_i and X_{1i} are the signals transmitted by the sender and by the relay, respectively. The receivers' noise processes $\{Z_{1i}\}$ and $\{Z_i\}$ are assumed to be independent white Gaussian noise processes each with power N, and the constants $a, b > 0$ represent the gain of the channels from the sender to the relay and from the relay to the receiver, respectively, relative to the gain of the direct channel (which is assumed to be equal to one). The frequency-division AWGN (FD-AWGN) model is motivated by the constraint that in practice the relay cannot send and receive information within the same frequency band (or at the same time). To satisfy this constraint, one can either split the channel from the sender to the relay and the receiver into two bands or alternatively split the channel to the receiver from the sender and the relay. The capacity of the first model has been established in [9]. In this paper, we focus on the second FD-AWGN model, depicted in Figure 1(b), the received signal at time i is given by $Y_i = \{Y_{Di}, Y_{Ri}\}$, where $Y_{Di} = X_i + Z_{Di}$ is the received signal from the sender and $Y_{Ri} = bX_{1i} + Z_{Ri}$ is the received signal from the relay, and $\{Z_{Di}\}$ and $\{Z_{Ri}\}$ are independent white Gaussian noise processes each with average noise power N.

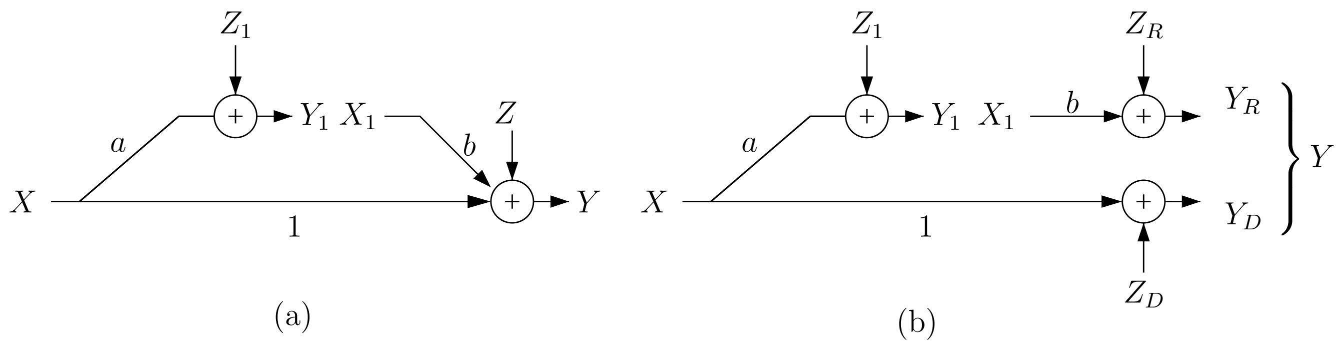


Figure 1: (a) General AWGN relay channel model. (b) FD-AWGN relay channel model. Path gains are normalized to 1 for the direct channel, $a > 0$ for the channel to the relay, and $b > 0$ for the channel from the relay to the receiver.

The paper establishes upper and lower bounds on the capacity and minimum energy-per-bit for the general and FD-AWGN relay channel models. In the following discussion we summarize our main results and provide an outline for the rest of the paper.

Bounds on capacity: In Section 2, we use the max-flow min-cut upper bound [4] and the generalized block Markov lower bound [4, 5] on the capacity of the relay channel to derive upper and lower bounds on the capacity of the general and FD-AWGN relay channels (see Table 1). The bounds are not tight for the general AWGN model for any $a, b > 0$ and are tight only for a restricted range of these parameters for the FD-AWGN model. We find that the gap between the upper and lower bounds is the largest when the channel to the relay is not much better than the channel from the sender to the receiver, i.e., a is close to 1. We argue that the reason for the large gap is that in the generalized block Markov coding scheme the relay is either required to decode the entire message or is not used at all. For a close to 1, this severely limits the achievable rate. Motivated by this observation, in Sections 4 and 5, we investigate achievable rates using two schemes where the relay cooperates in sending the message but without decoding any part of it. In Section 4, we explore achievable rates using the side information coding scheme in [4]. We find that this scheme can outperform block Markov coding and in fact becomes optimal as $b \to \infty$. We show that the achievable rate can be improved via time-sharing and provide a general expression for achievable rate with side information for relay channels in general (see Theorem 2). In Section 5, we investigate the achievable rates when the relay is restricted to sending linear combinations of past received signals. We show that when a is close to 1, a simple sub-optimal linear relaying scheme can significantly outperform the more sophisticated block Markov scheme (see Example 1). We show that the capacity with linear relaying functions can be found by solving a sequence of non-convex optimization problems. One of the main results in this paper is showing that this formulation can be reduced to a "single-letter" optimization problem for the FD-AWGN model (see Theorem 3).

Bounds on minimum energy-per-bit: In Section 3, we establish a general relationship between the minimum energy-per-bit and capacity (see Theorem 1). We use this relationship together with the lower bound on capacity based on the generalized block Markov coding scheme and the "max-flow min-cut" upper bound to establish upper and lower bounds on the minimum energy-per-bit (see Table 2). These bounds can be very close and do not differ by more than a factor of two. For the FD-AWGN model the upper and lower bounds coincide when the channel from the relay to the receiver is worse than the direct channel, i.e., $b \leq 1$. For the general AWGN model, the bounds are never tight. Using the lower bounds on capacity in Sections 4 and 5, we are able to close the gap between the lower and upper bounds to less than a factor of 1.45 for the FD-AWGN model and 1.7 for the general AWGN model.

2 Basic Bounds on Capacity

As in [4], we define a $(2^{nR_n}, n)$ code for the relay channel to consist of, (i) a set $\mathcal{M} = \{1, 2, \ldots, 2^{nR_n}\}$ of messages, (ii) a codebook $\{x^n(1), x^n(2), \ldots, x^n(2^{nR_n})\}$ consisting of codewords of length n, (iii) a set of relay functions $\{f_i\}_{i=1}^n$ such that $x_{1i} = f_i(y_{11}, y_{12}, \ldots, y_{1i-1})$, and (iv) a decoding rule $D(y^n) \in \mathcal{M}$. The average probability of decoding error is defined in the usual way as

$$\mathrm{P}_e^{(n)} = \frac{1}{2^{nR_n}} \sum_{k=1}^{2^{nR_n}} \mathrm{P}\{D(y^n) \neq k | x^n(k) \text{ transmitted}\}.$$

We assume the average transmitted power constraints

$$\sum_{i=1}^{n} x_i^2(k) \leq nP, \text{ for all } k \in \mathcal{M}, \text{ and } \sum_{i=1}^{n} x_{1i}^2 \leq n\gamma P, \text{ for all } y_1^n \in \mathbb{R}^n, \gamma \geq 0.$$

A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n)$ codes satisfying the average power constraints P on X and γP on X_1, such that $\mathrm{P}_e^{(n)} \to 0$ as $n \to \infty$. The capacity $C(P, \gamma P)$ of the AWGN relay channel with average power constraints is defined as the supremum of the set of achievable rates. To distinguish between the capacity for the general AWGN and the FD model, we label the capacity of the FD-AWGN relay channel as $C^{\mathrm{FD}}(P, \gamma P)$.

In this section, we evaluate the max-flow min-cut upper bound and the block Markov lower bounds on the capacity of the general and frequency-division AWGN channels. In particular we show that the lower bounds achieved with the block Markov encoding and the generalized block Markov encoding schemes are the same for AWGN relay channels.

First note that the capacity of both the general and FD-AWGN relay channels are lower bounded by the capacity of the direct link

$$\mathcal{C}\left(\frac{P}{N}\right) \triangleq \frac{1}{2} \log\left(1 + \frac{P}{N}\right).$$

Next, we use bounds from [4] on the capacity C of the discrete-memoryless relay channel to derive the upper and lower bounds on $C(P,\gamma P)$ and $C^{\rm FD}(P,\gamma P)$ given in Table 1.

Generalized Block Markov Lower Bound	Max-flow Min-cut Upper Bound
$R(P,\gamma P) \triangleq \begin{cases} \mathcal{C}\left(\frac{\left(b\sqrt{\gamma(a^2-1)}+\sqrt{a^2-b^2\gamma}\right)^2 P}{a^2 N}\right), & \text{if } \frac{a^2}{1+b^2\gamma} \geq 1 \\ \mathcal{C}\left(\frac{\max\{1,a^2\}P}{N}\right), & \text{otherwise} \end{cases}$	$\hat{C}(P,\gamma P) \triangleq \begin{cases} \mathcal{C}\left(\frac{\left(ab\sqrt{\gamma}+\sqrt{1+a^2-b^2\gamma}\right)^2 P}{(1+a^2)N}\right), & \text{if } \frac{a^2}{b^2\gamma} > 1 \\ \mathcal{C}\left(\frac{(1+a^2)P}{N}\right), & \text{otherwise} \end{cases}$
$R^{\rm FD}(P,\gamma P) \triangleq \begin{cases} \mathcal{C}\left(\frac{P}{N}\right)+\mathcal{C}\left(\frac{b^2\gamma P}{N}\right), & \text{if } \frac{a^2}{1+b^2\gamma\left(\frac{P}{N}+1\right)} \geq 1 \\ \mathcal{C}\left(\frac{\max\{1,a^2\}P}{N}\right), & \text{otherwise} \end{cases}$	$\hat{C}^{\rm FD}(P,\gamma P) \triangleq \begin{cases} \mathcal{C}\left(\frac{P}{N}\right)+\mathcal{C}\left(\frac{b^2\gamma P}{N}\right), & \text{if } \frac{a^2}{b^2\gamma\left(\frac{P}{N}+1\right)} \geq 1 \\ \mathcal{C}\left(\frac{(1+a^2)P}{N}\right), & \text{otherwise} \end{cases}$

Table 1: Max-flow min-cut upper bounds and generalized block Markov lower bounds on $C(P,\gamma P)$ and $C^{\rm FD}(P,\gamma P)$.

The upper bounds are derived using the "max-flow min-cut" bound on the capacity of the discrete-memoryless relay channel (Theorem 4 in [4])

$$C \leq \max_{p(x,x_1)} \min\{I(X,X_1;Y), I(X;Y,Y_1|X_1)\}. \tag{1}$$

This bound is the tightest upper bound to date on the capacity of the relay channel. For completion, derivations of the upper bounds are given in Appendix A.

The lower bounds in the table are obtained using a special case of Theorem 7 in [4] that yields

$$C \geq \max_{p(u,x,x_1)} \min\{I(X,X_1;Y), I(U;Y_1|X_1)+I(X;Y|X_1,U)\}, \tag{2}$$

This lower bound is achieved using a *generalized block Markov coding scheme*, where in each block the relay decodes part of the new message (represented by U) and cooperatively sends enough information with the sender to help the receiver decode the previous message (U then X). Note that if we set $U=X$, we obtain the rate for the block Markov scheme, which is optimal for the physically degraded relay channel, while if we set $U=\phi$, the bound reduces to the capacity of the direct channel, which is optimal for reversely degraded relay channels. In addition to these two special cases, this bound was also shown to be tight for semi-deterministic relay channels [5] and more recently for a class of relay channels with orthogonal components [9].

The bounds on the capacity of the general AWGN channel in the table are not tight for any $a,b>0$. In particular, when $a<1$, the generalized block Markov coding bound yields $\mathcal{C}\left(P/N\right)$, which is simply the capacity of the direct link. For the FD-AWGN, the bounds coincide for $a^2 \geq 1+b^2\gamma\left(1+\frac{P}{N}\right)$. In Sections 4 and 5, we show that side information and linear relaying coding schemes can provide much tighter lower bounds than the generalized block Markov bound for small values of a.

Derivation of the lower bounds: Consider the lower bound for the general AWGN case. Note that $R(P, \gamma P)$ is in fact achievable by evaluating the mutual information terms in (2) using a jointly Gaussian (U, X, X_1). We now show that the lower bound in (2) with the power constraints is upper bounded by $R(P, \gamma P)$ in Table 1. It is easy to verify that

$$I(X, X_1; Y) \le \mathcal{C}\left(\frac{(1 + b^2\gamma + 2b\rho\sqrt{\gamma})P}{N}\right),$$

where ρ is the correlation coefficient between X and X_1. Next consider

$$\begin{aligned}
I(U; Y_1|X_1) + I(X; Y|X_1, U) &= h(Y_1|X_1) - h(Y_1|X_1, U) + h(Y|X_1, U) - h(Y|X, X_1, U) \\
&\le \frac{1}{2}\log\frac{E\mathrm{Var}(Y_1|X_1)}{N} - h(Y_1|X_1, U) + h(Y|X_1, U) \\
&\le \frac{1}{2}\log\frac{a^2\left(E(X^2) - \frac{E(XX_1)^2}{E(X_1^2)}\right) + N}{N} - h(Y_1|X_1, U) + h(Y|X_1, U) \\
&\le \mathcal{C}\left(\frac{a^2(1-\rho^2)P}{N}\right) - h(Y_1|X_1, U) + h(Y|X_1, U).
\end{aligned}$$

We now find an upper bound on $h(Y|X_1, U) - h(Y_1|X_1, U)$. Note that

$$\frac{1}{2}\log(2\pi e N) \le h(Y_1|X_1, U) \le \frac{1}{2}\log(2\pi e(a^2P + N)).$$

Therefore, there exists a constant $0 \le \beta \le 1$ such that $h(Y_1|X_1, U) = \frac{1}{2}\log(2\pi e(a^2\beta P + N))$. First assume $a < 1$. Using the entropy power inequality we obtain

$$\begin{aligned}
h(aX + Z_1|X_1, U) &= h\left(a\left(X + \frac{Z_1}{a}\right)\middle| X_1, U\right) \\
&= h\left(X + \frac{Z_1}{a}\middle| X_1, U\right) + \log a \\
&= h\left(X + Z + Z'|\, X_1, U\right) + \log a \\
&\ge \frac{1}{2}\log\left(2^{2h(X+Z|X_1,U)} + 2^{2h(Z'|X_1,U)}\right) + \log a \\
&= \frac{1}{2}\log\left(2^{2h(X+Z|X_1,U)} + 2\pi e\left(\frac{1}{a^2} - 1\right)N\right) + \log a \\
&= \frac{1}{2}\log\left(2^{2h(Y|X_1,U)} + 2\pi e\left(\frac{1}{a^2} - 1\right)N\right) + \log a,
\end{aligned}$$

where $Z' \sim \mathcal{N}\left(0, \left(\frac{1}{a^2} - 1\right)N\right)$ and is independent of Z. Since $h(aX + Z_1|X_1, U) = \frac{1}{2}\log(2\pi e(a^2\beta P + N))$, we obtain

$$h(Y|X_1, U) \le \frac{1}{2}\log\left(2\pi e(\beta P + N)\right),$$

and

$$h(Y|X_1, U) - h(Y_1|X_1, U) \le \frac{1}{2}\log\left(\frac{\beta P + N}{a^2\beta P + N}\right) \le \frac{1}{2}\log\left(\frac{P + N}{a^2P + N}\right).$$

Hence, for $a < 1$,

$$I(U;Y_1|X_1) + I(X;Y|X_1,U) \leq \frac{1}{2}\log\left(\frac{a^2(1-\rho^2)P+N}{N}\right) + \frac{1}{2}\log\left(\frac{P+N}{a^2P+N}\right) \leq \mathcal{C}\left(\frac{P}{N}\right).$$

For $a > 1$, note that $h(Y_1|X_1,U) = h(aX+Z_1|X_1,U) = h(aX+Z|X_1,U) \geq h(X+Z|X_1,U) = h(Y|X_1,U)$ and hence

$$I(U;Y_1|X_1) + I(X;Y|X_1,U) \leq \mathcal{C}\left(\frac{a^2(1-\rho^2)P}{N}\right).$$

Note that the above bounds are achieved by choosing (U, X_1, X) to be jointly Gaussian with zero mean and appropriately chosen covariance matrix. Performing the maximization over ρ gives the lower bound result in Table 1. This completes the derivation of the lower bound. The lower bound for the FD-AWGN case can be similarly derived.

3 Basic Bounds on Minimum Energy-per-Bit

In this section we establish a general relationship between minimum energy-per-bit and capacity with average power constraints for the discrete time AWGN relay channel. We then use this relationship and the bounds on capacity established in the previous section to find lower and upper bounds on the minimum energy-per-bit.

The minimum energy-per-bit can be viewed as a special case of the reciprocal of the capacity per-unit-cost [19], when the cost is average power. In [22], Verdu established a relationship between capacity per-unit-cost and channel capacity for stationary memoryless channels. He also found the capacity per-unit cost region for multiple access and interference channels. Here, we define the minimum energy-per-bit directly and not as a special case of capacity per unit cost. We consider a sequence of codes where the rate $R_n \geq 1/n$ can vary with n. This allows us to define the minimum energy-per-bit in an unrestricted way. The energy for codeword k is given by

$$\mathcal{E}^{(n)}(k) = \sum_{i=1}^{n} x_i^2(k).$$

The maximum relay transmission energy is given by

$$\mathcal{E}_r^{(n)} = \max_{y_1^n}\left(\sum_{i=1}^{n} x_{1i}^2\right).$$

The energy-per-bit for the code is therefore given by

$$\mathcal{E}^{(n)} = \frac{1}{nR_n}\left(\max_k \mathcal{E}^{(n)}(k) + \mathcal{E}_r^{(n)}\right).$$

An energy-per-bit $\mathcal{E}$ is said to be achievable if there is a sequence of $(2^{nR_n}, n)$ codes with $\mathrm{P}_e^{(n)} \to 0$ and $\limsup \mathcal{E}^{(n)} \leq \mathcal{E}$. We define the minimum energy-per-bit as the energy-per-bit that can be achieved with no constraint on the rate. More formally

Definition 1 *The minimum energy-per-bit $\mathcal{E}_b$ is the infimum of the set of achievable energy-per-bit values.*

To distinguish between the minimum energy-per-bit for the general and FD-AWGN channel models, we label the minimum energy-per-bit for the FD-AWGN relay channel by $\mathcal{E}_b^{\rm FD}$. In the discussion leading to Theorem 1, we derive a relationship between $C(P,\gamma P)$ and $\mathcal{E}_b$. The statements and results including the theorem apply with no change if we replace $C(P,\gamma P)$ by $C^{\rm FD}(P,\gamma P)$ and $\mathcal{E}_b$ by $\mathcal{E}_b^{\rm FD}$.

First note that $C(P,\gamma P)$ can be expressed as

$$C(P,\gamma P) = \sup_k C_k(P,\gamma P) = \lim_{k\to\infty} C_k(P,\gamma P), \tag{3}$$

where

$$C_k(P,\gamma P) = \frac{1}{k} \sup_{\substack{\mathrm{P}_{\mathbf{X}^k},\{f_i\}_{i=1}^k :\ \sum_{i=1}^k E(X_i^2)\le kP, \\ \max_{y_1^n}\left(\sum_{i=1}^k x_{1i}^2\right)\le k\gamma P}} I(\mathbf{X}^k;\mathbf{Y}^k), \tag{4}$$

where $x_{1i} = f_i(y_{11},\ldots,y_{1(i-1)})$, and $\mathbf{X}^k = [X_1,X_2,\ldots,X_k]^T$. Note that by a standard random coding argument, any rate less than $C_k(P,\gamma P)$ is achievable. It is easy to argue that $kC_k(P,\gamma P)$ as defined in (4) is a super-additive sequence in k, i.e., $(k+m)C_{k+m}(P,\gamma P) \ge kC_k(P,\gamma P) + mC_m(P,\gamma P)$ for any $k,m\ge 1$. Hence the supremum in (3) can be replaced with the limit. We now establish the following properties of $C(P,\gamma P)$ as a function of P.

Lemma 1 *The capacity of the AWGN relay channel with average power constraints satisfies the following:*

(i) $C(P,\gamma P) > 0$ *if* $P>0$ *and approaches* ∞ *as* $P\to\infty$*.*

(ii) $C(P,\gamma P)\to 0$ *as* $P\to 0$*.*

(iii) $C(P,\gamma P)$ *is concave and strictly increasing in* P*.*

(iv) $\frac{(1+\gamma)P}{C(P,\gamma P)}$ *is non-decreasing in* P*, for all* $P>0$*.*

The proof of this lemma is provided in Appendix D.

We are now ready to establish the following relationship between the minimum energy-per-bit and capacity with average power constraint for the AWGN relay channel.

Theorem 1 *The minimum energy-per-bit for the AWGN relay channel is given by*

$$\mathcal{E}_b = \inf_{\gamma\ge 0}\lim_{P\to 0}\frac{(1+\gamma)P}{C(P,\gamma P)}. \tag{5}$$

Proof: We establish achievability and weak converse to show that

$$\mathcal{E}_b = \inf_{\gamma \geq 0} \inf_{P>0} \frac{(1+\gamma)P}{C(P, \gamma P)}.$$

From part (iv) of Lemma 1, the second inf can be replaced by lim.

To show achievability, we need to show that any $\mathcal{E} > \inf_{\gamma\geq 0} \inf_{P>0} \frac{(1+\gamma)P}{C(P,\gamma P)}$ is achievable. First note that there exist $P' > 0$ and $\gamma' \geq 0$ such that $\mathcal{E} > \frac{(1+\gamma')P'}{C(P',\gamma' P')} = \inf_{\gamma\geq 0} \inf_{P>0} \frac{(1+\gamma)P}{C(P,\gamma P)} + \epsilon$ for any small $\epsilon > 0$. Now we set $R = (1+\gamma')P'/\mathcal{E}$. Using standard random coding with power constraints argument, we can show that $R < C(P', \gamma' P')$ is achievable, which proves the achievability of $\mathcal{E}$.

To prove the weak converse we need to show that for any sequence of $(2^{nR_n}, n)$ codes with $\mathrm{P}_e^{(n)} \to 0$,

$$\liminf \mathcal{E}^{(n)} \geq \mathcal{E}_b = \inf_{\gamma \geq 0} \inf_{P>0} \frac{(1+\gamma)P}{C(P, \gamma P)}.$$

Using Fano's inequality, we can easily arrive at the bound

$$R_n \leq C(P_n, \gamma_n P_n) + \frac{1}{n} H(\mathrm{P}_e^{(n)}) + R_n \mathrm{P}_e^{(n)},$$

where $P_n > 0$ is the maximum average codeword power and $\gamma_n P_n$ is the maximum average transmitted relay power. Thus

$$R_n \leq \frac{C(P_n, \gamma_n P_n) + \frac{1}{n} H(\mathrm{P}_e^{(n)})}{1 - \mathrm{P}_e^{(n)}}.$$

Now, by definition the energy-per-bit for the code is $\mathcal{E}^{(n)} \geq P_n(1+\gamma_n)/R_n$. Using the bound on R_n we obtain the bound

$$\begin{aligned}
\mathcal{E}^{(n)} &\geq \frac{P_n(1+\gamma_n)(1-\mathrm{P}_e^{(n)})}{C(P_n, \gamma_n P_n) + \frac{1}{n} H(\mathrm{P}_e^{(n)})} \\
&= \frac{P_n(1+\gamma_n)(1-\mathrm{P}_e^{(n)})}{C(P_n, \gamma_n P_n)(1 + \frac{1}{n} H(\mathrm{P}_e^{(n)})/C(P_n, \gamma_n P_n))} \\
&\geq \frac{\mathcal{E}_b(1-\mathrm{P}_e^{(n)})}{1 + \frac{1}{n} H(\mathrm{P}_e^{(n)})/C(P_n, \gamma_n P_n)}.
\end{aligned}$$

Now since $\mathrm{P}_e^{(n)} \to 0$ and $H(\mathrm{P}_e^{(n)}), C(P_n, \gamma_n P_n) > 0$, we get that $\liminf \mathcal{E}^{(n)} \geq \mathcal{E}_b$. ■

We now use the above relationship and the bounds on capacity to establish lower and upper bounds on the minimum energy-per-bit. First note that the minimum energy-per-bit for the direct channel, given by $2N \ln 2$, is an upper bound on the minimum energy-per-bit for both relay channel models considered. Using Theorem 1 and the bounds on capacity given in Table 1, we obtain the lower and upper bounds on the minimum energy-per-bit

Normalized Minimum Energy-per-bit	Max-flow Min-cut Lower Bound	Block Markov Upper Bound
$\mathcal{E}_b/2N\ln 2$	$\frac{1+a^2+b^2}{(1+a^2)(1+b^2)}$	$\min\left\{1, \frac{a^2+b^2}{a^2(1+b^2)}\right\}$
$\mathcal{E}_b^{\mathrm{FD}}/2N\ln 2$	$\min\left\{1, \frac{a^2+b^2}{b^2(1+a^2)}\right\}$	$\frac{a^2+b^2-1}{a^2b^2}$, if $a,b>1$; 1, otherwise

Table 2: Lower and upper bounds on $\mathcal{E}_b/2N\ln 2$ and $\mathcal{E}_b^{\mathrm{FD}}/2N\ln 2$.

provided in Table 2. The bounds in the table are normalized with respect to $2N\ln 2$ and therefore represent the reduction in the energy-per-bit using relaying.

It is easy to verify that the ratio of the upper bound to the lower bound for each channel is always less than 2. This maximum ratio is approached for $a=1$ as $b\to\infty$, i.e., when the relay and receiver are at the same distance from the transmitter and the relay is very close to the receiver. Note that for the FD-AWGN channel, the lower and upper bounds coincide and are equal to $2N\ln 2$ for $b\leq 1$, and therefore relaying does not reduce the minimum energy-per-bit. For the general AWGN channel, the ratio is very close to 1 when a or b are small, or when a is large. We now derive the upper and lower bounds for the general AWGN model. The bounds for the FD-AWGN relay channel can be similarly established.

Derivation of bounds: Using Theorem 1 and the bounds on capacity derived in section 2, we now establish the bounds on the minimum energy-per-bit of the general AWGN relay channel given in table, i.e.,

$$\frac{1+a^2+b^2}{(1+a^2)(1+b^2)} \leq \frac{\mathcal{E}_b}{2N\ln 2} \leq \min\left\{1, \frac{a^2+b^2}{a^2(1+b^2)}\right\}.$$

To prove the lower bound we use the upper bound $\hat{C}(P,\gamma P)$ on capacity in Table 1 and the relationship of Theorem 1 to obtain the bound. Substituting the upper bound given in Table 1 and taking limits as $P\to 0$, we obtain the expression

$$\mathcal{E}_b \geq 2N\ln 2 \times \min\left\{\min_{0\leq\gamma<\frac{a^2}{b^2}} \frac{(1+\gamma)(1+a^2)}{\left(ab\sqrt{\gamma}+\sqrt{1+a^2-b^2\gamma}\right)^2}, \min_{\gamma\geq\frac{a^2}{b^2}} \frac{1+\gamma}{1+a^2}\right\}.$$

To complete the derivation of the lower bound, we analytically perform the minimization. For $\gamma\geq a^2/b^2$, it is easy to see that the minimum is achieved by making γ as small as possible, i.e., $\gamma=a^2/b^2$, and the minimum is given by $\frac{a^2+b^2}{b^2(1+a^2)}$. On the other hand, if $\gamma<a^2/b^2$, the minimum is achieved when $\gamma=a^2b^2/(a^2+b^4+2b^2+1)<a^2/b^2$ and is given by $\frac{1+a^2+b^2}{(1+a^2)(1+b^2)}$. Now, since $\frac{1+a^2+b^2}{(1+a^2)(1+b^2)}<\frac{a^2+b^2}{b^2(1+a^2)}$, the minimum is given by $\frac{1+a^2+b^2}{(1+a^2)(1+b^2)}$, which establishes the lower bound.

Now we turn our attention to upper bounds on minimum energy-per-bit. Using the lower bound on capacity given in Table 1 and the relationship in Theorem 1, we can obtain an upper bound for $a > 1$. Note that the capacity lower bound $R(P, \gamma P)$ in Table 1 satisfies the conditions on $C(P, \gamma P)$ in Lemma 1, and therefore, the best upper bound is given by

$$\mathcal{E}_b \leq \inf_{\gamma \geq 0} \lim_{P \to 0} \frac{(1+\gamma)P}{R(P, \gamma P)}.$$

Now we show that this bound gives

$$\mathcal{E}_b \leq 2N \ln 2 \times \min \left\{ 1, \frac{a^2 + b^2}{a^2(1+b^2)} \right\}.$$

Substituting the lower bound $R(P, \gamma P)$ from Table 1 in theorem 1 and taking the limit as $P \to 0$, for $a > 1$ we obtain

$$\mathcal{E}_b \leq 2N \ln 2 \times \min \left\{ \min_{0 \leq \gamma < \frac{a^2-1}{b^2}} \frac{(1+\gamma)a^2}{\left(b\sqrt{(a^2-1)\gamma} + \sqrt{a^2 - b^2\gamma}\right)^2}, \min_{\gamma \geq \frac{a^2-1}{b^2}} \frac{1+\gamma}{a^2} \right\}.$$

To evaluate this bound we use the same approach we used in evaluating the lower bound. We consider the two cases $\gamma < (a^2-1)/b^2$ and $\gamma \geq (a^2-1)/b^2$ and find that the minimization is achieved for $\gamma = (a^2-1)b^2/(b^4 + 2b^2 + a^2) < (a^2-1)/b^2$, and the bound is given by the expression in the theorem. For $a < 1$, the best upper bound using the generalized block Markov encoding is given by $2N \ln 2$. Since $\frac{a^2+b^2}{a^2(1+b^2)} > 1$ for $a < 1$, this completes the proof.

Remark: Note that the definition of the energy-per-bit assigned equal cost to the energy expended by the sender and the relay. One can easily extend the bounds in Theorem 1 to the case where the energy of the sender or relay is more costly by adding a weight $\lambda > 0$ to the relay part of the energy-per-bit definition for a code. The new definition is

$$\mathcal{E}^{(n)} = \frac{1}{nR_n} \left(\max_k \mathcal{E}^{(n)}(k) + \lambda \mathcal{E}_r^{(n)} \right).$$

It is easy to see that the bounds in this section hold with b replaced by $b/\sqrt{\lambda}$.

4 Side-Information Lower Bounds

The lower bounds on capacity given in Table 1 are based on the generalized block Markov encoding scheme. In this scheme, the relay node is either required to fully decode the message transmitted by the sender or is not used at all. When the channel from the transmitter to the relay is not much better than the channel from the sender to the receiver, i.e., $a \approx 1$, the achievable rate is very close to the capacity of the direct link. This motivates the investigation of coding schemes that use the relay to help in transmitting the message without decoding the message itself.

In this section, we consider the case where the relay node facilitates the transmission from the sender to the receiver by sending a "quantized" version of its received signal using the Wyner-Ziv coding scheme.

First we show that the lower bound on the capacity of the discrete-memoryless relay channel using this scheme (see Theorem 6 in [4]) can be recast into the bound

$$C \geq \max_{p(x)p(x_1)p(\hat{y}_1|y_1,x_1)} \min\left\{I(X;Y,\hat{Y}_1|X_1), I(X,X_1;Y) - I(Y_1;\hat{Y}_1|X,X_1)\right\}. \tag{6}$$

The derivation of this result is given in Appendix B.

Using this result and choosing $(X, X_1, \hat{Y}_1)$ to be jointly Gaussian, which is not necessarily the optimal choice, we can establish the following lower bound on capacity of the general AWGN relay channel

$$C(P,\gamma P) \geq \mathcal{C}\left(\frac{P}{N}\left(1 + \frac{a^2b^2\gamma P}{P(1+a^2+b^2\gamma)+N}\right)\right). \tag{7}$$

Achievability of this lower bound can be established by extending the achievability for the discrete-memoryless case given in [4] to the AWGN case using standard arguments found in [26].

Note that this achievable rate is not a concave function of P. The intuitive reason is that as P decreases, the received signal at the relay becomes too noisy in this case and therefore the relay transmission mostly contains information about the received noise sequence at the relay. The effectiveness of this scheme can be improved for low P by transmitting at power P/α for a fraction $0 < \alpha \leq 1$ of the time, and not transmitting the rest of the time. Using this time-sharing argument, the following tighter lower bound can be established.

$$C(P,\gamma P) \geq \max_{0<\alpha\leq 1} \alpha\mathcal{C}\left(\frac{P}{\alpha N}\left(1 + \frac{a^2b^2\gamma P}{P(1+a^2+b^2\gamma)+\alpha N}\right)\right). \tag{8}$$

Similar derivation using equation (6) for the FD-AWGN relay channel yields the lower bound

$$C^{\text{FD}}(P,\gamma P) \geq \mathcal{C}\left(\frac{P}{N}\left(1 + \frac{a^2b^2\gamma P(P+N)}{a^2PN+(P+N)(b^2\gamma P+N)}\right)\right). \tag{9}$$

It can be easily shown that the bounds in (7) and (9) become tight as $\sqrt{\gamma}b \to \infty$ or $P \to \infty$. Now, we show that we can do better by modifying the side information scheme in [4] to include time-sharing.

Theorem 2 *The capacity of the discrete-memoryless relay channel* $(\mathcal{X}\times\mathcal{X}_1, p(y,y_1|x,x_1), \mathcal{Y}\times\mathcal{Y}_1)$ *is lower bounded by*

$$C \geq \max_{p(q)p(x|q)p(x_1|q)p(\hat{y}_1|y_1,x_1,q)} \min\left\{I(X;Y,\hat{Y}_1|X_1,Q), I(X,X_1;Y|Q) - I(Y_1;\hat{Y}_1|X,X_1,Q)\right\}, \tag{10}$$

where Q is a discrete random variable with finite domain. Achievability for each value of the time-sharing random variable Q follows from Theorem 6 in [4]. The rate is then achieved by time-sharing among the codes for the different values of Q according to its optimal distribution.

Evaluating this bound for jointly Gaussian $(X, X_1, \hat{Y}_1)$ (see Appendix C) and using time-sharing only on the broadcast side of the channel, the following lower bound on the capacity of the FD-AWGN relay channel can be obtained

$$C^{\mathrm{FD}}(P, \gamma P) \geq \max_{0<\alpha\leq 1} \alpha \mathcal{C}\left(\frac{P}{\alpha N}\left(1+\frac{a^2}{1+\eta}\right)\right), \tag{11}$$

where η is given by

$$\eta = \frac{1+\frac{a^2 P}{P+\alpha N}}{-1+\left(1+\frac{b^2\gamma P}{N}\right)^{\frac{1}{\alpha}}}.$$

Setting $\alpha = 1$, gives the lower bound in (9). Hence this bound is also tight as $\sqrt{\gamma}b \to \infty$ or $P \to \infty$.

4.1 Upper bound on minimum energy-per-bit

The lower bound on capacity of the FD-AWGN relay channel using side information coding leads to a tighter upper bound on the minimum energy-per-bit for some values of a and b. Using the side information coding scheme, the maximum ratio reduces to less than 1.45 (versus 2 using the generalized block Markov coding). The same coding scheme also reduces the maximum ratio to less than 1.9 for the general AWGN relay channel.

5 Achievable Rates Using Linear Relaying Functions

In this section, we investigate a simple relaying scheme, where the relay sends a linear combination of its past received signals. We show that this relatively simple scheme can outperform the generalized block Markov encoding scheme when the channel from the transmitter to the relay is not much better than the direct channel. We show that it can also outperform the more sophisticated side information encoding scheme in some cases.

We restrict the relay functions to be linear in the past received sequence of symbols at relay. Thus at time i, the transmitted relay symbol can be expressed as $x_{1i} = \sum_{j=1}^{i-1} d_{ij} y_{1j}$, where the d_{ij}s are coefficients chosen such that the average power constraint for the relay is satisfied. Equivalently, for a transmission block length k, the transmitted relay vector $\mathbf{X}_1^k = [X_{11}, X_{12}, \ldots, X_{1k}]^T$ can be expressed in terms of the received relay vector $\mathbf{Y}_1^k = [Y_{11}, Y_{12}, \ldots, Y_{1k}]^T$ as $\mathbf{X}_1^k = D\mathbf{Y}_1^k$, where D is a strictly lower triangular weight matrix. We define the capacity with linear relaying for the general AWGN relay channel $C^{\mathrm{L}}(P, \gamma P)$ to be the supremum over the set of achievable rates with linear relaying. The capacity with linear relaying for the FD-AWGN channel is labeled as $C^{\mathrm{FD-L}}(P, \gamma P)$. First we establish a general expression for $C^{\mathrm{L}}(P, \gamma P)$. The same expression applies to $C^{\mathrm{FD-L}}(P, \gamma P)$ with a small modification that simplifies our derivations.

Similar to (3), the capacity with linear relaying for the general AWGN relay channel can be expressed as

$$C^{\mathrm{L}}(P,\gamma P)=\sup_k C_k^{\mathrm{L}}(P,\gamma P)=\lim_{k\to\infty} C_k^{\mathrm{L}}(P,\gamma P), \tag{12}$$

where

$$C_k^{\mathrm{L}}(P,\gamma P)=\frac{1}{k}\sup_{\substack{\mathrm{P}_{\mathbf{X}^k},D:\ \sum_{i=1}^k E(X_i^2)\le kP,\\ \max_{y_1^n}\left(\sum_{i=1}^k x_{1i}^2\right)\le k\gamma P}} I(\mathbf{X}^k;\mathbf{Y}^k). \tag{13}$$

Note that $C_1^{\mathrm{L}}(P,\gamma P)=\mathcal{C}(P/N)$. Also, by a standard random coding argument, any rate less than $C_k^{\mathrm{L}}(P,\gamma P)$ is achievable.

The maximization in Equation (13) is achieved when $\mathbf{X}^k$ is Gaussian. Denoting the covariance matrix of $\mathbf{X}^k$ by Σ, it is easy to show that

$$C_k^{\mathrm{L}}(P,\gamma P)=\frac{1}{2k}\max_{\Sigma,D}\log\frac{|(I+abD)\Sigma(I+abD)^T+N(I+b^2DD^T)|}{|N(I+b^2DD^T)|}, \tag{14}$$

subject to $\Sigma\succeq 0$, $\mathrm{tr}(\Sigma)\le kP$, $\mathrm{tr}(a^2D\Sigma D^T+NDD^T)\le k\gamma P$, and $d_{ij}=0$, for $j\ge i$. For $k>2$, this is a non-convex optimization problem. Thus finding $C^{\mathrm{L}}(P,\gamma P)$ involves solving a sequence of non-convex optimization problems, a daunting task indeed! Interestingly, we can show that even a sub-optimal linear relaying scheme can outperform generalized block Markov coding for small values of a.

Example 1: Consider the following linear relaying scheme for the general AWGN relay channel with block length $k=2$ (See Figure 2). In the first transmission, the sender's signal is $X_1\sim\mathcal{N}(0,2\beta P)$, for $0<\beta\le 1$, and the relay's signal is $X_{11}=0$. The received signal at the receiver and the relay receiver are $Y_1=X_1+Z_1$ and $Y_{11}=aX_1+Z_{11}$, respectively. In the second transmission, the sender's signal is $X_2=\sqrt{(1-\beta)/\beta}X_1$, i.e., a scaled version of X_1 with average power $2(1-\beta)P$. The relay cooperates with the sender by relaying a scaled version of Y_{11}, $X_{12}=dY_{11}$, where $d=\sqrt{2\gamma P/(2a^2\beta P+N)}$ is chosen to satisfy the relay sender power constraint. Thus,

$$\Sigma=2P\begin{bmatrix}\beta & \sqrt{\beta(1-\beta)}\\ \sqrt{\beta(1-\beta)} & 1-\beta\end{bmatrix},\ D=\begin{bmatrix}0 & 0\\ d & 0\end{bmatrix}.$$

The received signal after the second transmission is given by $Y_2=X_2+dbY_{11}+Z_2$.

It can be easily shown that the rate achieved by this scheme is given by

$$r^{\mathrm{L}}(P,\gamma P)=\frac{1}{2}I(X_1,X_2;Y_1,Y_2)=\max_{0\le\beta\le 1}\frac{1}{2}\mathcal{C}\left(\frac{2\beta P}{N}\left(1+\frac{\left(\sqrt{\frac{1-\beta}{\beta}}+abd\right)^2}{1+b^2d^2}\right)\right), \tag{15}$$

where $d=\sqrt{2\gamma P/(2a^2\beta P+N)}$.

This scheme is not optimal even among linear relaying schemes with block length $k=2$. However, as demonstrated in Figure 3, it achieves higher rate than the generalized block

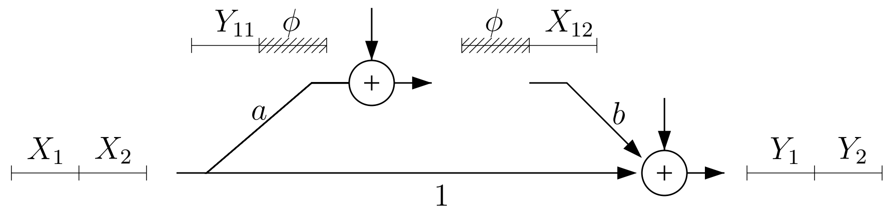


Figure 2: Sub-optimal linear relaying scheme for general AWGN relay channel with block length $k = 2$.

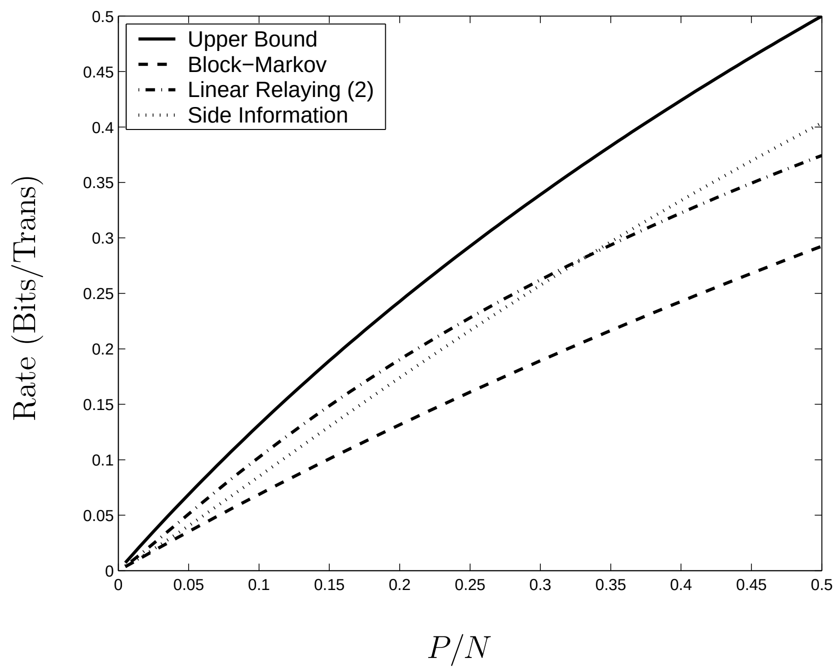


Figure 3: Comparison of achievable rates based on the generalized block Markov, linear relaying and side information encoding schemes for the general AWGN relay channel for $a = 1$ and $\sqrt{\gamma}b = 2$.

Markov coding when a is small and can also outperform the more sophisticated side information coding scheme when all distributions are Gaussians.

Next we consider the FD-AWGN relay channel with linear relaying functions. Since the channel from the relay to the receiver uses a different frequency band than the channel from the sender, without loss of generality we assume that ith relay transmission can depend on all received signals up to i (instead of $i-1$). With this relabeling, for block length k, the transmitted vector $\mathbf{X}^k = [X_1, X_2, \ldots, X_k]^T$, the transmitted relay vector $\mathbf{X}_1^k = [X_{11}, X_{12}, \ldots, X_{1k}]^T = D\mathbf{Y}_1^k$, where D is a lower triangular weight matrix with possibly non-zero diagonal elements, and the received vector $\mathbf{Y}^k = [\mathbf{Y}_D^{k\,T}, \mathbf{Y}_R^{k\,T}]^T$, where $\mathbf{Y}_D^k = [Y_{D1}, Y_{D2}, \ldots, Y_{Dk}]^T$ and $\mathbf{Y}_R^k = [Y_{R1}, Y_{R2}, \ldots, Y_{Rk}]^T$. The capacity with linear relaying can be expressed as in (12), where

$$C_k^{\mathrm{FD-L}}(P, \gamma P) = \frac{1}{k} \sup_{\substack{\mathrm{P}_{\mathbf{X}^k},\, D: \quad \sum_{i=1}^k E(X_i^2) \le kP, \\ \max_{y_1^n} \left(\sum_{i=1}^k x_{1i}^2\right) \le k\gamma P}} I(\mathbf{X}^k; \mathbf{Y}^k). \tag{16}$$

It can be easily shown that the above maximization is achieved when $\mathbf{X}^k$ is Gaussian. Denoting the covariance of $\mathbf{X}^k$ by Σ, we can reduce (16) to

$$C_k^{\mathrm{FD-L}}(P, \gamma P) = \frac{1}{2k} \max_{\Sigma, D} \log \frac{\left| \begin{array}{cc} \Sigma + NI & ab\Sigma D^T \\ abD\Sigma & a^2b^2 D\Sigma D^T + N(I + b^2 DD^T) \end{array} \right|}{\left| N \left[\begin{array}{cc} I & 0 \\ 0 & I + b^2 DD^T \end{array} \right] \right|}, \tag{17}$$

where the maximization is subject to $\Sigma \succeq 0$, $\mathrm{tr}(\Sigma) \le kP$, $\mathrm{tr}(a^2 D\Sigma D^T + NDD^T) \le k\gamma P$ and $d_{ij} = 0$ for $i > j$. This is again a non-convex optimization problem and finding $C^{\mathrm{FD-L}}(P, \gamma P)$ reduces to solving a sequence of such problems indexed by k. Luckily in this case we can reduce the problem to a "single-letter" optimization problem. Before proceeding to prove this result, consider the following simple amplify-and-forward scheme.

Example 2: We consider block length $k = 1$. It is easy to show that Equation (17) reduces to

$$C_1^{\mathrm{FD-L}}(P, \gamma P) = \mathcal{C}\left(\frac{P}{N} \left(1 + \frac{a^2 b^2 \gamma P}{(a^2 + b^2\gamma)P + N} \right) \right), \tag{18}$$

which can be achieved by the simple amplify-and-forward scheme depicted in Figure 4, with $X_1 \sim \mathcal{N}(0, P)$ and $X_{11} = \sqrt{\frac{\gamma P}{a^2 P + N}} Y_1$.

It can be shown that if $ab\sqrt{\gamma} < 1/\sqrt{2}$, then $C_1^{\mathrm{FD-L}}(P, \gamma P)$ is a concave function of P. Otherwise, it is convex for small values of P and concave for large values of P. The interpretation is that as P is decreased, linear relaying injects more noise than signal, thus becoming less helpful. In such cases, the performance of this scheme can be improved by time-sharing between amplify-and-forward for a fraction $0 \le \alpha \le 1$ of the time and direct transmission, where only the sender node transmits, for the rest of the time. The following lower bound can be established using this scheme and can be shown to be concave in P for all parameter values and all $P > 0$ (See Figure 6 for an illustration).

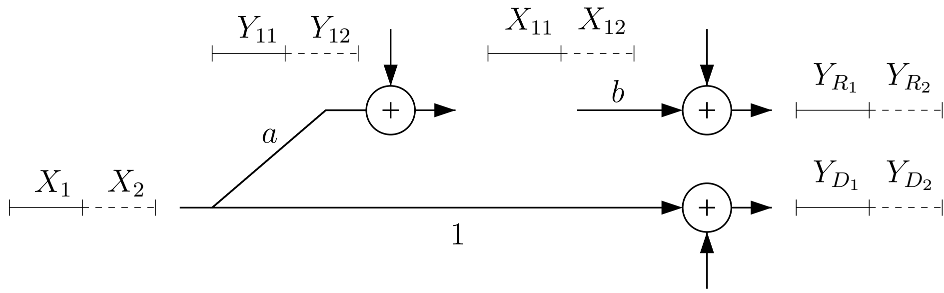


Figure 4: Linear relaying scheme for FD-AWGN relay channel with block length $k = 1$.

$$C^{\mathrm{FD-L}}(P, \gamma P) \geq \max_{0<\alpha,\theta\leq 1} \overline{\alpha}\mathcal{C}\left(\frac{\overline{\theta}P}{\overline{\alpha}N}\right) + \alpha\mathcal{C}\left(\frac{\theta P}{\alpha N}\left(1 + \frac{a^2 b^2 \gamma P}{a^2\theta P + b^2\gamma P + \alpha N}\right)\right), \tag{19}$$

where $\overline{\alpha} = 1 - \alpha$ and $\overline{\theta} = 1 - \theta$.

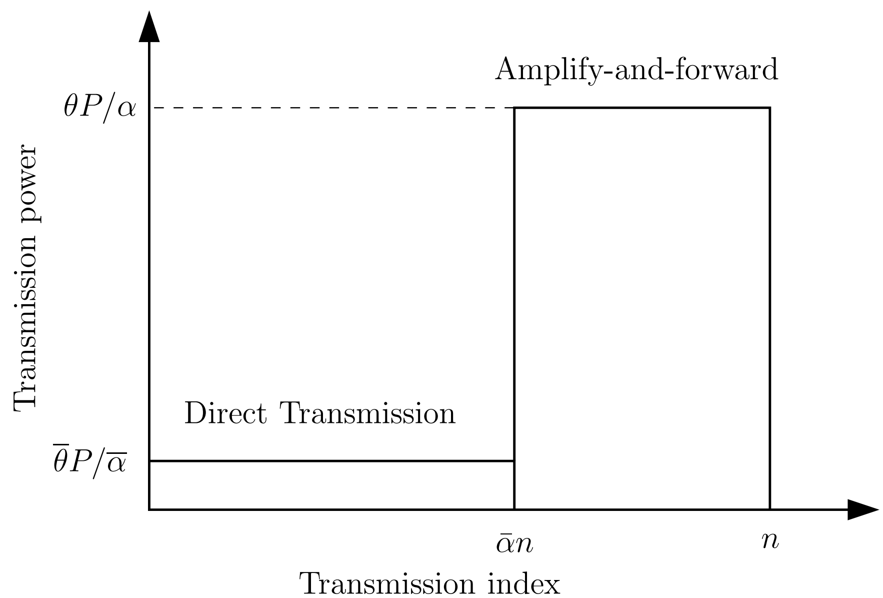


Figure 5: Time-sharing between amplify-and-forward and direct transmission only.

As proved in the following theorem, the capacity with linear relaying can be achieved using time-sharing between direct transmission and amplify-and-forward with at most four different power and relaying parameter settings.

Theorem 3 *The capacity of FD-AWGN relay channel with linear relaying functions is given by*

$$C^{\mathrm{FD-L}}(P, \gamma P) = \max_{\boldsymbol{\alpha},\, \boldsymbol{\theta},\, \boldsymbol{\eta}} \alpha_0 \mathcal{C}\left(\frac{\theta_0 P}{\alpha_0 N}\right) + \sum_{j=1}^{4} \alpha_j \mathcal{C}\left(\frac{\theta_j P}{\alpha_j N}\left(1 + \frac{a^2 b^2 \eta_j}{1 + b^2 \eta_j}\right)\right), \tag{20}$$

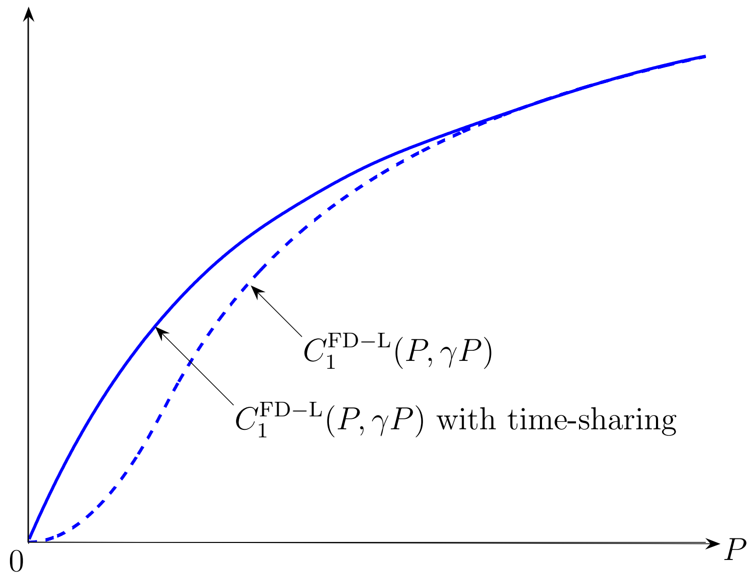


Figure 6: Comparison of achievable rates using amplify-and-forward and time-sharing between amplify-and-forward and direct transmission only.

where $\boldsymbol{\alpha} = [\alpha_0,\ \alpha_1, \dots, \alpha_4]$, $\boldsymbol{\theta} = [\theta_0,\ \theta_1, \dots, \theta_4]$, $\boldsymbol{\eta} = [\eta_1,\ \eta_2, \dots, \eta_4]$, *subject to* $\alpha_j, \theta_j \geq 0$, $\eta_j > 0$, $\sum_{j=0}^4 \alpha_j = \sum_{j=0}^4 \theta_j = 1$, *and* $\sum_{j=1}^4 \eta_j\left(a^2\theta_j P + N\alpha_j\right) = \gamma P$.

Proof: We first outline the proof. The main difficulty in solving the optimization problem is that it is not concave in Σ and D. However, for any fixed D, the problem is concave in Σ and the optimal solution can be readily obtained using convex optimization techniques [27]. We show that for any matrix D, there exists a diagonal matrix L whose associated optimal covariance matrix Ψ is also diagonal and such that the value of the objective function for the pair (D, Σ) is the same as that for the pair (L, Ψ). Hence the search for the optimal solution can be restricted to the set of diagonal D and Σ matrices. The reduced optimization problem, however, remains non-convex. We show, however, that it can be reduced to a non-convex constrained optimization problem with 14 variables and 3 equality constraints.

Simplifying the expression for $C_k^{\mathrm{FD-L}}(P, \gamma P)$ in (17), the optimization problem can be expressed as

$$C_k^{\mathrm{FD-L}}(P,\gamma P) = \max_{\Sigma, D} \frac{1}{2k} \log \left| I + \frac{1}{N} \begin{bmatrix} I \\ ab(I + b^2 DD^T)^{-1/2} D \end{bmatrix} \Sigma \begin{bmatrix} I \\ ab(I + b^2 DD^T)^{-1/2} D \end{bmatrix}^T \right|, \tag{21}$$

subject to $\Sigma \succeq 0$, $\mathrm{tr}(\Sigma) \leq kP$ and $\mathrm{tr}(a^2 D\Sigma D^T + NDD^T) \leq k\gamma P$, where D is a lower triangular matrix.

Finding the optimal Σ for any fixed D is a convex optimization problem [27]. Defining

$$G = \frac{1}{\sqrt{N}} \begin{bmatrix} I \\ ab(I + b^2 DD^T)^{-1/2} D \end{bmatrix},$$

and neglecting the $\frac{1}{2k}$ factor, the Lagrangian for this problem can be expressed as

$$L(\Sigma, \Phi, \lambda, \mu) = -\log|I + G\Sigma G^T| - \mathrm{tr}(\Phi\Sigma) + \lambda(\mathrm{tr}(\Sigma) - kP) + \mu(\mathrm{tr}(a^2 D\Sigma D^T + NDD^T) - k\gamma P).$$

The Karush-Kuhn-Tucker (KKT) conditions for this problem are given by

$$\begin{aligned} \lambda I + \mu a^2 D^T D &= G^T \left(I + G\Sigma G^T\right)^{-1} G + \Phi, \\ -\Sigma \preceq 0 &\Leftrightarrow \mathrm{tr}(\Phi\Sigma) = 0, \\ \mathrm{tr}(\Sigma) \le kP &\Leftrightarrow \lambda(\mathrm{tr}(\Sigma) - kP) = 0, \\ \mathrm{tr}(a^2 D\Sigma D^T + NDD^T) \le k\gamma P &\Leftrightarrow \mu(\mathrm{tr}(a^2 D\Sigma D^T + NDD^T) - k\gamma P) = 0, \end{aligned}$$

where $\lambda, \mu \ge 0$. Now if UHV^T is the singular value decomposition of G and FLQ^T is the singular value decomposition of D, then

$$\begin{aligned} VH^2V^T &= G^T G \\ &= \frac{1}{N} Q(I + a^2 b^2 L(I + b^2 L^2)^{-1} L)Q^T. \end{aligned}$$

Columns of matrices V and Q are eigenvectors of the symmetric matrix $G^T G$. Hence $V^T Q$ and $Q^T V$ are permutation matrices. As a result, $V^T D^T DV = V^T Q L^2 Q^T V = L^2$ is a diagonal matrix and therefore, $D^T D = V\Lambda V^T$ where $\Lambda = L^2$. Therefore, the first KKT condition can be simplified to

$$\lambda I + \mu a^2 \Lambda = H\left(I + HV^T \Sigma V H\right)^{-1} H + V^T \Phi V.$$

Since the KKT conditions are necessary and sufficient for optimality, it follows that Σ must have the same set of eigenvectors V, i.e., $\Sigma = V\Psi V^T$, where Ψ is a diagonal matrix. The dual matrix Φ can be chosen to have the same structure, i.e., $\Phi = V\Theta V^T$ and $\mathrm{tr}(\Sigma\Phi) = \mathrm{tr}(\Psi\Theta) = 0$, hence satisfying the KKT conditions. As a result, the expression in (21) can be simplified to

$$C_k^{\mathrm{FD-L}}(P, \gamma P) = \frac{1}{2k} \max_{\Psi, \Lambda} \log \left| I + \frac{1}{N} \Psi (I + a^2 b^2 \Lambda (I + b^2 \Lambda)^{-1}) \right|, \tag{22}$$

where Ψ and Λ are diagonal matrices. Since this expression is independent of V, we can set $V = I$. Hence, the search for the optimal matrices D and Σ can be limited to the space of diagonal matrices Ψ and Λ. In particular, if $D^T D = V\Lambda V^T$, then the diagonal matrix $L = \sqrt{\Lambda}$.

Thus, the maximization problem can be simplified to the following

$$C_k^{\mathrm{FD-L}}(P, \gamma P) = \frac{1}{2k} \max_{\substack{\psi_1, \ldots, \psi_k \\ \lambda_1, \ldots, \lambda_k}} \log \prod_{i=1}^{k} \left(1 + \frac{1}{N} \psi_i \left(1 + \frac{a^2 b^2 \lambda_i}{1 + b^2 \lambda_i}\right)\right), \tag{23}$$

subject to $\psi_i \geq 0$, $\lambda_i \geq 0$, for $i = 1, 2, \ldots, k$, $\sum_{i=1}^{k} \psi_i \leq kP$, and $\sum_{i=1}^{k} \lambda_i(a^2\psi_i + N) \leq k\gamma P$.

First it is easy to see that at the optimum point, $\sum_{i=1}^{k} \psi_i = kP$ and $\sum_{i=1}^{k} \lambda_i(a^2\psi_i + N) = k\gamma P$. Note that the objective function is an increasing function of λ_i. Therefore, if $\sum_{i=1}^{k} \lambda_i(a^2\psi_i + N) < k\gamma P$, we can always increase the value of objective function by increasing λ_j assuming that $\psi_j \neq 0$. Now we show that $\sum_{i=1}^{k} \psi_i = kP$. It is easy to verify that $\frac{\partial}{\partial \psi_j}\left(1 + \frac{1}{N}\psi_j\left(1 + \frac{a^2b^2\lambda_j}{1+b^2\lambda_j}\right)\right)$ is positive along the curve $\lambda_j(a^2\psi_j + N) = const.$. Therefore, while keeping $\sum_{i=1}^{k} \lambda_i(a^2\psi_i + N)$ fixed at $k\gamma P$, we can always increase the objective function by increasing ψ_j and hence $\sum_{i=1}^{k} \psi_i = kP$.

Note that at the optimum, if $\psi_j = 0$, then $\lambda_j = 0$. However, if $\lambda_j = 0$, the value of ψ_j is not necessarily equal to zero. Without loss of generality, assume that at the optimum, $\lambda_j = 0$ for the first $0 \leq k_0 \leq k$ indices, and that the total power assigned to the k_0 indices is given by $\sum_{j=1}^{k_0} \psi_j = \theta_0 kP$, for $0 \leq \theta_0 \leq 1$. Then by convexity

$$\log \prod_{j=1}^{k_0} \left(1 + \frac{1}{N}\psi_j\left(1 + \frac{a^2b^2\lambda_j}{1 + b^2\lambda_j}\right)\right) = \log \prod_{j=1}^{k_0} \left(1 + \frac{1}{N}\psi_j\right) \leq k_0 \log\left(1 + \frac{\theta_0 kP}{k_0 N}\right),$$

where the upper bound can be achieved by redistributing the power as $\psi_j = \frac{\theta_0 kP}{k_0}$, for $1 \leq j \leq k_0$. In Appendix E, we show that at the optimum, there are no more than four (ψ_j, λ_j) distinct pairs such that $\psi_j > 0$ and $\lambda_j > 0$. Including the case where $\lambda_j = 0$, we therefore conclude that there are at most five distinct (ψ_j, λ_j) pairs for any $k \geq 5$. Thus, in general, the capacity can be expressed as

$$C_k^{\mathrm{FD-L}}(P, \gamma P) = \frac{1}{2k} \max_{\mathbf{k},\, \boldsymbol{\theta},\, \boldsymbol{\eta}} \log\left[\left(1 + \frac{\theta_0 kP}{k_0 N}\right)^{k_0} \prod_{j=1}^{4}\left(1 + \frac{\theta_j kP}{k_j N}\left(1 + \frac{a^2b^2\eta_j}{1 + b^2\eta_j}\right)\right)^{k_j}\right], \quad (24)$$

where $\mathbf{k} = [k_0,\ k_1, \ldots, k_4]$, $\boldsymbol{\theta} = [\theta_0,\ \theta_1, \ldots, \theta_4]$, $\boldsymbol{\eta} = [\eta_1,\ \eta_2, \ldots, \eta_4]$, subject to $\theta_j \geq 0$, $\eta_j > 0$, $\sum_{j=0}^{4} k_j = k$, $\sum_{j=0}^{4} \theta_j = 1$, and $\sum_{j=1}^{4} k_j\eta_j\left(a^2\frac{k\theta_j P}{k_j} + N\right) = k\gamma P$.

To find $C^{\mathrm{FD-L}}(P, \gamma P)$ we need to find $\lim_{k\to\infty} C_k^{\mathrm{FD-L}}(P, \gamma P)$. Taking the limit of the above expression as $k \to \infty$, we obtain

$$C^{\mathrm{FD-L}}(P, \gamma P) = \max_{\boldsymbol{\alpha},\, \boldsymbol{\theta},\, \boldsymbol{\eta}} \alpha_0 \mathcal{C}\left(\frac{\theta_0 P}{\alpha_0 N}\right) + \sum_{j=1}^{4} \alpha_j \mathcal{C}\left(\frac{\theta_j P}{\alpha_j N}\left(1 + \frac{a^2b^2\eta_j}{1 + b^2\eta_j}\right)\right), \quad (25)$$

where $\boldsymbol{\alpha} = [\alpha_0,\ \alpha_1, \ldots, \alpha_4]$, subject to $\alpha_j, \theta_j \geq 0$, $\eta_j > 0$, $\sum_{j=0}^{4} \alpha_j = \sum_{j=0}^{4} \theta_j = 1$, and $\sum_{j=1}^{4} \eta_j\left(a^2\theta_j P + N\alpha_j\right) = \gamma P$. This completes the proof of the theorem. ■

We have shown that the capacity with linear relaying can be computed by solving a single-letter cosntrained optimization problem. The optimization problem, however, is non-convex and involves 14 variables (or 11 if we use the three equality constraints). Finding the solution for this optimization problem in general requires exhaustive search, which is computationally extensive for 11 variables. Noting that the problem is convex in each set of variables $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ if we fix the other two, the following fast Monte Carlo algorithm can be used to find a good lower bound to the solution of the problem. Randomly choose initial values for the three sets of variables, fix two of them and optimize over the third set. This process is continued by cycling through the variables sets, until the rate converges to a local maximum. The process is repeated many times for randomly chosen initial points and local maximas are found.

Figure 7 compares the lower bound on the capacity of the FD-AWGN relay channel with linear relaying to the max-flow min-cut upper bound and the generalized block Markov and side-information coding lower bound. Note that when a is small, the capacity with linear relaying becomes very close to the upper bound. Further, as $\sqrt{\gamma}b \to \infty$, the capacity with linear relaying becomes tight. On the other hand, as $a \to \infty$, the generalized block Markov lower bound becomes tight. Note that if $a^2 \geq 1 + b^2\gamma\left(1 + \frac{P}{N}\right)$, the capacity is given by the block Markov lower bound. This is similar to the result reported in [17].

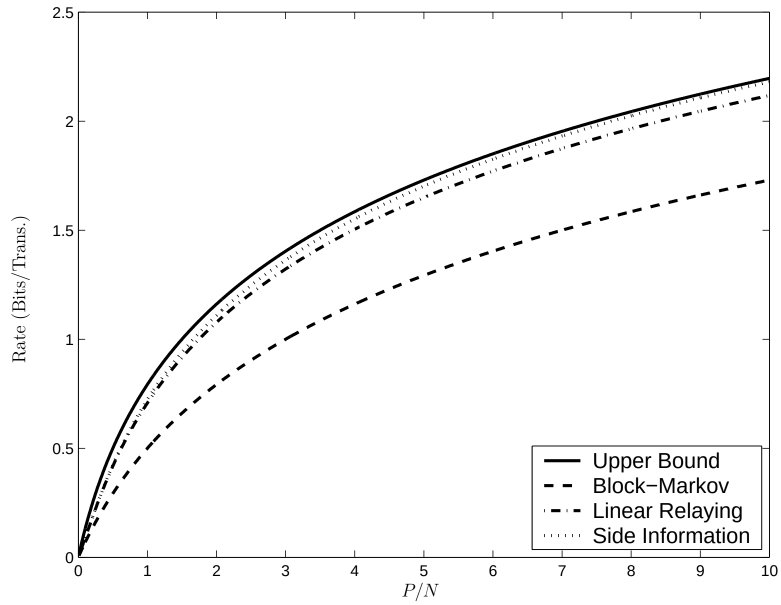
5.1 Upper bound on minimum energy-per-bit

Linear relaying can improve the upper bound on the minimum energy-per-bit for the general AWGN relay model. To demonstrate this, consider the achievable rate by the scheme in Example 1. It can be shown that the rate function (15) is convex for small P and therefore, as in Example 2, the rate can be improved by time-sharing. The rate function with time-sharing can be used to obtain an upper bound on the minimum energy-per-bit. Figure 8 plots the ratio of the best upper to lower bounds on the minimum energy-per-bit. Note that the simple scheme in Example 1 with time-sharing reduces the maximum ratio to 1.7. The minimum energy-per-bit using the linear relaying is usually lower than the minimum energy-per-bit using the side-information scheme for the general AWGN relay channel.

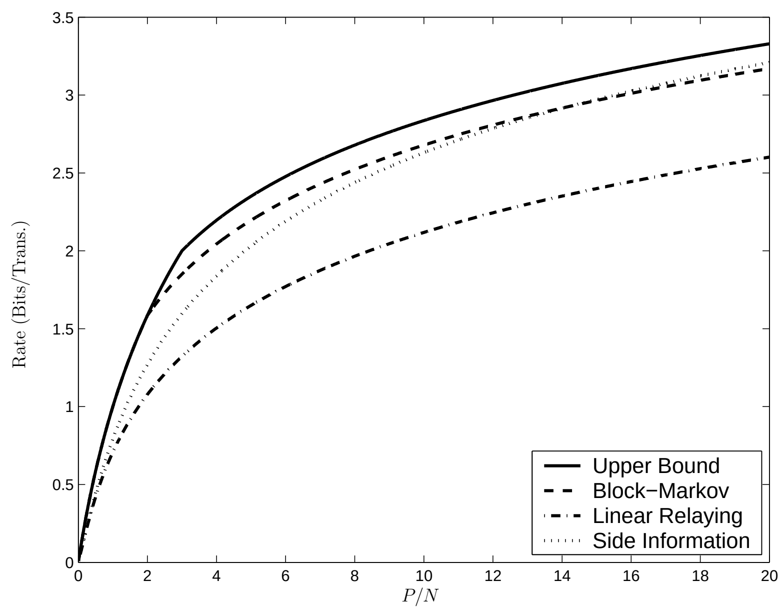
Using the relationship between the minimum energy-per-bit and capacity in Theorem 1 and the capacity with linear relaying for the FD-AWGN model established in Theorem 3, we can readily obtain an upper bound on minimum energy-per-bit of the FD-AWGN relay channel. Linear relaying improves the ratio for large b and $a \approx 1$. The maximum ratio using the linear relaying can be reduced to 1.87. The maximum ratio can be further reduced by using the side information coding scheme to 1.45. Figure 9 plots the ratio of the best upper bound on the minimum energy-per-bit to the lower bound in Table 2 for the FD-AWGN relay channel.

6 Conclusion

The paper establishes upper and lower bounds on the capacity and minimum energy-per-bit for general and FD-AWGN relay channel models. The max-flow min-cut upper bound and the generalized block Markov lower bound on capacity of the relay channel are first used to derive corresponding upper and lower bounds on capacity. These bounds are never tight for the general AWGN model and are tight only under certain conditions for the FD-AWGN



(a) $a = 1$, and $\sqrt{\gamma}b = 2$



(b) $a = 2$, and $\sqrt{\gamma}b = 1$

Figure 7: Comparison of achievable rates based on the generalized block Markov, linear relaying, and side information encoding schemes for the FD-AWGN relay channel.

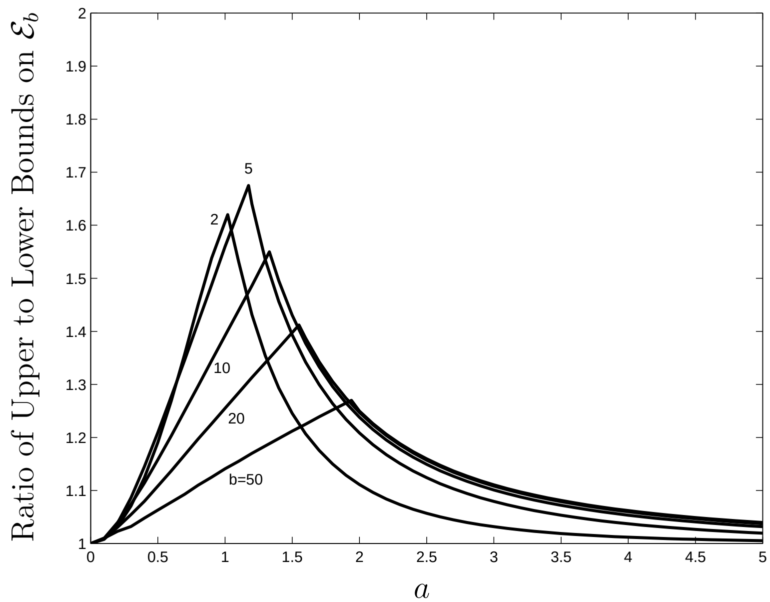


Figure 8: Ratio of the best upper bound to the lower bound of Theorem 2 for various values of a and b for general AWGN relay channel.

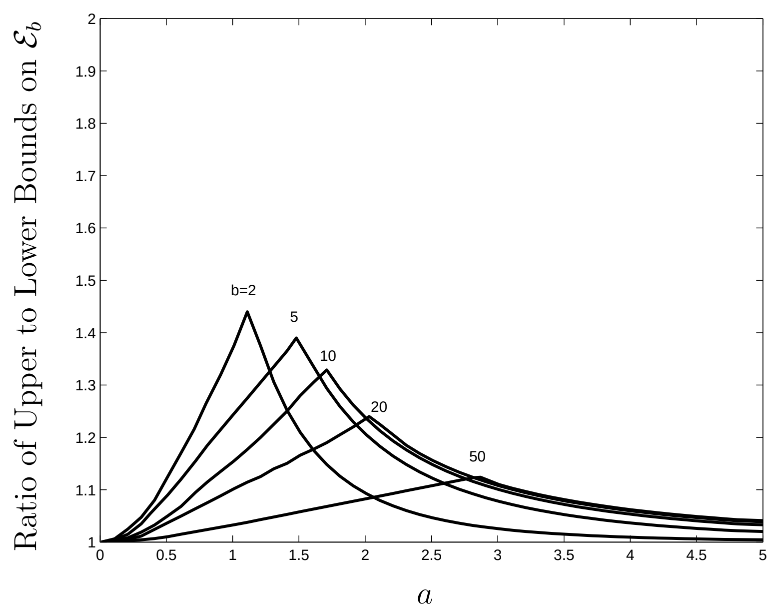


Figure 9: Ratio of the best upper bound to the lower bound of Theorem 2 for various values of a and b for FD-AWGN relay channel.

model. The gap between the upper and lower bounds is largest when the gain of the channel to the relay is comparable or worse than that of the direct channel. We argue that the reason for this large gap is that in the generalized block Markov scheme, the relay either decodes the entire message or it is not used at all. When $a \approx 1$ or less than 1, this restricts the achievable rate to be close to the capacity of the direct channel. To obtain tighter lower bounds for this case, two coding schemes are investigated where the relay cooperates with the sender but without decoding any part of the message. First the side information coding scheme is shown to outperform the block Markov coding when the gain of the channel to the relay is comparable to that of the direct channel. We show that the achievable rate can be improved via time-sharing and provide a general expression for the achievable rate using side information coding for relay channels in general. In the second scheme, the relaying functions are restricted to be linear. For the general AWGN model, a simple linear-relaying scheme is shown to significantly outperform the more sophisticated generalized block Markov and side information schemes in some cases. It is shown that the capacity with linear relaying can be found by solving a sequence of non-convex optimization problems. One of our main results in the paper is reducing this formulation to a "single-letter" expression for the FD-AWGN model. Figures 3 and 7 compare the rates for the different schemes.

The paper also established a general relationship between the minimum energy-per-bit and the capacity of the AWGN relay channel. This relationship together with the lower and upper bounds on capacity are used to establish corresponding lower and upper bounds on the minimum energy-per-bit for the general and FD-AWGN relay channels. The bounds are very close and do not differ by more than a factor of 1.45 for the FD-AWGN relay channel model and by 1.7 for the general AWGN model.

Two open problems are suggested by the work in this paper. The first is to find the distribution on $(Q, X, X_1, \hat{Y})$ that optimizes the achievable rate using side information coding given in Theorem 2. Our bounds are obtained with the assumption that $(X, X_1, \hat{Y})$ is Gaussian and with specific choices of the time-sharing random variable Q. The second open problem is finding a "single-letter" characterization of the capacity with linear relaying for the general AWGN relay model. We have been able to find such characterization only for the FD-AWGN case.

In conclusion, the upper and lower bounds for the capacity and minimum energy-per-bit established in this paper are still not tight for the general AWGN relay model and are only tight under certain conditions for the FD-AWGN relay channel. Establishing capacity and the minimum energy-per-bit is likely to require a combination of new coding schemes and a tighter upper bound than the max-flow min-cut bound.

References

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

[2] H. Sato, "Information transmission through a channel with relay," The Aloha System, University of Hawaii, Honolulu, Tech. Rep B76-7, March 1976.

[3] A. El Gamal, *Results in Multiple User Channel Capacity*, PhD. Thesis, Stanford University, May 1978.

[4] T. M. Cover and A. El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Transactions on Information Theory*, Vol. 25, No. 5, pp. 572-584, September 1979.

[5] A. El Gamal and M. Aref, "The Capacity of the Semi-Deterministic Relay Channel," *IEEE Transactions on Information Theory,* Vol. IT-28, No. 3, pp. 536, May 1982.

[6] A. El Gamal, "On Information Flow in Relay Networks," *IEEE National Telecommunications Conference,* Vol. 2, pp. D4.1.1 - D4.1.4, November 1981.

[7] M. Aref, *Information flow in relay networks*, PhD. Thesis, Stanford University, October 1980.

[8] B. Schein and R. G. Gallager, "The Gaussian parallel relay network," *Proc IEEE Int Symp Info Theory*, Sorrento, Italy, p. 22, July 2000.

[9] A. El Gamal and S. Zahedi, "Capacity of a Class of Relay Channels with Orthogonal Components," *IEEE Transactions on Information Theory*, Vol. IT-51, No. 5, pp. 1815-1817, May 2005.

[10] S. Zahedi, M. Mohseni and A. El Gamal, "On the Capacity of AWGN Relay Channels with Linear Relaying Functions," *Proc. International Symposium on Information Theory*, June 27-July 2, 2004, Chicago, IL, P. 399.

[11] P. Gupta and P.R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, Vol. 46, pp. 388-404, March 2000.

[12] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *Proc IEEE Int Symp Info Theory*, Washington DC, p. 159, June 2001.

[13] M. Gastpar and M. Vetterli, "On The Capacity of Wireless Networks: the relay case," *Proc IEEE INFOCOM 2002*, New York city, NY, vol. 3, pp. 1577-1586, June 2002.

[14] M. Gastpar and M. Vetterli, "On The Asymptotic Capacity of Gaussian Relay Channels," *Proc IEEE Int Symp Info Theory*, Lausanne, Switzerland, June 30-July 5 2002.

[15] G. Kramer, M. Gastpar and P. Gupta," Cooperative strategies and capacity theorems for relay networks," to appear in *IEEE Transactions on Information Theory*.

[16] V. Rodoplu and T.H. Meng, "Minimum energy mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, pp. 1333 -1344, August 1999.

[17] Y. Liang and V.V. Veeravalli, "Gaussian frequency division relay channels: optimal bandwidth allocation and capacity," In *Proc CISS*, Princeton, NJ, March 2004.

[18] A. El Gamal and S. Zahedi, "Minimum Energy Communication Over a Relay Channel," *Proc. International Symposium on Information Theory*, June 29-July 4, 2003, Yokohama, Japan, P. 344.

[19] R.J. McEliece, *The Theory of Information and Coding*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.

[20] R. G. Gallager, "Energy Limited Channels:Coding, Multiaccess, and Spread Spectrum," Tech. Report LIDS-P-1714, LIDS, MIT, Cambridge, Mass., November 1987.

[21] R. G. Gallager, "Energy limited channels, coding, multiaccess and spread spectrum," *Proc. Conf. Inform. Sci. Syst.*, p. 372, Princeton, NJ, Mar. 1988.

[22] S. Verdu, "On Channel Capacity per Unit Cost," *IEEE Transactions on Information Theory*, Vol. 36, pp. 1019-1030, September 1990.

[23] S. Verdu, G. Caire and D. Tuninetti, "Is TDMA optimal in the low power regime?" *Proc IEEE Int Symp Info Theory*, Lausanne, Switzerland, p. 193, July 2002.

[24] A. Hst-Madsen and J. Zhang: "Capacity Bounds and Power Allocation for Wireless Relay Channel," submitted to *IEEE Trans. Information theory*.

[25] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, Vol. 48, pp. 1319 -1343, Jun 2002.

[26] T. M. Cover and J. A. Thomas, "Elements of Information Theory," Wiley, New York, 1991.

[27] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, 2004.

Appendix A

In this appendix, we evaluate upper bounds on the capacity of the general and FD-AWGN relay channels. The max-flow min-cut bound gives the

$$C \leq \max_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y, Y_1|X_1)\}.$$

First consider the general AWGN relay channel. To prove the upper bound we begin with the first bound. Using standard arguments, it can be easily shown that

$$\begin{aligned}
I(X, X_1; Y) &= h(Y) - h(Y|X, X_1) \\
&= h(Y) - \frac{1}{2}\log(2\pi e N) \\
&\leq \frac{1}{2}\log\frac{\mathrm{Var}(Y)}{N} \\
&\leq \frac{1}{2}\log\left(1 + \frac{\mathrm{Var}(X) + b^2\mathrm{Var}(X_1) + 2bE(XX_1)}{N}\right) \\
&\leq \frac{1}{2}\log\left(1 + \frac{(1 + b^2\gamma + 2b\rho\sqrt{\gamma})P}{N}\right),
\end{aligned}$$

where we define

$$\rho = \frac{E(XX_1)}{\sqrt{E(X^2)E(X_1^2)}}.$$

Now, consider the second bound

$$\begin{aligned}
I(X; Y, Y_1|X_1) &= h(Y, Y_1|X_1) - h(Y, Y_1|X, X_1) \\
&\leq h(Y, Y_1|X_1) - \log 2\pi e N \\
&= h(Y|X_1) + h(Y_1|Y, X_1) - \log 2\pi e N \\
&\leq \frac{1}{2}\log 2\pi e E\mathrm{Var}(Y|X_1) + \frac{1}{2}\log 2\pi e E\mathrm{Var}(Y_1|Y, X_{1i}) - \log 2\pi e N \\
&\leq \frac{1}{2}\log\left(2\pi e\left(E(X^2) - \frac{(E(XX_1))^2}{E(X_1^2)} + N\right)\right) \\
&\quad + \frac{1}{2}\log\left(2\pi e\left(\frac{(1+a^2)\left(E(X^2) - \frac{(E(XX_1))^2}{E(X_1^2)}\right)N + N^2}{\left(E(X^2) - \frac{(E(XX_1))^2}{E(X_1^2)}\right) + N}\right)\right) - \log 2\pi e N \\
&= \frac{1}{2}\log\left((1+a^2)\left(E(X^2) - \frac{(E(XX_1))^2}{E(X_1^2)}\right)N + N^2\right) \\
&\leq \frac{1}{2}\log\left(1 + \frac{(1+a^2)(1-\rho^2)P}{N}\right).
\end{aligned}$$

Therefore, the upper bound can be expressed as

$$C \leq \max_{0\leq\rho\leq 1} \min\left\{\mathcal{C}\left(\frac{(1+a^2)(1-\rho^2)P}{N}\right), \mathcal{C}\left(\frac{(1+b^2\gamma+2b\rho\sqrt{\gamma})P}{N}\right)\right\}.$$

Performing the maximization over ρ, we can easily obtain the upper bound given in Table 1.

Now consider the bound for the FD-AWGN relay channel. Substituting Y by $\{Y_D, Y_R\}$ in (1) yields the max-flow min-cut upper bound on capacity of the FD-AWGN channel. Note that

$$\begin{aligned}
I(X;Y_1,Y_D,Y_R|X_1) &= I(X;Y_1|X_1) + I(X;Y_D|X_1,Y_1) + I(X;Y_R|X_1,Y_1,Y_D) \\
&= I(X;Y_1|X_1) + I(X;Y_D|X_1,Y_1) \\
&= h(Y_1|X_1) - h(Y_1|X,X_1) + h(Y_D|X_1,Y_1) - h(Y_D|X,X_1,Y_1) \\
&\leq h(Y_1|X_1) + h(Y_D|X_1,Y_1) - \log 2\pi e N \\
&= h(Y_1|X_1) + h(Y_D|Y_1) - \log 2\pi e N \\
&\leq \frac{1}{2}\log 2\pi e \mathrm{Var}(Y_1|X_1) + \frac{1}{2}\log 2\pi e \mathrm{Var}(Y_D|Y_1) - \log 2\pi e N \\
&= \mathcal{C}\left(\frac{a^2P(1-\rho^2)}{N}\right) + \mathcal{C}\left(\frac{P}{a^2P+N}\right).
\end{aligned}$$

Similarly, it can be shown that

$$\begin{aligned}
I(X,X_1;Y_D,Y_R) &= I(X;Y_D,Y_R) + I(X_1;Y_D,Y_R|X) \\
&= I(X;Y_D) + I(X;Y_R|Y_D) + I(X_1;Y_R|X) + I(X_1;Y_D|X,Y_R) \\
&= I(X;Y_D) + I(X;Y_R|Y_D) + I(X_1;Y_R|X) \\
&\leq \mathcal{C}\left(\frac{P}{N}\right) + \mathcal{C}\left(\frac{b^2\gamma\rho^2NP}{(b^2\gamma P(1-\rho^2)+N)(P+N)}\right) + \mathcal{C}\left(\frac{b^2\gamma P(1-\rho^2)}{N}\right).
\end{aligned}$$

Again both terms are maximized for $\rho = 0$. As a result the following upper bound on capacity can be established

$$C \leq \min\left\{\mathcal{C}\left(\frac{P}{N}\right) + \mathcal{C}\left(\frac{b^2\gamma P}{N}\right), \mathcal{C}\left(\frac{(1+a^2)P}{N}\right)\right\}.$$

Upper and lower bounds in Table 1 can be readily established.

Appendix B

We show that the lower bound on the capacity of the discrete-memoryless relay channel of Theorem 6 in [4] can be recast into the bound

$$C \geq \max_{p(x)p(x_1)p(\hat{y}_1|y_1,x_1)} \min \left\{ I(X;Y,\hat{Y}_1|X_1), I(X,X_1;Y) - I(Y_1;\hat{Y}_1|X,X_1) \right\}. \tag{26}$$

Achievability of any rate $R \leq I(X;Y,\hat{Y}_1|X_1)$ subject to the constraint $I(X_1;Y) \geq I(Y_1;\hat{Y}_1|X_1,Y)$, for any distribution $p(x,x_1,y,y_1,\hat{y}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|y_1,x_1)$ was proved in [4]. We show the converse, i.e., any rate satisfying the original conditions $R \leq I(X;Y,\hat{Y}_1|X_1)$ and $I(X_1;Y) \geq I(Y_1;\hat{Y}_1|X_1,Y)$, also satisfies (26). Consider

$$\begin{aligned}
R &= I(X;Y,\hat{Y}_1|X_1) \\
&= H(Y,\hat{Y}_1|X_1) - H(Y,\hat{Y}_1|X,X_1) \\
&= H(Y|X_1) + H(\hat{Y}_1|X_1,Y) - H(\hat{Y}_1|X,X_1) - H(Y|X,X_1,\hat{Y}_1) \\
&= H(Y|X_1) + H(\hat{Y}_1|X_1,Y) - H(\hat{Y}_1|X,X_1) - H(Y|X,X_1) \\
&= H(Y) - I(X_1;Y) + I(Y_1;\hat{Y}_1|X_1,Y) + H(\hat{Y}_1|X_1,Y,Y_1) - H(\hat{Y}_1|X,X_1) - H(Y|X,X_1) \\
&\leq H(Y) + H(\hat{Y}_1|X_1,Y,Y_1) - H(\hat{Y}_1|X,X_1) - H(Y|X,X_1) \\
&= H(Y) + H(\hat{Y}_1|X_1,Y_1) - H(\hat{Y}_1|X,X_1) - H(Y|X,X_1) \\
&= H(Y) + H(\hat{Y}_1|X,X_1,Y_1) - H(\hat{Y}_1|X,X_1) - H(Y|X,X_1) \\
&= I(X,X_1;Y) - I(Y_1;\hat{Y}_1|X,X_1).
\end{aligned}$$

It is not difficult to show the achievability, i.e., that any rate satisfying the above inequality also satisfies the original conditions.

Appendix C

Achievability of any rate

$$R \le \max_{p(q)p(x|q)p(x_1|q)p(\hat{y}_1|y_1,x_1,q)} \min \left\{ I(X;Y_D,Y_R,\hat{Y}_1|X_1,Q), I(X,X_1;Y_D,Y_R|Q) - I(Y_1;\hat{Y}_1|X,X_1,Q) \right\},$$

was shown in section 4. We now evaluate the mutual information terms for the AWGN relay channel. The optimal choice of probability mass functions are not known. We assume the random variable Q has cardinality 2 and takes values in $\{0,1\}$. We further assume $\mathrm{P}\{Q=1\}=\alpha$. Consider X as a Gaussian random variable with variance $\frac{P}{\alpha}$ if $Q=1$ and zero otherwise. Furthermore, assume X_1 is a Gaussian random variable with variance γP irrespective of the value of random variable Q and independent of X. Define the random variable $\hat{Y}_1 = 0$ if $Q=0$, and $\hat{Y}_1 = \theta(Y_1+Z')$ if $Q=1$, where θ is a constant and $Z' \sim \mathcal{N}(0,N')$, is independent of Q, X, X_1, Z, and Z_1.

Now consider

$$\begin{aligned}
h(\hat{Y}_1|X_1,Y_D,Q=1) &= E_{X_1,Y_D}\left(\frac{1}{2}\log 2\pi e \mathrm{Var}(\hat{Y}_1|y_D,x_1)\right) \\
&= \frac{1}{2}\log 2\pi e\theta^2\left(N+N'+\frac{a^2PN}{P+\alpha N}\right), \\
h(\hat{Y}_1|Y_1,Q=1) &= E_{Y_1}\left(\frac{1}{2}\log 2\pi e \mathrm{Var}(\hat{Y}_1|y_1)\right) \\
&= \frac{1}{2}\log 2\pi e\theta^2 N', \\
h(\hat{Y}_1|X,X_1,Q=1) &= E_{X,X_1}\left(\frac{1}{2}\log 2\pi e \mathrm{Var}(\hat{Y}_1|x,x_1)\right) \\
&= \frac{1}{2}\log 2\pi e\theta^2 (N+N').
\end{aligned}$$

Using the above results we can easily show that

$$\begin{aligned}
I(X;Y_D,Y_R,\hat{Y}_1|X_1,Q) &= I(X;Y_D|X_1,Q) + I(X;\hat{Y}_1|X_1,Y_D,Q) + I(X;Y_R|X_1,\hat{Y}_1,Y_D,Q) \\
&= I(X;Y_D|X_1,Q) + I(X;\hat{Y}_1|X_1,Y_D,Q) \\
&= \alpha\mathcal{C}\left(\frac{P}{\alpha N}\right) + \alpha\mathcal{C}\left(\frac{a^2PN}{(P+\alpha N)(N+N')}\right), \\
I(X,X_1;Y_D,Y_R|Q) &= I(X,X_1;Y_D|Q) + I(X,X_1;Y_R|Y_D,Q) \\
&= I(X;Y_D|Q) + I(X_1;Y_D|X,Q) + I(X_1;Y_R|Y_D,Q) + I(X;Y_R|X_1,Y_D,Q) \\
&= I(X;Y_D|Q) + I(X_1;Y_R|Y_D,Q) \\
&= I(X;Y_D|Q) + I(X_1;Y_R|Q) \\
&= \alpha\mathcal{C}\left(\frac{P}{\alpha N}\right) + \mathcal{C}\left(\frac{b^2\gamma P}{N}\right), \\
I(Y_1;\hat{Y}_1|X,X_1,Q) &= \alpha\mathcal{C}\left(\frac{N}{N'}\right).
\end{aligned}$$

Combining the above results, it can be shown that any rate

$$R \leq \max_{0<\alpha\leq 1, N'\geq 0} \alpha\mathcal{C}\left(\frac{P}{\alpha N}\right) + \min\left\{\mathcal{C}\left(\frac{b^2\gamma P}{N}\right) - \alpha\mathcal{C}\left(\frac{N}{N'}\right), \alpha\mathcal{C}\left(\frac{a^2PN}{(P+\alpha N)(N+N')}\right)\right\},$$

is achievable. It can be shown that the value of N' that maximizes the above expression is given by,

$$N' = \frac{N + \frac{a^2PN}{P+\alpha N}}{-1 + \left(1 + \frac{b^2\gamma P}{N}\right)^{\frac{1}{\alpha}}}.$$

Replacing N' and simplifying the lower bound expression, we obtain the lower bound given in (11).

Appendix D

Lemma: The capacity of the AWGN relay channel with average power constraints satisfies the following:

(i) $C(P, \gamma P) > 0$ if $P > 0$ and approaches ∞ as $P \to \infty$.

(ii) $C(P, \gamma P) \to 0$ as $P \to 0$.

(iii) $C(P, \gamma P)$ is concave and strictly increasing in P.

(iv) $\frac{(1+\gamma)P}{C(P,\gamma P)}$ is non-decreasing in P, for all $P > 0$.

Proof:

(i) This follows from the fact that $\mathcal{C}(P/N)$, which is less than or equal to $C(P, \gamma P)$, is strictly greater than zero for $P > 0$, and approaches infinity as $P \to \infty$.

(ii) This follows from the fact that $\hat{C}(P, \gamma P)$ in table 1 which is greater than or equal to $C(P, \gamma P)$ approaches zero as $P \to 0$.

(iii) Concavity follows by the following "time-sharing" argument. For any $P, P' > 0$ and $\epsilon > 0$, there exists k and k' such that $C(P, \gamma P) < C_k(P, \gamma P) + \epsilon$ and $C(P', \gamma P') < C_{k'}(P', \gamma P') + \epsilon$. Now, for any $\alpha = nk/(nk + mk')$, where n and m are integers,

$$\begin{aligned} \alpha C(P, \gamma P) + (1-\alpha) C(P', \gamma P') &\leq \alpha C_k(P, \gamma P) + (1-\alpha) C_{k'}(P', \gamma P') + \epsilon \\ &\leq C_{nk+mk'}\left(\alpha P + (1-\alpha)P', \gamma(\alpha P + (1-\alpha)P')\right) + \epsilon \\ &\leq C\left(\alpha P + (1-\alpha)P', \gamma(\alpha P + (1-\alpha)P')\right) + \epsilon, \end{aligned}$$

where the second inequality follows from the fact that $\alpha C_k(P, \gamma P) + (1-\alpha) C_{k'}(P', \gamma P')$ is achieved arbitrarily closely for some $\mathrm{P}_{\mathbf{X}^{nk+mk'}}$ and $\{f_i\}_{i=1}^{nk+mk'}$ that is a mixture of the $\mathrm{P}_{\mathbf{X}^k}$ and $\{g_i\}_{i=1}^{k}$ that achieve $C_k(P, \gamma P)$ and the $\mathrm{P}_{\mathbf{X}^{k'}}$ and $\{h_i\}_{i=1}^{k'}$ that achieve $C_{k'}(P', \gamma P')$ (corresponding to "time-sharing"). Clearly, the set of "time-sharing" distributions is a subset of the set of all possible $\mathrm{P}_{\mathbf{X}^{nk+mk'}}$ and $\{f_i\}_{i=1}^{nk+mk'}$. Note here that even though $C_k(P, \gamma P)$ may not be concave, $C(P, \gamma P) = \sup_k C_k(P, \gamma P)$ is concave.

That $C(P, \gamma P)$ is strictly monotonically increasing in P follows from parts (i), (ii), and concavity.

(iv) For any $\gamma \geq 0$, and $0 < P_1 < P_2$, it follows from the concavity of $C(P, \gamma P)$ that

$$\frac{P_1}{P_2} C(P_2, \gamma P_2) + \frac{P_2 - P_1}{P_2} C(0, 0) \leq C(P_1, \gamma P_1).$$

But since $C(0, 0) = 0$, this implies that $\frac{P_1}{P_2} C(P_2, \gamma P_2) \leq C(P_1, \gamma P_1)$, or $\frac{P_1}{C(P_1, \gamma P_1)} \leq \frac{P_2}{C(P_2, \gamma P_2)}$. Thus for any $\gamma \geq 0$, $\frac{P(1+\gamma)}{C(P, \gamma P)}$ is non-decreasing in P.

Appendix E

In this appendix, we prove that $C_k^{\mathrm{FD-L}}(P,\gamma P)$ can be expressed as

$$C_k^{\mathrm{FD-L}}(P,\gamma P) = \frac{1}{2k} \max_{\substack{k_0,k_1,k_2,k_3,k_4\\ \theta_0,\theta_1,\theta_2,\theta_3,\theta_4\\ \eta_1,\eta_2,\eta_3,\eta_4}} \log\left(1+\frac{k\theta_0 P}{k_0 N}\right)^{k_0} \prod_{j=1}^{4}\left(1+\frac{k\theta_j P}{k_j N}\left(1+\frac{a^2b^2\eta_j}{1+b^2\eta_j}\right)\right)^{k_j}.$$

The starting point is the expression

$$C_k^{\mathrm{FD-L}}(P,\gamma P) = \frac{1}{2k} \max_{\substack{k_0,\theta_0\\ \psi_{k_0+1},\dots,\psi_k\\ \lambda_{k_0+1},\dots,\lambda_k}} \log\left(1+\frac{k\theta_0 P}{k_0 N}\right)^{k_0} \prod_{i=k_0+1}^{k}\left(1+\frac{1}{N}\psi_i\left(1+\frac{a^2b^2\lambda_i}{1+b^2\lambda_i}\right)\right),$$

subject to $\psi_i > 0$, $\lambda_i > 0$, for $i = k_0+1,\dots,k$, $\sum_{i=k_0+1}^{k}\psi_i = k(1-\theta_0)P$, and $\sum_{i=k_0+1}^{k}\lambda_i(a^2\psi_i + N) = k\gamma P$.

For a given θ_0 and k_0, this optimization problem is equivalent to finding the maximum of

$$\log\prod_{i=k_0+1}^{k}\left(1+\frac{1}{N}\psi_i\left(1+\frac{a^2b^2\lambda_i}{1+b^2\lambda_i}\right)\right),$$

subject to $\psi_i > 0$, $\lambda_i > 0$, for $i = k_0+1,\dots,k$, $\sum_{i=k_0+1}^{k}\psi_i = k(1-\theta_0)P$, and $\sum_{i=k_0+1}^{k}\lambda_i(a^2\psi_i + N) = k\gamma P$.

To find the optimality condition for this problem, we form the Lagrangian

$$\mathcal{L}(\psi_{k_0+1}^k,\lambda_{k_0+1}^k,\alpha,\beta) = \sum_{i=k_0+1}^{k}\log\left(1+\frac{1}{N}\psi_i\left(1+\frac{a^2b^2\lambda_i}{1+b^2\lambda_i}\right)\right)+\alpha\sum_{i=k_0+1}^{k}\psi_i+\beta\sum_{i=k_0+1}^{k}\lambda_i(a^2\psi_i+N)$$

where α and β are Lagrange multipliers for the two equality constraints (the Lagrange multipliers for the inequality constraints are all equal to zero at the optimum, since by assumption $\psi_i > 0$ and $\lambda_i > 0$ for all $i > k_0$).

At the optimum, we must have

$$\frac{\partial\mathcal{L}}{\partial\psi_i} = 0, \text{ and } \frac{\partial\mathcal{L}}{\partial\lambda_i} = 0, \text{ for all } i = k_0+1,\dots,k.$$

Computing the derivatives, we obtain the conditions

$$\frac{1+(1+a^2)b^2\lambda_i)}{1+b^2\lambda_i+\psi_i(1+b^2\lambda_i(1+a^2))}+\alpha+a^2\beta\lambda_i = 0,$$

and

$$\frac{a^2b^2\psi_i}{(1+b^2\lambda_i)(1+b^2\lambda_i+\psi_i(1+b^2\lambda_i(1+a^2)))}+\beta(a^2\psi_i+N) = 0.$$

Solving this set of equations, we obtain

$$\psi_i = \frac{\beta N(1+b^2\lambda_i)(1+(1+a^2)b^2\lambda_i))}{a^2b^2\alpha - a^2\beta - 2a^2b^2\beta\lambda_i - (1+a^2)a^2b^4\beta\lambda_i^2},$$

where λ_is are the positive roots of the fourth order polynomial equation

$$C_4Z^4 + C_3Z^3 + C_2Z^2 + C_1Z + C_0 = 0,$$

with coefficients

$$\begin{aligned}
C_0 &= a^2b^6\beta^2(1+a^2)(N(1+a^2)-a^2) \\
C_1 &= (1+a^2)b^4\beta((1+a^2)(\alpha Nb^2 - a^2b^2 + a^2\beta N) - \alpha a^2b^2 + 2a^2\beta N - a^4\beta) - 2a^4b^4\beta^2 \\
C_2 &= (1+a^2)b^2\beta(-3a^2b^2 + 3\alpha Nb^2 + \alpha Na^2b^2 + 2a^2\beta N) + a^2\beta^2b^2N - 3a^2\beta^2b^2(a^2+\alpha) \\
C_3 &= (1+a^2)b^2(\alpha a^2b^2 + 2\alpha\beta N - a^2\beta) + a^2b^2(-2\beta + \alpha\beta(a^2-3)\alpha^2b^2) + \alpha\beta Nb^2 - a^4\beta^2 \\
C_4 &= a^2(1+\alpha)(b^2\alpha - \beta) + \alpha\beta N.
\end{aligned}$$

This polynomial equation has at most four distinct roots for any given channel gain coefficients a and b. Denote the roots by $\eta_1, \eta_2, \eta_3, \eta_4$. Substituting in the optimality conditions, we obtain at most four distinct values of ψ, which we denote by by $\psi_1, \psi_2, \psi_3, \psi_4$. Note that only pairs such that $\psi_j > 0$ and $\eta_j > 0$ are feasible.