Capacity Bounds for Diamond Networks with an Orthogonal Broadcast Channel

Shirin Saeedi Bidokhti, *Member, IEEE,* and Gerhard Kramer, *Fellow, IEEE,*

Abstract

A class of diamond networks is studied where the broadcast component is orthogonal and modeled by two independent bit-pipes. New upper and lower bounds on the capacity are derived. The proof technique for the upper bound generalizes bounding techniques of Ozarow for the Gaussian multiple description problem (1981) and Kang and Liu for the Gaussian diamond network (2011). The lower bound is based on Marton's coding technique and superposition coding. The bounds are evaluated for Gaussian and binary adder multiple access channels (MACs). For Gaussian MACs, both the lower and upper bounds strengthen the Kang-Liu bounds and establish capacity for interesting ranges of bit-pipe capacities. For binary adder MACs, the capacity is established for all ranges of bit-pipe capacities.

I. Introduction

The diamond network [1] is a two-hop network that is a cascade of a broadcast channel (BC) and a multiple access channel (MAC). The two-relay diamond network has a source communicate with a sink through two relay nodes that do not have information of their own to communicate. The underlying challenge may be described as follows. In order to fully utilize the MAC to the receiver, we would like to achieve full cooperation at the relay nodes. On the other hand, to better use the diversity that is offered by the relays, we would like to send independent information to the relay nodes over the BC.

The problem of finding the capacity of this network is unresolved. Lower and upper bounds on the capacity are given in [1]. An interesting class of networks is when the BC and/or MAC are modelled via orthogonal links [2], [3], [4], [5]. The problem is solved for linear deterministic relay networks, and the capacity of Gaussian relay networks has been approximated within a constant number of bits [6]. The capacity of Gaussian diamond networks with n relays is studied in [7], [8], [9]. These works propose relaying strategies that achieve the cut-set upper bound up to an additive (or multiplicative) gap.

In this paper, we study capacity bounds when there are two relays and the BC is orthogonal, which means that the BC may as well have two independent bit-pipes. This problem was studied in [2] where lower and upper bounds were derived on the capacity. Recently, [3] studied a Gaussian MAC and derived a new upper bound that constrains the mutual information between the MAC inputs. The bounding technique in [3] is motivated by [10] that treats the Gaussian multiple description problem. Unfortunately, neither result seems to apply to discrete memoryless channels.

This paper is organized as follows. We state the problem setup in Section II. In Section III, we improve the achievable rates of [2] by communicating a common piece of information from the source to both relays using superposition coding and Marton's coding. In Section IV, we prove new capacity upper bounds by generalizing and improving the bounding technique of [3]. Our upper bounds apply to the general class of discrete memoryless MACs, and strictly improve the cut-set bound. We study the bounds for networks with a Gaussian MAC (Section V) and a binary adder MAC (Section VI). For networks with a Gaussian MAC, we find conditions on the bit-pipe capacities such that the upper and lower bounds meet. For networks with a binary adder MAC, we find the capacity for all ranges of bit-pipe capacities.

II. Preliminaries

A. Notation

Random variables are denoted by capital letters, e.g. X, and their realizations are denoted by small letters, e.g. x. The probability mass function (pmf) describing X is denoted by $p_X(x)$ or $p(x)$. The entropy of X is denoted by $H(X)$, the conditional entropy of X given Y is denoted by $H(X|Y)$, and the mutual information between X and Y is denoted by $I(X;Y)$. Differential entropies are denoted by $h(X)$ and conditional differential entropies are denoted by $h(X|Y)$. Sets are denoted by script letters and matrices are denoted by bold capital letters. The random sequence $X_1,\dots,X_n$ is denoted by X^n. $\mathcal{T}_\epsilon^n(X)$ denotes the set of sequences x^n that are $\epsilon-$ typical with respect to $P_X(.)$ [11]. When $P_X(.)$ is clear from the context we write $\mathcal{T}_\epsilon^n$.

When X is a Bernoulli random variable with $p_X(0)=q$, its entropy in bits is $h_2(q)=-q\log_2(q)-(1-q)\log_2(1-q)$. The pair of random variables (X,Y) is said to be a doubly symmetric binary source with parameter p if $p_{XY}(0,0)=p_{XY}(1,1)=\frac{1-p}{2}$, and $p_{XY}(0,1)=p_{XY}(1,0)=\frac{p}{2}$. Throughout this paper, all logarithms are to the base 2. For a real number x, we denote $\max(x,0)$ by x^+.

S. Saeedi Bidokhti and G. Kramer are with the Department for Electrical and Computer Engineering, Technische Universität München, Germany (shirin.saeedi@tum.de, gerhard.kramer@tum.de).

The authors are supported by the German Federal Ministry of Education and Research in the framework of the Alexander von Humboldt-Professorship. The work of S. Saeedi Bidokhti was also supported by the Swiss National Science Foundation Fellowship no. 146617.

The material in this paper was presented in part at the 2014 IEEE International Symposium on Information Theory, Honolulu, USA.

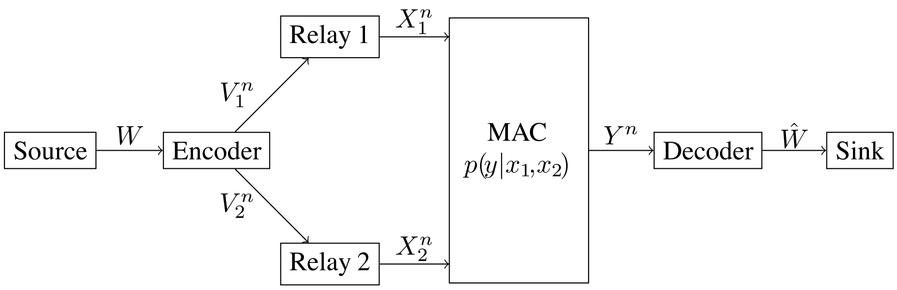


Fig. 1: Problem setup.

B. Model

Consider the diamond network in Fig. 1. A source communicates a message W with nR bits to a sink. The source encodes W into the sequence V_1^n, which is available at relay 1, and the sequence V_2^n, which is available at relay 2. V_1^n and V_2^n are such that $H(V_1^n) \leq nC_1$ and $H(V_2^n) \leq nC_2$. Each relay i, $i = 1, 2$, maps its received sequence V_i^n into a sequence X_i^n which is sent over a MAC with transition probabilities $p(y|x_1, x_2)$, for each $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, $y \in \mathcal{Y}$. From the received sequence Y^n, the sink decodes an estimate $\hat{W}$ of W.

A coding scheme consists of an encoder, two relay mappings, and a decoder, and is said to achieve the rate R if, by choosing n sufficiently large, we can make the error probability $\Pr(\hat{W} \neq W)$ as small as desired. We are interested in characterizing the largest achievable rate R. We refer to the maximum achievable rate as the capacity $C^{\diamond}$ of the network.

III. Lower bound

Our coding scheme is based on [2], but we further send a common message to both relaying nodes. This is done by rate splitting, superposition coding and Marton's coding and is summarized in the following theorem.

Theorem 1. *The rate R is achievable if it satisfies the following condition for some pmf $p(u, x_1, x_2, y) = p(u, x_1, x_2)p(y|x_1, x_2)$, and $U \in \mathcal{U}$ with $|\mathcal{U}| \leq \min\{|\mathcal{X}_1||\mathcal{X}_2| + 3, |\mathcal{Y}| + 4\}$.*

$$R \leq \min \left\{ \begin{array}{l} C_1 + C_2 - I(X_1; X_2|U), \\ C_2 + I(X_1; Y|X_2 U), \\ C_1 + I(X_2; Y|X_1 U), \\ \frac{1}{2}(C_1 + C_2 + I(X_1 X_2; Y|U) - I(X_1; X_2|U)), \\ I(X_1 X_2; Y) \end{array} \right\} \tag{1}$$

Remark 1. *If U is a constant then the fourth bound in* (1) *is redundant as it is half the sum of the first and fifth bounds. This shows that Theorem 1 without a U reduces to [2, Theorem 1]. U turns out to be useful for Gaussian MACS, as shown in Fig. 3. Theorem 1 appeared in [12, Theorem 2] and also in [13, Theorem 2].*

Remark 2. *One could add a time-sharing random variable Q to* (1)*. However, by combining Q with U, one can check that Theorem 1 is at least as large as this region.*

Sketch of proof:

a) Codebook construction: Fix the joint pmf $p(u, x_1, x_2)$ and $R_{12}, R_1, R_2, R_1', R_2' \geq 0$. Let

$$R = R_{12} + R_1 + R_2. \tag{2}$$

Generate $2^{nR_{12}}$ sequences $u^n(m_{12})$ independently, each in an i.i.d manner according to $\prod_l P_U(u_l)$. For each sequence $u^n(m_{12})$, generate (i) $2^{n(R_1+R_1')}$ sequences $x_1^n(m_{12}, m_1, m_1')$, $m_1 = 1, \ldots, 2^{nR_1}$, $m_1' = 1, \ldots, 2^{nR_1'}$, conditionally independently, each in an i.i.d manner according to $\prod_l P_{X_1|U}(x_{1,l}|u_l(m_{12}))$ and (ii) $2^{n(R_2+R_2')}$ sequences $x_2^n(m_{12}, m_2, m_2')$, $m_2 = 1, \ldots, 2^{nR_2}$, $m_2' = 1, \ldots, 2^{nR_2'}$, conditionally independently, each in an i.i.d manner according to $\prod_l P_{X_2|U}(x_{2,l}|u_l(m_{12}))$. For each bin index (m_{12}, m_1, m_2), pick a sequence pair $(x_1^n(m_{12}, m_1, m_1'), x_2^n(m_{12}, m_2, m_2'))$ that is jointly typical.

b) Encoding: To communicate message $W = (m_{12}, m_1, m_2)$, communicate (m_{12}, m_1, m_1') to relay 1 and (m_{12}, m_2, m_2') to relay 2; here $(x_1^n(m_{12}, m_1, m_1'), x_2^n(m_{12}, m_2, m_2'))$ is the jointly typical pair picked in the bin indexed by (m_{12}, m_1, m_2).

c) Decoding: Upon receiving y^n, the receiver looks for indices $\hat{m}_{12}, \hat{m}_1, \hat{m}_2$ for which the following tuple is jointly typical for some $\hat{m}_1'$, $\hat{m}_2'$:

$$(u^n(\hat{m}_{12}), x_1^n(\hat{m}_{12}, \hat{m}_1, \hat{m}_1'), x_2^n(\hat{m}_{12}, \hat{m}_2, \hat{m}_2'), y^n) \in \mathcal{T}_\epsilon^n.$$

d) Error Analysis: The error analysis is standard and is deferred to Appendix A. Eliminating $R_{12}, R_1, R_2, R_1', R_2'$ by Fourier-Motzkin elimination, we arrive at Theorem 1. Cardinality bounds follow by using the standard method via the Fenchel-Eggleston-Carathéodory theorem [11, Appendix C], [14, Appendix B]. ■

Proposition 1. *The lower bound of Theorem 1 is concave in C_1, C_2.*

Proof: We prove the statement for $C_1 = C_2 = C$. The same argument holds in general. We express the lower bound of Theorem 1 in terms of the following maximization problem.

$$f_\ell(C) = \max_{p(u,x_1,x_2)} f_\ell^p(C, p(u,x_1,x_2)) \tag{3}$$

In this formulation, $f_\ell^p(C, p(u,x_1,x_2))$ is the minimum term on the right hand side (RHS) of (1).

The proof is by contradiction. Suppose that the lower bound is not concave in C; i.e., there exist values $C^{(1)}, C^{(2)}$, and α, $0 \le \alpha \le 1$, such that $C^\star = \alpha C^{(1)} + (1-\alpha)C^{(2)}$ and $f_\ell(C^\star) < \alpha f_\ell(C^{(1)}) + (1-\alpha) f_\ell(C^{(2)})$. Let $p^{(1)}(u,x_1,x_2)$ (resp. $p^{(2)}(u,x_1,x_2)$) be the pmf that maximizes $f_\ell^p(C^{(1)}, p(u,x_1,x_2))$ (resp. $f_\ell^p(C^{(2)}, p(u,x_1,x_2))$). Let $p_Q(1) = \alpha$, $p_Q(2) = 1-\alpha$, and define $p_{UX_1X_2|Q}(u,x_1,x_2|1) = p^{(1)}(u,x_1,x_2)$ and $p_{UX_1X_2|Q}(u,x_1,x_2|2) = p^{(2)}(u,x_1,x_2)$. Then we have

$$\begin{aligned}
& f_\ell(C^\star) \\
& < \alpha f_\ell(C^{(1)}) + (1-\alpha) f_\ell(C^{(2)}) \\
& \le \min \left\{ \begin{array}{l} 2C^\star - I(X_1;X_2|UQ), \\ C^\star + I(X_1;Y|X_2UQ), \\ C^\star + I(X_2;Y|X_1UQ), \\ \frac{1}{2}(2C^\star + I(X_1X_2;Y|UQ) - I(X_1;X_2|UQ)), \\ I(UX_1X_2;Y|Q) \end{array} \right\} \\
& \le \min \left\{ \begin{array}{l} 2C^\star - I(X_1;X_2|UQ), \\ C^\star + I(X_1;Y|X_2UQ), \\ C^\star + I(X_2;Y|X_1UQ), \\ \frac{1}{2}(2C^\star + I(X_1X_2;Y|UQ) - I(X_1;X_2|UQ)), \\ I(UX_1X_2;Y) \end{array} \right\} && (4) \\
& \overset{(a)}{\le} f_\ell(C^\star). && (5)
\end{aligned}$$

Step (a) follows by renaming (U,Q) a U and comparing (4) with the lower bound of Theorem 1. ■

IV. An upper bound

The idea behind our upper bound is motivated by [3], [10]. The proposed bound applies not only to Gaussian channels, but also to general discrete memoryless channels. It strictly improves the cut-set bound as we show via two examples. The cut-set bound [15, Theorem 15.10.1] is given by the following Lemma.

Lemma 1 (Cut-Set Bound)**.** *The capacity $C^\diamond$ satisfies*

$$C^\diamond \le \max_{p(x_1,x_2)} \min \left\{ \begin{array}{l} C_1 + C_2, \\ C_1 + I(X_2;Y|X_1), \\ C_2 + I(X_1;Y|X_2), \\ I(X_1X_2;Y) \end{array} \right\}. \tag{6}$$

The cut-set bound disregards the potential correlation between the inputs in the first term of (6). More precisely, we have

$$\begin{aligned}
nR &\le H(V_1^n, V_2^n) \\
&= H(V_1^n) + H(V_1^n) - I(V_1^n; V_2^n) \\
&\le nC_1 + nC_2 - I(X_1^n; X_2^n). && (7)
\end{aligned}$$

It is noted in [2] that optimizing the following n-letter characterization gives the capacity of the network when $n \to \infty$:

$$nR \le nC_1 + nC_2 - I(X_1^n; X_2^n) \tag{8}$$
$$nR \le nC_1 + I(X_2^n; Y^n | X_1^n) \tag{9}$$
$$nR \le nC_2 + I(X_1^n; Y^n | X_2^n) \tag{10}$$
$$nR \le I(X_1^n X_2^n; Y^n). \tag{11}$$

But, infinite letter characterizations are usually non-computable and we would like to find computable bounds.

We prove the following upper bound.

Theorem 2. *The capacity $C^\diamond$ satisfies*

$$C^\diamond \le \max_{p(x_1,x_2)} \min_{p(u|x_1,x_2,y)=p(u|y)} \min \left\{ \begin{array}{l} C_1 + C_2, \\ C_1 + I(X_2;Y|X_1), \\ C_2 + I(X_1;Y|X_2), \\ I(X_1X_2;Y), \\ \frac{1}{2}\left(C_1 + C_2 + I(X_1X_2;Y|U) + I(X_1;U|X_2) + I(X_2;U|X_1)\right) \end{array} \right\}. \tag{12}$$

Remark 3. *For a fixed auxiliary channel $p(u|x_1,x_2,y)$ and a fixed MAC $p(y|x_1,x_2)$, all RHS terms in* (12) *are concave in $p(x_1,x_2)$. See Appendix B.*

Remark 4. *The last term of the minimum in* (12) *may be written as*

$$R \le \frac{1}{2}\left(C_1 + C_2 + I(X_1X_2;YU) - I(X_1;X_2) + I(X_1;X_2|U)\right). \tag{13}$$

Since we choose $p(u|x_1,x_2,y) = p(u|y)$, the bound (13) *becomes*

$$2R \le C_1 + C_2 + I(X_1X_2;Y) - I(X_1;X_2) + I(X_1;X_2|U). \tag{14}$$

Proof of Theorem 2: It is observed in [3] that $I(X_1^n;X_2^n)$ may be written in the following form for *any* integer n, and *any* random sequence U^n:

$$I(X_1^n;X_2^n) = I(X_1^nX_2^n;U^n) - I(X_1^n;U^n|X_2^n) - I(X_2^n;U^n|X_1^n) + I(X_1^n;X_2^n|U^n). \tag{15}$$

Therefore, using (7) and the non-negativity of mutual information we have

$$nR \le nC_1 + nC_2 - I(X_1^nX_2^n;U^n) + I(X_1^n;U^n|X_2^n) + I(X_2^n;U^n|X_1^n). \tag{16}$$

To see the usefulness of (16), we proceed as follows. First, note that

$$nR \le I(X_1^nX_2^n;Y^n) \le I(X_1^nX_2^n;Y^nU^n). \tag{17}$$

Combining inequalities (16) and (17), we have

$$2nR \le nC_1 + nC_2 + I(X_1^nX_2^n;Y^n|U^n) + I(X_1^n;U^n|X_2^n) + I(X_2^n;U^n|X_1^n). \tag{18}$$

Define U_i from X_{1i}, X_{2i}, Y_i through the channel $p_{U|X_1X_2Y}(u_i|x_{1i},x_{2i},y_i)$, $i = 1,2,\ldots,n$. With this choice of U_i we have the following chain of inequalities:

$$\begin{aligned} 2nR &\le nC_1 + nC_2 + I(X_1^nX_2^n;Y^n|U^n) + I(X_1^n;U^n|X_2^n) + I(X_2^n;U^n|X_1^n) \\ &= nC_1 + nC_2 + \sum_{i=1}^n I(X_1^nX_2^n;Y_i|U^nY^{i-1}) + \sum_{i=1}^n I(X_1^n;U_i|X_2^nU^{i-1}) + \sum_{i=1}^n I(X_2^n;U_i|X_1^nU^{i-1}) \\ &\overset{(a)}{\le} nC_1 + nC_2 + \sum_{i=1}^n I(X_{1i}X_{2i};Y_i|U_i) + \sum_{i=1}^n I(X_{1i};U_i|X_{2i}) + \sum_{i=1}^n I(X_{2i};U_i|X_{1i}) \\ &\le nC_1 + nC_2 + nI(X_{1I},X_{2I};Y_I|U_I) + nI(X_{1I};U_I|X_{2I}) + nI(X_{2I};U_I|X_{1I}), \end{aligned} \tag{19}$$

where I is a time-sharing random variable with $p_I(i) = \frac{1}{n}$ for all $i = 1,\ldots,n$. Step (a) holds because of the following two Markov chains:

$$(X_1^nX_2^nU^nY^{i-1}) - (X_{1i}X_{2i}U_i) - Y_i \tag{20}$$

$$(X_1^nX_2^nU^{i-1}) - (X_{1i}X_{2i}) - U_i. \tag{21}$$

From here on, for simplicity we restrict the auxiliary channel to satisfy

$$p_{U|X_1X_2Y}(u|x_1,x_2,y) = p_{U|Y}(u|y), \quad \forall x_1 \in \mathcal{X}_1,\ x_2 \in \mathcal{X}_2,\ u \in \mathcal{U},\ y \in \mathcal{Y}.$$

This proves Theorem 2. ■

We now refine our bounding technique to derive a stronger upper bound in Theorem 3.

Theorem 3. *The capacity $C^\diamond$ satisfies*

$$C^\diamond \le \max_{p(x_1,x_2)} \min_{p(u|x_1,x_2,y)=p(u|y)} \max_{p(q|x_1,x_2,y,u)=p(q|x_1,x_2)} \min \begin{Bmatrix} C_1 + C_2, \\ C_1 + I(X_2;Y|X_1Q), \\ C_2 + I(X_1;Y|X_2Q), \\ I(X_1X_2;Y|Q), \\ C_1 + C_2 - I(X_1X_2;U|Q) + I(X_2;U|X_1Q) + I(X_1;U|X_2Q) \end{Bmatrix}. \tag{22}$$

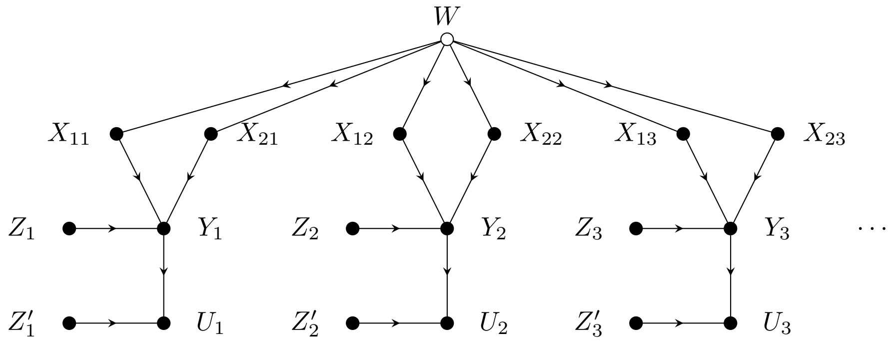


Fig. 2: The associated FDG for Theorem 3 and $n = 3$. The random variables Z_1, Z_2, Z_3 and Z'_1, Z'_2, Z'_3 are appropriate noise random variables with the distributions $p_Z(.)$ and $p_{Z'}(.)$, respectively.

Remark 5. *In the above characterization, it suffices to consider $|\mathcal{Q}| \leq |\mathcal{X}_1||\mathcal{X}_2| + 3$. See Appendix C.*

Remark 6. *The upper bound in* (22) *may be loosened by exchanging the order of the fist maximization and the second minimization. In this case, it suffices to consider $|\mathcal{Q}| \leq 4$. See Remark 21 in Appendix C.*

Remark 7. *The last term of the minimum in* (22) *may be re-written as*

$$C_1 + C_2 - I(X_1; X_2|Q) + I(X_1; X_2|UQ). \tag{23}$$

Observe that the difference of mutual information terms in (23) *also appears in the Hekstra-Willems dependence balance bound [16].*

Remark 8. *The upper bound given in Theorem 3 is tighter than Theorem 2 (see Appendix D). We show through the examples of Sections V and VI that Theorem 3 can strictly improve on Theorem 2.*

Proof of Theorem 3: We start with the multi-letter bound in (8)-(11). We use the identity in (15) to expand inequality (8) for *any random sequence U^n* as follows:

$$\begin{aligned} nR &\leq nC_1 + nC_2 - I(X_1^n; X_2^n) \\ &= nC_1 + nC_2 - I(X_1^n X_2^n; U^n) + I(X_2^n; U^n|X_1^n) + I(X_1^n; U^n|X_2^n) - I(X_1^n; X_2^n|U^n). \end{aligned} \tag{24}$$

In particular, we choose U^n to be such that each symbol U_i is the output of the channel $p_{U|Y}(u_i|y_i)$ with input y_i, $i = 1, \ldots, n$. Thus, we have the functional dependence graph (FDG) depicted in Fig. 2. Furthermore, we have

$$\begin{aligned} nR &\overset{(a)}{\leq} nC_1 + nC_2 - \sum_i \left[I(X_{1i}X_{2i}; U_i|U^{i-1}) + I(X_{2i}; U_i|U^{i-1}X_{1i}) + I(X_{1i}; U_i|U^{i-1}X_{2i}) \right] \\ &= nC_1 + nC_2 - nI(X_{1I}X_{2I}; U_I|U^{I-1}I) + nI(X_{2I}; U_I|U^{I-1}X_{1I}I) + nI(X_{1I}; U_I|U^{I-1}X_{2I}I) \\ &\overset{(b)}{=} nC_1 + nC_2 - nI(X_{1I}X_{2I}; U_I|Q) + nI(X_{2I}; U_I|X_{1I}Q) + nI(X_{1I}; U_I|X_{2I}Q). \end{aligned} \tag{25}$$

Step (a) follows because $U_i - X_{1i}X_{2i}U^{i-1} - X_1^n X_2^n$ forms a Markov chain (see (21)). Step (b) follows by defining $Q = U^{I-1}I$.

We single-letterize (9)-(11) next:

$$\begin{aligned} nR &\leq nC_1 + I(X_2^n; Y^n|X_1^n) \\ &\leq nC_1 + \sum_{i=1}^{n} I(X_{2i}; Y_i|X_{1i}Y^{i-1}) \\ &\overset{(a)}{=} nC_1 + \sum_{i=1}^{n} I(X_{2i}; Y_i|X_{1i}Y^{i-1}U^{i-1}) \\ &\leq nC_1 + \sum_{i=1}^{n} I(X_{2i}; Y_i|X_{1i}U^{i-1}) \\ &= nC_1 + nI(X_{2I}; Y_I|X_{1I}U^{I-1}I) \\ &= nC_1 + nI(X_{2I}; Y_I|X_{1I}Q). \end{aligned} \tag{26}$$

Step (a) follows because $X_{1i}X_{2i}Y_i - Y^{i-1} - U^{i-1}$ forms a Markov chain. Similarly, we have

$$R \leq C_2 + I(X_{1I}; Y_I | X_{2I} Q) \tag{27}$$

$$R \leq I(X_{1I} X_{2I}; Y_I | Q). \tag{28}$$

We further have

$$p_{X_{1I}X_{2I}Y_IU_IQ}(x_1, x_2, y, u, q) = p_{X_{1I}X_{2I}}(x_1, x_2) p_{Q|X_{1I}X_{2I}}(q|x_1, x_2) p_{Y|X_1X_2}(y|x_1, x_2) p_{U|Y}(u|y). \tag{29}$$

Renaming $(X_{1I}, X_{2I}, Y_I, U_I, Q)$ as (X_1, X_2, Y, U, Q) concludes the proof of Theorem 3. ■

Remark 9. *Note that Q is defined based on U. That is, $p(q|x_1, x_2)$ could be a function of $p(u|y)$ and we cannot necessarily change the order in which we minimize over $p(u|y)$ and maximize over $p(q|x_1, x_2)$.*

V. The Gaussian MAC

The output of the Gaussian MAC is

$$Y = X_1 + X_2 + Z$$

where $Z \sim \mathcal{N}(0, 1)$ and the transmitters have average block power constraints P_1, P_2; i.e., we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_{1,i}^2) \leq P_1 \tag{30}$$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_{2,i}^2) \leq P_2. \tag{31}$$

When $C_1 = C_2 = C$ and $P_1 = P_2 = P$, we call the network symmetric.

To find a lower bound on the maximum achievable rate, we use Theorem 1. We choose (U, X_1, X_2) to be jointly Gaussian with zero mean and covariance matrix $\mathbf{K}_{UX_1X_2}$. A special case is when U is null and (X_1, X_2) is jointly Gaussian with the correlation coefficient ρ. The rates that satisfy the following constraints for some ρ, $0 \leq \rho \leq 1$, are thus achievable.

$$R \leq C_1 + C_2 - \frac{1}{2}\log\frac{1}{1-\rho^2} \tag{32}$$

$$R \leq C_1 + \frac{1}{2}\log\left(1 + P_2\left(1-\rho^2\right)\right) \tag{33}$$

$$R \leq C_2 + \frac{1}{2}\log\left(1 + P_1\left(1-\rho^2\right)\right) \tag{34}$$

$$R \leq \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\sqrt{P_1P_2}\rho\right) \tag{35}$$

This choice of (U, X_1, X_2) is not optimal in general. For example when C_1 and C_2 are large (i.e., $C_1, C_2 > \frac{1}{2}\log(1 + P_1 + P_2 + 2\sqrt{P_1P_2})$), the rate

$$R = \frac{1}{2}\log(1 + P_1 + P_2 + 2\sqrt{P_1P_2})$$

is not achievable by (32)-(35) but is achievable by Theorem 1 if we choose (U, X_1, X_2) to be jointly Gaussian and such that $\frac{U}{\sqrt{P_1}} = \frac{X_1}{\sqrt{P_1}} = \frac{X_2}{\sqrt{P_2}} \sim \mathcal{N}(0, 1)$. Theorem 1 therefore gives a strictly larger lower bound compared to [2, Theorem 1], [3, Theorem 2]. More interestingly, in certain regimes of C_1, C_2 the optimal (U, X_1, X_2) is not jointly Gaussian.

Fig. 3 shows the lower bound as a function of C for a symmetric network with $P = 1$. The dotted curve in Fig. 3 shows the rates achieved using the scheme of Section III with jointly Gaussian random variables (U, X_1, X_2) (see [3, Fig. 2] and also [13, Fig. 4]). It is interesting that the obtained lower bound is not concave in C. This does not contradict Proposition 1 because Gaussian distributions are sub-optimal. The improved solid curve shows rates that are achievable using a mixture of two Gaussian distributions. These rates are slightly larger than the rates achieved by time-sharing between two Gaussian distributions with powers $P_1 = P_2 = 1$. If one permits both time-sharing and power control, then one achieves similar rates as for mixture distributions.

Theorems 2 and 3 give upper bounds on the capacity. From Remark 8, Theorem 3 is stronger than Theorem 2. Nevertheless, the bound in Theorem 2 is simpler to evaluate analytically because we can use the maximum entropy lemma to bound all terms. We study both bounds for the Gaussian MAC.

First, we find an upper bound using Theorem 2. We choose $U = Y + Z'$, where Z' is Gaussian noise with zero mean and variance N (to be optimized later). The constrains in (12) are written as follows using maximum entropy lemmas:

$$R \leq C_1 + C_2 \tag{36}$$

$$R \leq C_1 + \frac{1}{2}\log\left(1 + P_2(1-\rho^2)\right) \tag{37}$$

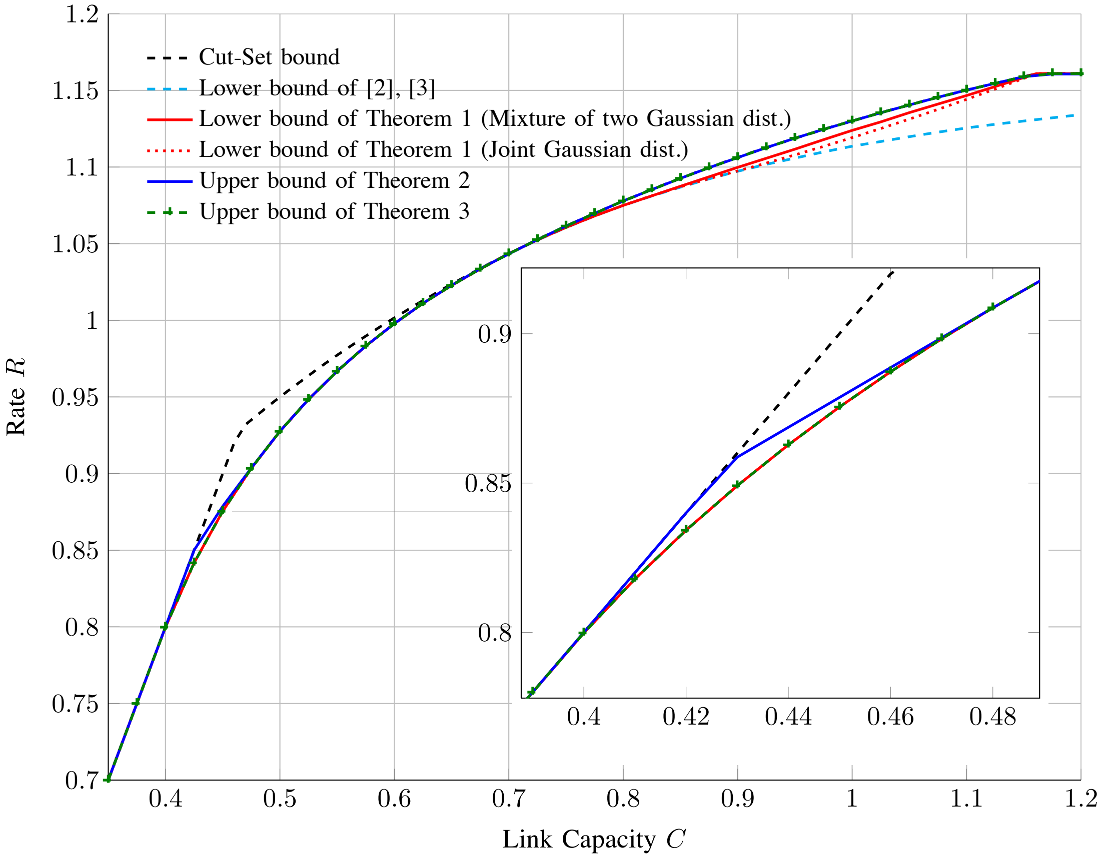


Fig. 3: Upper and lower bounds on R as functions of C for the Gaussian MAC with $P_1 = P_2 = 1$.

$$R \leq C_2 + \frac{1}{2}\log\left(1 + P_1(1-\rho^2)\right) \tag{38}$$

$$R \leq \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\rho\sqrt{P_1P_2}\right) \tag{39}$$

$$2R \overset{(a)}{\leq} C_1 + C_2 + \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\rho\sqrt{P_1P_2}\right) + \frac{1}{2}\log\left(\frac{(1+N+P_1(1-\rho^2))(1+N+P_2(1-\rho^2))}{(1+N+P_1+P_2+2\rho\sqrt{P_1P_2})(1+N)}\right). \tag{40}$$

To obtain inequality (a) above, write the last constraint of (12) as

$$2R \leq C_1 + C_2 + h(Y|U) - h(YU|X_1X_2) + h(U|X_1) + h(U|X_2) - h(U|X_1X_2). \tag{41}$$

The negative terms are easy to calculate because of the Gaussian nature of the channel and the choice of U. The positive terms are bounded from above using the conditional version of the maximum entropy lemma [17]. It remains to solve a max-min problem (max over ρ and min over N). So the rate R is achievable only if there exists some $\rho \geq 0$ for which for every $N \geq 0$ inequalities (36)-(40) hold.

We choose N to be (see [3, eqn. (21)])

$$N = \left(\sqrt{P_1P_2}\left(\frac{1}{\rho} - \rho\right) - 1\right)^+. \tag{42}$$

Let us first motivate this choice. From (14), the inequality in (40) is

$$2R \leq C_1 + C_2 + I(X_1X_2;Y) - I(X_1;X_2) + I(X_1;X_2|U) \tag{43}$$

evaluated for the joint Gaussian distribution $p(x_1, x_2)$ with covariance matrix

$$\begin{bmatrix} P_1 & \rho\sqrt{P_1P_2} \\ \rho\sqrt{P_1P_2} & P_2 \end{bmatrix}. \tag{44}$$

The choice (42) makes U satisfy the Markov chain $X_1 - U - X_2$ for the regime where

$$\sqrt{P_1P_2}\left(\frac{1}{\rho} - \rho\right) - 1 \geq 0 \tag{45}$$

and thus minimizes the RHS of (43). Otherwise, we choose $U = Y$ which results in a redundant bound. The resulting upper bound is summarized in Corollary 1.

Corollary 1. *Rate R is achievable only if there are $\rho \geq 0$ such that*

$$\rho \leq \sqrt{1 + \frac{1}{4P_1P_2}} - \sqrt{\frac{1}{4P_1P_2}}, \tag{46}$$

$$R \leq C_1 + C_2 \tag{47}$$

$$R \leq C_2 + \frac{1}{2}\log\left(1 + P_1(1-\rho^2)\right) \tag{48}$$

$$R \leq C_1 + \frac{1}{2}\log\left(1 + P_2(1-\rho^2)\right) \tag{49}$$

$$R \leq \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\rho\sqrt{P_1P_2}\right) \tag{50}$$

$$2R \leq C_1 + C_2 + \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\rho\sqrt{P_1P_2}\right) - \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right), \tag{51}$$

or

$$\sqrt{1 + \frac{1}{4P_1P_2}} - \sqrt{\frac{1}{4P_1P_2}} \leq \rho \leq 1, \tag{52}$$

$$R \leq C_1 + C_2 \tag{53}$$

$$R \leq C_2 + \frac{1}{2}\log\left(1 + P_1(1-\rho^2)\right) \tag{54}$$

$$R \leq C_1 + \frac{1}{2}\log\left(1 + P_2(1-\rho^2)\right) \tag{55}$$

$$R \leq \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\rho\sqrt{P_1P_2}\right). \tag{56}$$

The above upper bound is plotted in Fig. 3 for different values of C and for $P = 1$. For symmetric diamond networks, we specify a regime of C for which the above upper bound meets the lower bound in Theorem 1 and thus characterizes the capacity. This is summarized in Theorem 4 and its proof is deferred to Appendix E.

Theorem 4. *For a symmetric Gaussian diamond network with orthogonal broadcast links, the upper bound in Theorem 2 is tight if $C \leq \frac{1}{4}\log(1+2P)$, $C \geq \frac{1}{2}\log(1+4P)$, or*

$$\frac{1}{4}\log\frac{1+2P(1+\rho^{(1)})}{1-\left(\rho^{(1)}\right)^2} \leq C \leq \frac{1}{4}\log\frac{1+2P(1+\rho^{(2)})}{1-\left(\rho^{(2)}\right)^2} \tag{57}$$

where

$$\rho^{(1)} = \frac{-(1+2P) + \sqrt{12P^2 + (1+2P)^2}}{6P} \tag{58}$$

$$\rho^{(2)} = \sqrt{1 + \frac{1}{4P^2}} - \frac{1}{2P}. \tag{59}$$

Remark 10. *The $\rho^{(1)}$ given in* (58) *maximizes the RHS of* (51)*. The $\rho^{(2)}$ given in* (59) *is the solution of* (45) *with equality. Note that $\rho^{(2)}$ forms the RHS of* (46) *and the LHS of* (52)*. In other words, for $\rho \leq \rho^{(2)}$ one can find U as a degraded version of Y such that $X_1 - U - X_2$ forms a Markov chain. This is not possible for $\rho > \rho^{(2)}$.*

Remark 11. *For $C \leq \frac{1}{4}\log(1+2P)$, the capacity is equal to $2C$ and is achieved by* (32)-(35) *with $\rho = 0$ (no cooperation among the relays). In the regime* (57)*, the capacity is given by* (32)-(35) *with partial cooperation among the relays. For $C \geq \frac{1}{2}\log(1+4P)$, the capacity is equal to $\frac{1}{2}\log(1+4P)$ and is achieved using Theorem 1 with $X_1 = X_2 = U \sim \mathcal{N}(0,P)$ (full cooperation among the relays).*

Remark 12. *The bound in Corollary 1 and the bound in [3, Theorem 1] are closely related. The bound in [3, Theorem 1] is tighter than Corollary 1 in certain regimes of operation. We will see that Theorem 3 strengthens Corollary 1 and is in general tighter than [3, Theorem 1].*

Based on Theorem 4, the upper and lower bounds match in Fig. 3 (where $P = 1$) for $C \leq 0.3962$, $0.4807 \leq C \leq 0.6942$, and $C \geq 1.1610$. Theorem 3 tightens the above upper bound as we show next. We again choose $U = Y + Z'$ where Z' is a Gaussian random variable with zero mean and variance N (to be optimized). In contrast to Theorem 2, it is not clear whether Gaussian distributions are optimal in Theorem 3. To compute the bound in Theorem 3, we proceed as follows.

The first four bounds of (22) may be loosened by dropping the time-sharing random variable Q and using the maximum entropy lemma:

$$R \leq C_1 + C_2 \tag{60}$$

$$R \leq C_1 + I(X_2;Y|X_1Q) \leq C_1 + \frac{1}{2}\log\left(1 + P_2\left(1-\rho^2\right)\right) \tag{61}$$

$$R \leq C_2 + I(X_1;Y|X_2Q) \leq C_2 + \frac{1}{2}\log\left(1 + P_1\left(1-\rho^2\right)\right) \tag{62}$$

$$R \leq I(X_1X_2;Y|Q) \leq \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\sqrt{P_1P_2}\rho\right). \tag{63}$$

To bound the last constraint in (22), we use both the entropy power inequality [15, Theorem 17.7.3] and the maximum entropy lemma:

$$\begin{aligned}
R \leq & C_1 + C_2 - I(X_1X_2;U|Q) + I(X_1;U|X_2Q) + I(X_2;U|X_1Q) \\
= & C_1 + C_2 - h(U|Q) - h(U|X_1X_2) + h(U|X_2Q) + h(U|X_1Q) \\
\leq & C_1 + C_2 - h(U|Q) - h(U|X_1X_2) + h(U|X_2) + h(U|X_1) \\
\overset{(a)}{\leq} & C_1 + C_2 - \frac{1}{2}\log\left(2\pi e N + 2^{2h(Y|Q)}\right) - h(U|X_1X_2) + h(U|X_2) + h(U|X_1) \\
\overset{(b)}{\leq} & C_1 + C_2 - \frac{1}{2}\log\left(2\pi e N + 2^{2h(Y|Q)}\right) - \frac{1}{2}\log\left(2\pi e(1+N)\right) \\
& + \frac{1}{2}\log\left(2\pi e\left(1 + N + P_1\left(1-\rho^2\right)\right)\right) + \frac{1}{2}\log\left(2\pi e\left(1 + N + P_2\left(1-\rho^2\right)\right)\right)
\end{aligned} \tag{64}$$

where (a) holds by the entropy power inequality and (b) holds by the maximum entropy lemma. We now use $R \leq I(X_1X_2;Y|Q)$ to write

$$\begin{aligned}
h(Y|Q) &= \frac{1}{2}\log(2\pi e) + I(X_1X_2;Y|Q) \\
&\geq \frac{1}{2}\log(2\pi e) + R.
\end{aligned} \tag{65}$$

From (64) and (65) we obtain

$$R \leq C_1 + C_2 - \frac{1}{2}\log\left(N + 2^{2R}\right) - \frac{1}{2}\log\left(1+N\right) + \frac{1}{2}\log\left(1 + N + P_1\left(1-\rho^2\right)\right) + \frac{1}{2}\log\left(1 + N + P_2\left(1-\rho^2\right)\right). \tag{66}$$

Remark 13. *The above argument is similar to the argument used in [10], and it is also related to [18, Section X].*

Remark 14. *Expression* (66) *may be re-written as*

$$R \leq \frac{1}{2}\log\frac{-N + \sqrt{N^2 + 2^{2(C_1+C_2+1)}\frac{(1+N+P_1(1-\rho^2))(1+N+P_2(1-\rho^2))}{1+N}}}{2}. \tag{67}$$

Recall that (67) holds for any value of $N \geq 0$. We choose N as a function of ρ to minimize the RHS of (67). It remains to maximize over ρ and find the maximum rate R admissible by (60)-(63), (67). We solve this optimization problem numerically for the symmetric Gaussian network with $P = 1$, and plot the resulting upper bound in Fig. 3. Note that the upper bound of Theorem 3 is strictly tighter than Theorem 2 for $0.3962 < C < 0.4807$. Furthermore, from the numerical evaluation of the bound, the upper bound of Theorem 3 is tight for $C \leq 0.6942$ and $C \geq 1.1610$. This is made precise for symmetric Gaussian networks in the following theorem which we prove in Appendix F.

Theorem 5. *For a symmetric Gaussian diamond network, the upper bound in Theorem 3 meets the lower bound in Theorem 1 for all C such that $C \geq \frac{1}{2}\log(1+4P)$, or*

$$C \leq \frac{1}{4}\log\frac{1 + 2P(1+\rho^{(2)})}{1 - \left(\rho^{(2)}\right)^2} \tag{68}$$

where

$$\rho^{(2)} = \sqrt{1 + \frac{1}{4P^2}} - \frac{1}{2P}. \tag{69}$$

Sketch of proof: The regime $C \geq \frac{1}{2}\log(1+4P)$ is addressed in Remark 11. We briefly outline the proof for the regime in (68). Consider the lower bound in (32)-(35) and let $R^{(l)}_{\max}$ be the maximum achievable rate. This lower bound meets the cut-set bound (and is thus tight) unless (32) and (35) are both active in which case we have

$$R^{(l)}_{\max} = C_1 + C_2 - \frac{1}{2}\log\frac{1}{1-\lambda^2} = \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\lambda\sqrt{P_1P_2}\right) \tag{70}$$

where λ is the optimal correlation coefficient in (32)-(35). We show in Appendix F that the upper bound given by (60)-(63), (66) meets the lower bound $R_{\max}^{(l)}$ when we have (70) and $\lambda \leq \rho^{(2)}$. One can check for symmetric networks that $\lambda \leq \rho^{(2)}$ if and only if (68) is satisfied. ■

More generally, we have the following result for asymmetric networks. This is addressed in Remark 26 in Appendix F.

Theorem 6. *The upper bound in Theorem 3 meets the lower bound in Theorem 1 if any of the following conditions hold:*

$$C_1 + C_2 \leq \frac{1}{2}\log\left(\frac{1 + P_1 + P_2 + 2\rho^{(2)}\sqrt{P_1P_2}}{1 - \left(\rho^{(2)}\right)^2}\right) \tag{71}$$

$$C_1 \leq \frac{1}{2}\log\left(\frac{1 + P_1 + P_2 + 2\rho_0\sqrt{P_1P_2}}{1 + P_2(1-\rho_0^2)}\right) \tag{72}$$

$$C_2 \leq \frac{1}{2}\log\left(\frac{1 + P_1 + P_2 + 2\rho_0\sqrt{P_1P_2}}{1 + P_1(1-\rho_0^2)}\right) \tag{73}$$

$$\min(C_1, C_2) \geq \frac{1}{2}\log\left(1 + P_1 + P_2 + 2\sqrt{P_1P_2}\right) \tag{74}$$

where $\rho^{(2)}$ is given by (69) *and ρ_0 is given by*

$$\rho_0 = \frac{-\sqrt{P_1P_2} + \sqrt{P_1P_2 + 2^{2(C_1+C_2)}\left(2^{2(C_1+C_2)} - 1 - P_1 - P_2\right)}}{2^{2(C_1+C_2)}}.$$

Remark 15. *ρ_0 is defined such that $C_1 + C_2 = \frac{1}{2}\log\left(\frac{1+P_1+P_2+2\rho_0\sqrt{P_1P_2}}{1-\rho_0^2}\right)$. Note that we have $\rho_0 \leq \rho^{(2)}$ if and only if* (71) *is satisfied. In defining ρ_0, we have implicitly assumed that $C_1 + C_2 \geq \frac{1}{2}\log(1 + P_1 + P_2)$; this is without loss of generality because otherwise C_1, C_2 are in the regime defined by* (71).

Remark 16. *In the regime* (74)*, the cut-set bound is achievable using Theorem 1 with $\frac{U}{\sqrt{P_1}} = \frac{X_1}{\sqrt{P_1}} = \frac{X_2}{\sqrt{P_2}} \sim \mathcal{N}(0,1)$ and the lower bound in* (32)-(35) *is loose.*

Remark 17. *Theorem 6 reduces to Theorem 5 when $P_1 = P_2 = P$ and $C_1 = C_2 = C$.*

Remark 18. *Theorem 3 is strictly tighter than [3, Theorem 1] and [13, Theorem 1]. The regime of interest is given by* (70) *because otherwise both upper bounds reduce to the cut-set bound which is tight. First suppose $\lambda > \rho^{(2)}$. In this case, [3, Theorem 1] reduces to the cut-set bound and is larger than or equal to the upper bound of Theorem 3. Next suppose $\lambda \leq \rho^{(2)}$. Here, the upper bound given by* (60)-(63), (66) *can be shown to be equal to $R_{\max}^{(l)}$ and is thus tight but [3, Theorem 1] may not be tight, see [13, Theorem 3]. For example, when $P_1 = P_2 = 0.25$ and $C_1 = C_2 = 0.15$ Theorem 3 gives $C^{\diamond} \leq .2994$ (which is tight) whereas [3, Theorem 1] gives $C^{\diamond} \leq 0.3$. The looseness of [3, Theorem 1] in comparison to our result seems to be due to the relaxation of [3, eqn. (28)] in the final theorem statement of [3, Theorem 1].*

VI. The Binary Adder Channel

Consider the binary adder channel defined by $\mathcal{X}_1 = \{0,1\}$, $\mathcal{X}_2 = \{0,1\}$, $\mathcal{Y} = \{0,1,2\}$, and $Y = X_1 + X_2$. Suppose without loss of generality that $C_1 \leq C_2$. When $C_1 = C_2 = C$, we call the network symmetric. The best known upper bound for this channel is the cut-set bound and the best known lower bound is given by [2, Theorem 1]. More precisely, using a doubly symmetric input distribution, R is achievable if it satisfies the following inequalities for some p, $0 \leq p \leq 1$.

$$R \leq C_1 + C_2 - 1 + h_2(p) \tag{75}$$

$$R \leq C_1 + h_2(p) \tag{76}$$

$$R \leq h_2(p) + 1 - p \tag{77}$$

This lower bound is a special case of Theorem 1 with U a constant. The bound is plotted in Fig. 4 as a function of C for symmetric networks where $C_1 = C_2 = C$.

We evaluate Theorems 2 and 3 to derive new upper bounds on the achievable rate. The obtained upper bounds are plotted in Fig. 4. It turns out that the upper bound of Theorem 3 meets the lower bound for all ranges of C. Theorem 3 is better than Theorem 2, but Theorem 2 is simpler to analyze and gives capacity for $C \leq 0.75$ and $C \geq .7929$.

Consider first Theorem 2. Let $p(u|y)$ be a symmetric channel as shown in Fig. 5 with parameter α, $\alpha \leq \frac{1}{2}$, to be optimized. From Theorem 2, we must solve a max-min problem (max over $p(x_1, x_2)$, min over α). For a fixed α, the upper bound is concave in $p(x_1, x_2)$ (see Remark 3). The concavity together with the symmetry of the problem and the auxiliary channel in $p(x_1, x_2)$ imply the following lemma. We defer the proof to Appendix G.

Lemma 2. *An optimizing pmf $p(x_1, x_2)$ in* (12) *is that of a doubly symmetric binary source.*

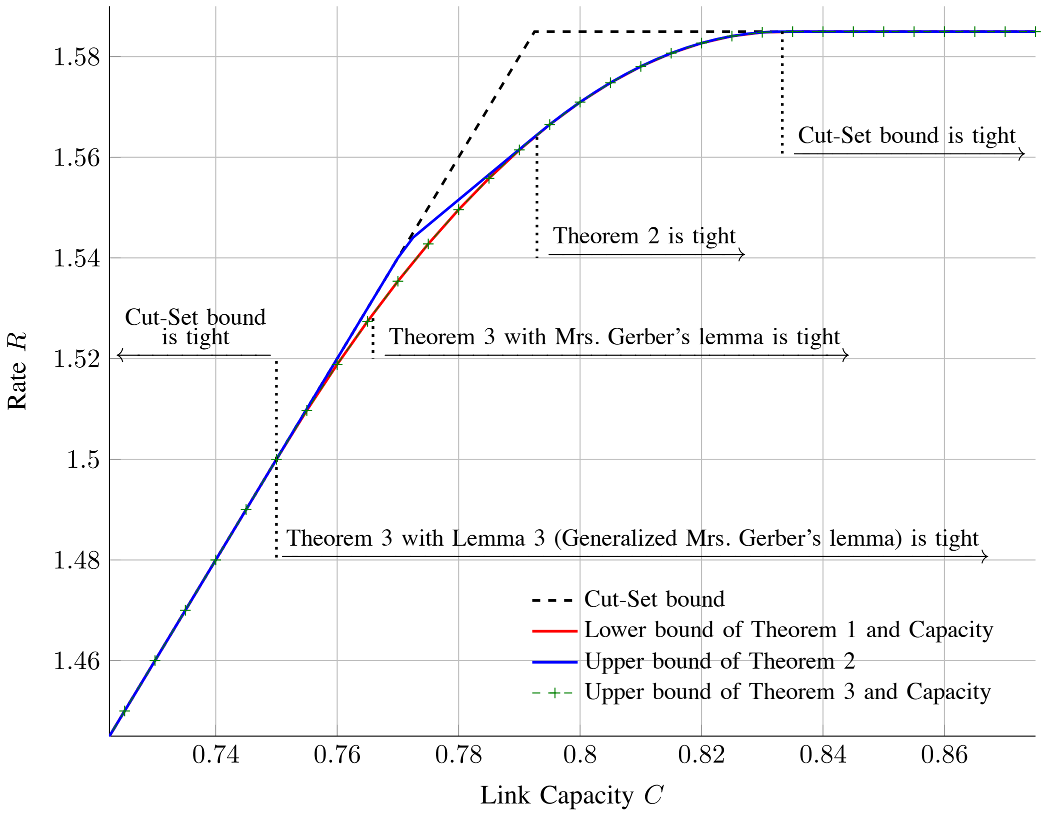


Fig. 4: Upper and lower bounds on R as functions of C for the binary adder MAC.

So suppose $p(x_1, x_2)$ is a doubly symmetric binary source with parameter p. The upper bound in Theorem 2 with $p(u|y)$ in Fig. 5 reduces to

$$\max_{0\le p\le \frac{1}{2}} \min_{0\le\alpha\le\frac{1}{2}} \min \left\{ \begin{array}{l} C_1 + C_2, \\ C_1 + h_2(p), \\ h_2(p) + 1 - p, \\ \frac{C_1+C_2}{2} + h_2(p) - \frac{p}{2} + \frac{1}{2} I(X_1;X_2|U) \end{array} \right\} \tag{78}$$

where the last term of (78) is written using (14), and where the range of p is $[0, \frac{1}{2}]$. Note that α is implicit in $I(X_1;X_2|U)$:

$$I(X_1;X_2|U) = 2h_2\left(\alpha \star \frac{p}{2}\right) - (1-p)h_2(\alpha) - h_2(p) - p. \tag{79}$$

Here the $\star$ operator is defined by $\alpha \star \beta = \alpha(1-\beta) + \beta(1-\alpha)$, $\beta \le 1$. To obtain the best bound, we choose U such that $X_1 - U - X_2$ forms a Markov chain. This requires

$$\alpha(1-\alpha) = \left(\frac{p}{2(1-p)}\right)^2 \tag{80}$$

which has a solution for α because $p \le \frac{1}{2}$.

Corollary 2. *Rate R is achievable only if there is some p, $0 \le p \le \frac{1}{2}$, such that*

$$R \le C_1 + C_2 \tag{81}$$
$$R \le C_1 + h_2(p) \tag{82}$$
$$R \le h_2(p) + 1 - p \tag{83}$$
$$R \le \frac{C_1+C_2}{2} + h_2(p) - \frac{p}{2}. \tag{84}$$

We compare Corollary 2 with the lower bound in (75)-(77) for symmetric networks and find the capacity for some ranges of C. This is summarized in Theorem 7 and the proof is deferred to Appendix H.

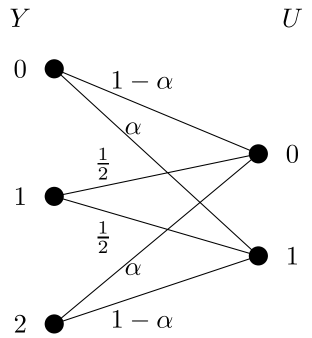


Fig. 5: Auxiliary channel $p(u|y)$ for the binary adder MAC.

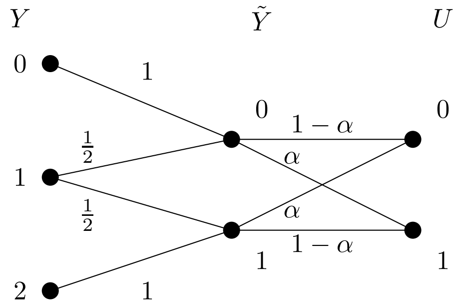


Fig. 6: The auxiliary channel $p(u|y)$ as the cascade of $p(\tilde{y}|y)$ and $p(u|\tilde{y})$ for the binary adder MAC.

Theorem 7. *The upper bound in Theorem 2 meets the lower bound in Theorem 1 for the symmetric diamond network with a binary adder channel if* $C \leq .75$ *or*

$$C \geq 1 - \frac{p^{(1)}}{2} \approx 0.7929 \tag{85}$$

where $p^{(1)} = \frac{1}{1+\sqrt{2}} \approx 0.4142$.

In the rest of this section, we show that Theorem 3 gives the capacity of the diamond network with a binary adder MAC for all ranges of C_1, C_2. We first state a generalization of Mrs. Gerber's Lemma [19] that we prove in Appendix J. For other generalizations, please see [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. Our generalization is different than previous ones in that it establishes the convexity of a *difference* of entropies, rather than an individual entropy. In this sense, Lemma 3 seems similar to an extension [30] of Shannon's entropy power inequality [31].

Lemma 3 (Generalization of Mrs. Gerber's Lemma)**.** *The function*

$$g(x,y) = h_2\left(\alpha \star \left(\frac{y}{2} + (1-y)h_2^{-1}\left(\frac{(x-h_2(y))^+}{1-y}\right)\right)\right) - h_2\left(\alpha \star \frac{y}{2}\right) \tag{86}$$

is jointly convex in x *and* y, $0 \leq x \leq 1 + h_2(y) - y$, $0 \leq y \leq 1$. *We recover Mrs. Gerber's Lemma by choosing* $y = 0$.

Theorem 8. *The upper bound of Theorem 3 matches the lower bound of Theorem 1; i.e., the capacity* $C^\diamond$ *of diamond networks with binary adder MACs and* $C_1 \leq C_2$ *is*

$$C^\diamond = \max_{0 \leq p \leq \frac{1}{2}} \min \begin{cases} C_1 + C_2 - 1 + h_2(p) \\ C_1 + h_2(p) \\ h_2(p) + 1 - p. \end{cases} \tag{87}$$

Proof: We again use the auxiliary channel $p(u|y)$ depicted in Fig. 5. This channel may be viewed as the cascade of the channels $p(\tilde{y}|y)$ and $p(u|\tilde{y})$ shown in Fig. 6, where $p(u|\tilde{y})$ is a BSC with cross over probability α. Define p_i, q_i, $i \in \mathcal{Q}$, and q by

$$p_i = p_{Y|Q}(0|i) \tag{88}$$

$$q_i = p_{Y|Q}(1|i) \tag{89}$$

$$q = p_Y(1). \tag{90}$$

The first four terms of (22) may be loosened by dropping the time sharing random variable Q. We use the symmetry and concavity of those terms in $p(x_1,x_2)$ to write

$$R \le C_1 + C_2 \tag{91}$$
$$R \le C_1 + h_2(q) \tag{92}$$
$$R \le h_2(q) + 1 - q. \tag{93}$$

It remains to bound the last term of (22):

$$\begin{aligned} R \le& C_1 + C_2 - I(X_1X_2;U|Q) + I(X_2;U|X_1Q) + I(X_1;U|X_2Q) \\ =& C_1 + C_2 - H(U|Q) - H(U|X_1X_2) + H(U|X_1Q) + H(U|X_2Q). \end{aligned} \tag{94}$$

We optimize the RHS of (94) under the constraint

$$R \le I(X_1X_2;Y|Q) = H(Y|Q) \tag{95}$$

that is imposed by the fourth term of (22). We have $H(U|X_1X2) = (1-q)h_2(\alpha)+q$ and can upper bound both $H(U|X_1Q=i)$ and $H(U|X_1Q=i)$ by

$$h_2\left(\alpha \star \frac{q_i}{2}\right)$$

by the concavity of $h_2(.)$ and symmetry of $p_{U|Y}$ (see Appendix I). But how to bound (94) from above is not obvious because $H(U|Q)$ appears with a negative sign. We start with

$$H(U|Q=i) = h_2\left(\alpha \star \left(\frac{q_i}{2} + p_i\right)\right) \tag{96}$$
$$H(Y|Q=i) = h_2(q_i) + (1-q_i)h_2\left(\frac{p_i}{1-q_i}\right). \tag{97}$$

Note that both (96) and (97) are symmetric in p_i with respect to $\frac{1-q_i}{2}$. We may therefore choose $p_i \le \frac{1-q_i}{2}$, find p_i from (97) and insert it into (96) to obtain

$$H(U|Q=i) = h_2\left(\alpha \star \left(\frac{q_i}{2} + (1-q_i)h_2^{-1}\left(\frac{H(Y|Q=i) - h_2(q_i)}{1-q_i}\right)\right)\right). \tag{98}$$

Combining the above bounds and inserting in (94), we have

$$\begin{aligned} R \le& C_1 + C_2 + \sum_{i=1}^{|\mathcal{Q}|} p_Q(i)\left(-h_2\left(\alpha \star \left(\frac{q_i}{2} + (1-q_i)h_2^{-1}\left(\frac{H(Y|Q=i) - h_2(q_i)}{1-q_i}\right)\right)\right) - (1-q_i)h_2(\alpha) - q_i + 2h_2\left(\alpha \star \frac{q_i}{2}\right)\right) \\ \overset{(a)}{\le}& C_1 + C_2 - h_2\left(\alpha \star \left(\frac{q}{2} + (1-q)h_2^{-1}\left(\frac{(H(Y|Q) - h_2(q))^+}{1-q}\right)\right)\right) - (1-q)h_2(\alpha) - q + 2h_2\left(\alpha \star \frac{q}{2}\right) \\ \overset{(b)}{\le}& C_1 + C_2 - h_2\left(\alpha \star \left(\frac{q}{2} + (1-q)h_2^{-1}\left(\min\left(1, \frac{(R - h_2(q))^+}{1-q}\right)\right)\right)\right) - (1-q)h_2(\alpha) - q + 2h_2\left(\alpha \star \frac{q}{2}\right) \end{aligned} \tag{99}$$

where (a) follows by concavity of the binary entropy function and Lemma 3, and (b) follows from (95) because $h_2(\alpha \star \left(\frac{q}{2} + (1-q)h_2^{-1}(x)\right))$ is non-decreasing in x for $\alpha \le \frac{1}{2}$. Choosing α appropriately, we show in Appendix K that the upper bound in (91)-(93) and (99) matches the lower bound in (75)-(77).

■

Remark 19. *One may use Mrs. Gerber's lemma [19] to obtain*

$$\begin{aligned} H(U|Q) \ge& h_2\left(\alpha \star h_2^{-1}\left(H(\tilde{Y}|Q)\right)\right) \\ \ge& h_2\left(\alpha \star h_2^{-1}\left(\min\left(1, (R - h_2(q) + q)^+\right)\right)\right) \end{aligned} \tag{100}$$

where the second inequality follows by

$$\begin{aligned} R &\le I(X_1X_2;Y|Q) \\ &= I(X_1X_2;Y\tilde{Y}|Q) \\ &= H(\tilde{Y}|Q) + H(Y|\tilde{Y}Q) - H(\tilde{Y}|X_1X_2) \\ &\le H(\tilde{Y}|Q) + h_2(q) - q \end{aligned} \tag{101}$$

and the monotonicity of $h_2(\alpha \star h_2^{-1}(x))$ in x. This approach gives an upper bound that is tight when $C_1 + C_2 \ge 1.5317$ (and when $C_1 + C_2 \le 1.5$). The range of symmetric bit-pipe capacities C for which Mrs. Gerber's lemma is tight is shown in Fig. 4. In fact, Mrs. Gerber's lemma is within less than 10^{-3} bits of capacity for all C in Fig. 4.

VII. Conclusion

We studied diamond networks with an orthogonal broadcast channel and found new upper and lower bounds on their capacities. The lower bound is based on Marton's coding technique and superposition coding. We showed through an example with a Gaussian MAC that the new lower bound strictly improves the previous bounds in [2], [3], [13]. The proof technique for the upper bound generalizes bounding techniques of Ozarow [10] and Kang and Liu [3] and applies to discrete memoryless MACs. We specialized the upper bound for networks with a Gaussian MAC and a binary adder MAC. We strengthened the results of Kang and Liu [3], [13] for Gaussian MACs and found the capacity for binary adder MACs.

Appendix A
Error probability analysis for Section III

We give a proof for discrete alphabet MACs. A proof for the AWGN MAC follows in the usual way be quantizing alphabets and taking limits [11, Page 50]. We calculate the average error probability P_e, averaged over the codebook and the message set, and show that P_e approaches zero, as n gets large, if we have

$$\begin{aligned}
& R_1' + R_2' > I(X_1;X_2|U) && (102)\\
& R_{12} + R_1 + R_1' < C_1 && (103)\\
& R_{12} + R_2 + R_2' < C_2 && (104)\\
& R_{12} + R_1 + R_1' + R_2 + R_2' < I(X_1X_2;Y) + I(X_1;X_2|U) && (105)\\
& R_1 + R_1' + R_2 + R_2' < I(X_1X_2;Y|U) + I(X_1;X_2|U) && (106)\\
& R_2 + R_2' < I(X_2;Y|X_1,U) + I(X_1;X_2|U) && (107)\\
& R_1 + R_1' < I(X_1;Y|X_2,U) + I(X_1;X_2|U). && (108)
\end{aligned}$$

Conditions (102)-(108), together with $R_1', R_2', R_1, R_2, R_{12} \geq 0$ and the rate-splitting condition in (2), characterize an achievable rate. By the symmetry of the codebook construction and the encoding/decoding scheme, we have

$$\begin{aligned}
P_e &= \Pr\Big((M_{12},M_1,M_2) \neq (\hat{M}_{12},\hat{M}_1,\hat{M}_2)\Big)\\
&= \Pr\Big((M_{12},M_1,M_2) \neq (\hat{M}_{12},\hat{M}_1,\hat{M}_2) | (M_{12},M_1,M_2) = (1,1,1)\Big). \quad (109)
\end{aligned}$$

Conditioned on $(M_{12},M_1,M_2) = (1,1,1)$, an error occurs only if one of the following events occurs:

- $\mathcal{E}_1$: There is no index pair (m_1', m_2') such that $(U^n(1), X_1^n(1,1,\hat{m}_1'), X_2^n(1,1,\hat{m}_2'), Y^n) \in \mathcal{T}_\epsilon^n$.
- $\mathcal{E}_2$: There are $\tilde{m}_{12}, \tilde{m}_1, \tilde{m}_2, \tilde{m}_1', \tilde{m}_2'$ such that $(\tilde{m}_{12}, \tilde{m}_1, \tilde{m}_2) \neq (1,1,1)$ and

$$(U^n(\tilde{m}_{12}), X_1^n(\tilde{m}_{12},\tilde{m}_1,\tilde{m}_1'), X_2^n(\tilde{m}_{12},\tilde{m}_2,\tilde{m}_2'), Y^n) \in \mathcal{T}_\epsilon^n.$$

We have

$$\begin{aligned}
P_e &\leq \Pr(\mathcal{E}_1 | (M_{12},M_1,M_2) = (1,1,1)) + \Pr(\mathcal{E}_2 | (M_{12},M_1,M_2) = (1,1,1))\\
&= \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2). \quad (110)
\end{aligned}$$

Using the Mutual Covering lemma [11, Lemma 8.1], $\Pr(\mathcal{E}_1)$ can be made small for large n if (102) is satisfied. To analyze $\Pr(\mathcal{E}_2)$, consider the following partition of $\mathcal{E}_2$:

- $\tilde{m}_{12} \neq 1$
- $\tilde{m}_{12} = 1$, $\tilde{m}_1 \neq 1$, $\tilde{m}_2 \neq 1$
- $\tilde{m}_{12} = 1$, $\tilde{m}_1 = 1$, $\tilde{m}_2 \neq 1$
- $\tilde{m}_{12} = 1$, $\tilde{m}_1 \neq 1$, $\tilde{m}_2 = 1$.

The first case has a small error probability, for large n, if (105) is satisfied. Similarly, the second, third, fourth cases have small error probabilities, for large n, if (106), (107), (108) are satisfied, respectively. We illustrate the analysis for the second case here:

$$\begin{aligned}
&\Pr\left(\bigcup_{\tilde{m}_1\neq 1,\tilde{m}_2\neq 1,\tilde{m}_1',\tilde{m}_2'} (U^n(1), X_1^n(1,\tilde{m}_1,\tilde{m}_1'), X_2^n(1,\tilde{m}_2,\tilde{m}_2'), Y^n) \in \mathcal{T}_\epsilon^n\right)\\
\leq\ & \sum_{\tilde{m}_1\neq 1,\tilde{m}_2\neq 1,\tilde{m}_1',\tilde{m}_2'} \Pr\left((U^n(1), X_1^n(1,\tilde{m}_1,\tilde{m}_1'), X_2^n(1,\tilde{m}_2,\tilde{m}_2'), Y^n) \in \mathcal{T}_\epsilon^n\right)\\
\leq\ & 2^{n(R_1+R_1'+R_2+R_2')} \sum_{(u^n,x_1^n,x_2^n,y^n)\in\mathcal{T}_\epsilon} p(u^n)p(x_1^n|u^n)p(x_2^n|u^n)p(y^n|u^n)\\
\leq\ & 2^{n(R_1+R_1'+R_2+R_2')} 2^{n(H(UX_1X_2Y)+\epsilon H(UX_1X_2Y))}
\end{aligned}$$

$$\times\, 2^{-n(H(U)-\epsilon H(U))} 2^{-n(H(X_1|U)-\epsilon H(X_1|U))} 2^{-n(H(X_2|U)-\epsilon H(X_2|U))} 2^{-n(H(Y|U)-\epsilon H(Y|U))}$$
$$= \quad 2^{n(R_1+R_1'+R_2+R_2')} 2^{-n(I(X_1X_2;Y|U)+I(X_1;X_2|U)-\delta(\epsilon))} \tag{111}$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. For rates that satisfy (106), the RHS in (111) approaches zero as n grows large.

APPENDIX B
CONCAVITY IN $p(x_1, x_2)$

Consider (12) and

$$I(X_1;Y|X_2) = H(Y|X_2) - H(Y|X_1X_2). \tag{112}$$

$H(Y|X_2)$ is a concave function of $p(x_2, y)$ which is a linear function of $p(x_1, x_2)$. $H(Y|X_1X_2)$ is a linear function of $p(x_1, x_2)$. So $I(X_1;Y|X_2)$ is concave in $p(x_1, x_2)$. A similar result holds for $I(X_2;Y|X_1)$ and $I(X_1, X_2;Y)$ when $p(y|x_1, x_2)$ is fixed.

Finally, consider the last RHS term in (12) when $p(y|x_1, x_2)$ and $p(u|x_1, x_2, y)$ are fixed. We have

$$\begin{aligned} &I(X_1X_2;Y|U) + I(X_1;U|X_2) + I(X_2;U|X_1) \\ &= H(Y|U) - H(YU|X_1X_2) + H(U|X_2) + I(X_2;U|X_1). \end{aligned} \tag{113}$$

$H(Y|U)$ is concave in $p(u, y)$, $H(YU|X_1X_2)$ is linear in $p(x_1, x_2)$, $H(U|X_2)$ is concave in $p(u, x_2)$ and $I(X_2;U|X_2)$ is concave in $p(x_1, x_2)$. Since $p(u, y)$ and $p(u, x_2)$ are both linear in $p(x_1, x_2)$, (113) is concave in $p(x_1, x_2)$.

APPENDIX C
CARDINALITY BOUND FOR THEOREM 3

We follow the line of argument in [14, Appendix B]. Denote by $\mathcal{P}$ the set of all probability vectors $p(x_1, x_2)$ and let P be an element of $\mathcal{P}$. Suppose that R is such that for a certain distribution $p_0(x_1, x_2, u, y, q) = p_0(x_1, x_2)p^\star(y|x_1, x_2)p^\star(u|y)p_0(q|x_1x_2)$ the following inequalities hold:

$$R \le C_1 + I_0(X_2;Y|X_1Q), \tag{114}$$
$$R \le C_2 + I_0(X_1;Y|X_2Q), \tag{115}$$
$$R \le I_0(X_1X_2;Y|Q), \tag{116}$$
$$R \le C_1 + C_2 - I_0(X_1X_2;U|Q) + I_0(X_2;U|X_1Q) + I_0(X_1;U|X_2Q) \tag{117}$$

In the above inequalities, the index 0 on the mutual information terms emphasizes that the mutual information is evaluated for $p_0(x_1, x_2, u, y, q)$. We now interpret $p_0(x_1, x_2|q)$ for every $q \in \mathcal{Q}$ as an element P_0^q of $\mathcal{P}$ with a corresponding probability $p_0(q)$. Consider the following continuous functions that map an element of $\mathcal{P}$ into an element of $\mathbb{R}$.

$$f_{x_1,x_2}(P) = \Pr_P\{X_1 = x_1, X_2 = x_2\}, \quad \forall (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \text{ except one} \tag{118}$$
$$f_I(P) = I_P(X_2;Y|X_1), \tag{119}$$
$$f_{II}(P) = I_P(X_1;Y|X_2), \tag{120}$$
$$f_{III}(P) = I_P(X_1X_2;Y), \tag{121}$$
$$f_{IV}(P) = -I_P(X_1X_2;U) + I_P(X_2;U|X_1) + I_P(X_1;U|X_2) \tag{122}$$

We are interested in the following terms.

$$p_0(x_1, x_2) = \sum_{q\in\mathcal{Q}} f_{x_1,x_2}(P_0^q)p_0(q), \quad \forall (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \text{ except one} \tag{123}$$
$$I_0(X_2;Y|X_1Q) = \sum_{q\in\mathcal{Q}} f_I(P_0^q)p_0(q), \tag{124}$$
$$I_0(X_1;Y|X_2Q) = \sum_{q\in\mathcal{Q}} f_{II}(P_0^q)p_0(q), \tag{125}$$
$$I_0(X_1X_2;Y|Q) = \sum_{q\in\mathcal{Q}} f_{III}(P_0^q)p_0(q), \tag{126}$$
$$-I_0(X_1X_2;U|Q) + I_0(X_2;U|X_1Q) + I_0(X_1;U|X_2Q) = \sum_{q\in\mathcal{Q}} f_{IV}(P_0^q)p_0(q) \tag{127}$$

The Fenchel-Eggleston-Carathéodory theorem [11, Appendix A] ensures that there are $|\mathcal{X}_1||\mathcal{X}_2| + 3$ vectors $P_k \in \mathcal{P}$, $k = 1, \ldots, |\mathcal{X}_1||\mathcal{X}_2| + 3$, whose convex combination gives (123)-(127). In other words, we may restrict attention to $|\mathcal{Q}| \le |\mathcal{X}_1||\mathcal{X}_2| + 3$.

Remark 20. *We need to keep $p(x_1, x_2)$ fixed because $p^\star(u|y)$ can be a function of $p(x_1, x_2)$ and we don't want to change $p^\star(u|y)$.*

Remark 21. *If $p^\star(u|y)$ is fixed (and not a function of $p(x_1, x_2)$), then it suffices to have $|\mathcal{Q}| \leq 4$ because we must fix only* (119)-(122).

APPENDIX D
THEOREM 3 IS TIGHTER THAN THEOREM 2

Let R be less than or equal to the upper bound of Theorem 3. Therefore, there is a $p(x_1, x_2)$ for which for all $p(u|x_1, x_2, y) = p(u|y)$ there is a $p(q|x_1, x_2, y, u) = p(q|x_1, x_2)$ such that the constraints in (22) hold. Combining the last two bounds in (22), we obtain

$$\begin{aligned} 2R &\leq C_1 + C_2 + I(X_1X_2;Y|Q) - I(X_1X_2;U|Q) + I(X_1;U|X_2Q) + I(X_2;U|X_1Q) \\ &= C_1 + C_2 + I(X_1X_2;Y|UQ) + I(X_1;U|X_2Q) + I(X_2;U|X_1Q) \\ &\leq C_1 + C_2 + I(X_1X_2;Y|U) + I(X_1;U|X_2) + I(X_2;U|X_1). \end{aligned} \tag{128}$$

Furthermore, we have

$$I(X_2;Y|X_1Q) \leq I(X_2;Y|X_1) \tag{129}$$
$$I(X_1;Y|X_2Q) \leq I(X_1;Y|X_2) \tag{130}$$
$$I(X_1X_2;Y|Q) \leq I(X_1X_2;Y). \tag{131}$$

APPENDIX E
PROOF OF THEOREM 4

Consider first the regime

$$C \leq \frac{1}{4}\log(1 + 2P).$$

The lower bound of Theorem 1 meets the upper bound of Corollary 1 for $U = \phi$, $X_1 \sim \mathcal{N}(0, P)$, $X_2 \sim \mathcal{N}(0, P)$, and X_1 independent of X_2. Consider next

$$C \geq \frac{1}{2}\log(1 + 4P).$$

The lower bound of Theorem 1 meets the upper bound of Corollary 1 for $X_1 = X_2 = U \sim \mathcal{N}(0, P)$.

The more interesting regime of C is given in (57). Consider the upper bound in Corollary 1 and define the functions

$$\begin{array}{l} f_1(C) = 2C \\ f_2(C,\rho) = C + \frac{1}{2}\log\left(1 + P(1-\rho^2)\right) \\ f_3(\rho) = \frac{1}{2}\log\left(1 + 2P(1+\rho)\right) \\ f_4(C,\rho) = \frac{1}{2}\left(2C + \frac{1}{2}\log\left(1 + 2P(1+\rho)\right) - \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right)\right) \\ f_4'(C,\rho) = \left\{ \begin{array}{ll} f_4(C,\rho) & \rho \leq \rho^{(2)} \\ f_2(C,\rho) & \rho > \rho^{(2)}. \end{array} \right. \end{array} \tag{132}$$

The functions $f_1(C)$, $f_2(C, \rho)$, $f_3(\rho)$, $f_4'(C, \rho)$ are plotted in Fig. 7 for different values of C (where C increases from Fig. 7a to Fig. 7c).

Remark 22. *For symmetric diamond networks, one can check that $f_4(C, \rho) \leq f_2(C, \rho)$ for all $0 \leq \rho \leq 1$. Recall from Corollary 1 that $f_4(C, \rho)$ is not limiting for $\rho > \rho^{(2)}$. This is reflected in the definition of $f_4'(C, \rho)$. So we can write the upper bound of Corollary 1 as*

$$\max_{\rho} \min\{f_1(C), f_3(\rho), f_4'(C, \rho)\}. \tag{133}$$

Remark 23. *The function $f_4(C, \rho)$ is concave in ρ and it attains its maximum at $\rho^{(1)}$ given in* (58). *$f_2(C, \rho)$ is concave and decreasing in ρ. One can check by substitution and differentiation that $f_4'(C, \rho)$ is continuous and differentiable with respect to ρ at $\rho = \rho^{(2)}$. Furthermore, $f_4'(C, \rho)$ is concave and attains its maximum at $\rho^{(1)}$. The derivative of $f_4'(C, \rho)$ is non-positive at $\rho^{(2)}$ and thus we have $\rho^{(1)} \leq \rho^{(2)}$.*

Remark 24. *In the regime $\frac{1}{4}\log(1 + 2P) < C < \frac{1}{2}\log(1 + 4P)$ the functions $f_3(\rho)$ and $f_4'(C, \rho)$ have exactly one point ρ where $f_3(\rho) = f_4'(C, \rho)$. To see this we study the zeros of the function $g(C, \rho) = f_3(\rho) - f_4'(C, \rho)$. Since $C > \frac{1}{4}\log(1 + 2P)$, we have $g(C, 0) < 0$. Since $C < \frac{1}{2}\log(1 + 4P)$, we have $g(C, 1) > 0$. Furthermore, we have*

$$\frac{\partial g}{\partial \rho}(C, \rho) = \frac{\partial f_3}{\partial \rho}(\rho) - \frac{\partial f_4'}{\partial \rho}(C, \rho) > 0, \quad \forall \rho \in [0, 1] \tag{134}$$

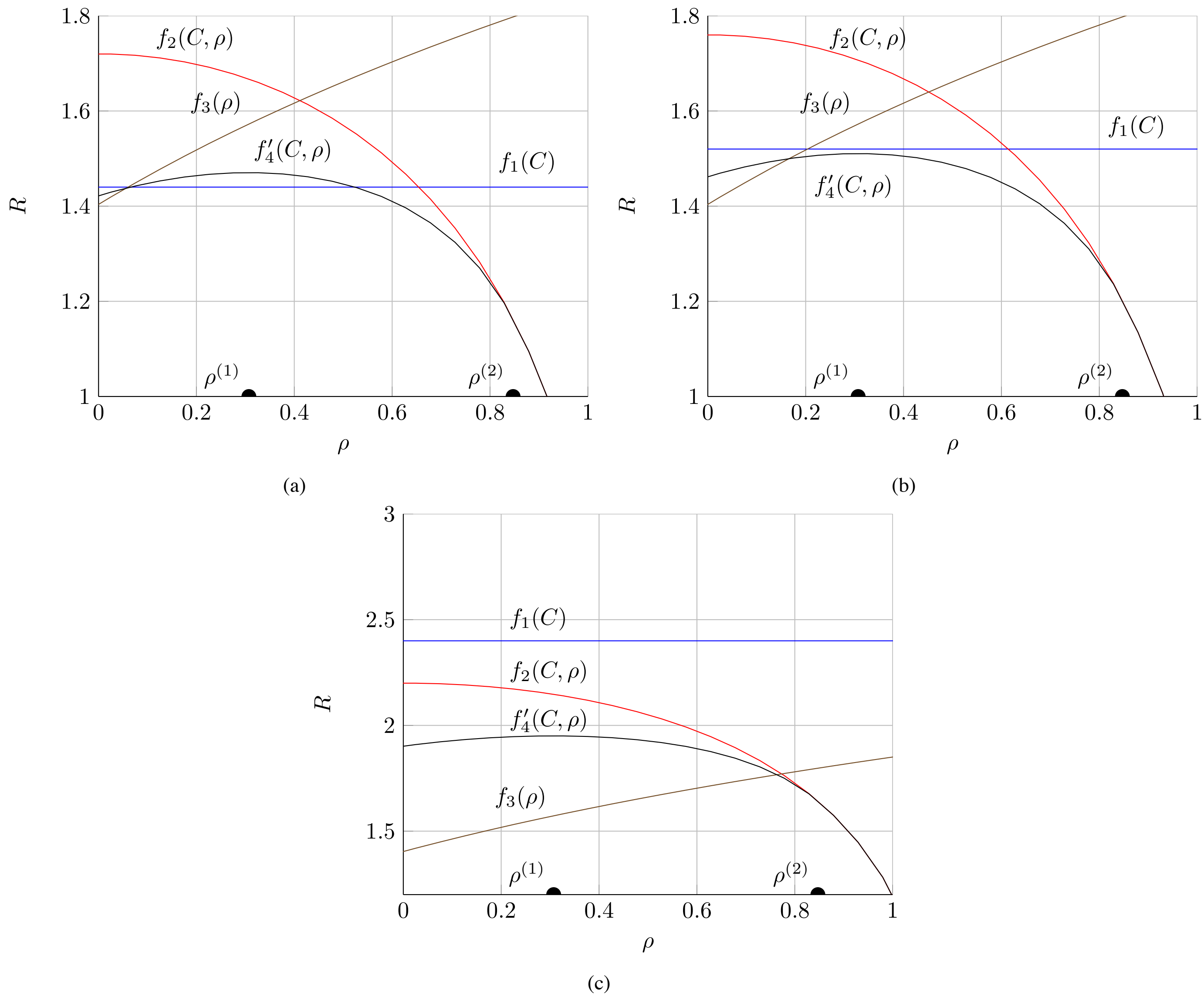


Fig. 7: The functions in (132) for different values of C where C increases from Fig. 7a to Fig. 7c. Note the range of the $R-$axis in Fig. 7c.

and thus $g(C,\rho)$ is increasing in ρ. So $g(C,\rho)$ has exactly one zero.

Remark 25. *In the regime*

$$2C \leq f_4'(C,\rho^{(1)}) \tag{135}$$

we have

$$\begin{aligned} f_3(\rho^{(1)}) &= \frac{1}{2}\log\left(1+2P(1+\rho^{(1)})\right) \\ &\geq \frac{1}{2}\log\left(1+2P(1+\rho^{(1)})\right) - \frac{1}{2}\log\left(\frac{1}{1-\left(\rho^{(1)}\right)^2}\right) \\ &= 2f_4'(C,\rho^{(1)}) - 2C \\ &\geq f_4'(C,\rho^{(1)}). \end{aligned} \tag{136}$$

Inequality (136) *follows by* (135)*. The implication is that $f_3(\rho)$ is not "limiting" in* (133) *for the the regime given by* (135)*.*

Fix the value of C. Define $\rho^\star$ to be the optimal correlation coefficient in Corollary 1 and let $R_{\max}$ be the maximum value it attains. We have one of the following cases.

(a) $\rho^\star$ is such that $R_{\max} = f_1(C)$.

(b) $\rho^\star$ is the unique ρ that maximizes $f_4'(C,\rho)$ and $R_{\max} = f_4'(C,\rho^{(1)})$.
(c) $\rho^\star$ is such that $R_{\max} = f_3(\rho^\star)$.

When $C = 0$, we are in case (a). As C increases, $f_3(\rho)$ remains unchanged but $f_1(C)$ and $f_4'(C,\rho)$ increase. We remain in case (a) as long as $2C \le f_4'(C,\rho^{(1)})$. This is illustrated in Fig. 7a. We then transit to case (b) where $R_{\max} = f_4'(C,\rho^{(1)})$, see Fig. 7b. To see this, keep increasing C until $2C = f_4'(C,\rho^{(1)})$. At this point, we are about to leave case (a), and we still have $f_3(\rho^{(1)}) \ge f_4'(C,\rho^{(1)})$ (see Remark 25). Moreover, $f_4'(C,\rho)$ is decreasing in ρ for $\rho > \rho^{(1)}$, and $f_3(\rho)$ is increasing in ρ. So the crossing point of the two curves should be at or before $\rho^{(1)}$. As C further increases, $\rho^\star$ remains equal to $\rho^{(1)}$ until $f_3(\rho^{(1)}) = f_4'(C,\rho^{(1)})$. From that point on, we have

$$C \ge \frac{1}{4}\log\frac{1+2P\left(1+\rho^{(1)}\right)}{1-{\rho^{(1)}}^2} \tag{137}$$

and we move into case (c). In this case, $\rho^\star$ is such that $R_{\max} = f_3(\rho^\star) = f_4'(C,\rho^\star)$ (see Fig. 7c). We note that as C increases, so does $\rho^\star$. We thus have $\rho^\star \le \rho^{(2)}$ if

$$C \le \frac{1}{4}\log\frac{1+2P\left(1+\rho^{(2)}\right)}{1-{\rho^{(2)}}^2}. \tag{138}$$

In this regime, besides the bounds (47)-(50), we have

$$\begin{aligned} R_{\max} &= f_3(\rho^\star) \\ &= 2f_4'(C,\rho^\star) - f_3(\rho^\star) \\ &= \begin{cases} 2C - \frac{1}{2}\log\left(\frac{1}{1-{\rho^\star}^2}\right) & \rho^\star \le \rho^{(2)} \\ 2f_2(C,\rho^\star) - f_3(\rho^\star) & \rho^\star > \rho^{(2)} \end{cases}. \end{aligned} \tag{139}$$

Therefore, when C satisfies (137) and (138) the upper bound meets the lower bound of (32)-(35).

Appendix F
Proof of Theorem 5

Consider (66) and suppose $C^\diamond$ is the capacity of the network. For symmetric networks, we thus have

$$\begin{aligned} C^\diamond &\le 2C - \frac{1}{2}\log\left(N + 2^{2C^\diamond}\right) - \frac{1}{2}\log\left(1+N\right) + \log\left(1+N+P\left(1-\rho^2\right)\right) \\ &\overset{(a)}{\le} 2C - \frac{1}{2}\log\left(N + 2^{2R_{\max}^{(l)}}\right) - \frac{1}{2}\log\left(1+N\right) + \log\left(1+N+P\left(1-\rho^2\right)\right) \end{aligned} \tag{140}$$

where $R_{\max}^{(l)}$ is the maximum admissible rate in the lower bound of (32)-(35). Note that (a) holds because $R_{\max}^{(l)} \le C^\diamond$. The upper bound of Theorem 3 is thus loosened to:

$$R_{\max}^{(u)} = \max_{0\le\rho\le1}\min_{N}\min\left\{f_1(C), f_2(C,\rho), f_3(\rho), f_5(C,\rho,N)\right\} \tag{141}$$

where f_1, f_2, f_3 are defined in (132) and $f_5(C,\rho,N)$ is the RHS of (140). Furthermore, define

$$f_0(C,\rho) = 2C - \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right) \tag{142}$$

so that we have

$$R_{\max}^{(l)} = \max_{0\le\rho\le1}\min\left\{f_0(C,\rho), f_2(C,\rho), f_3(\rho)\right\}. \tag{143}$$

We shall prove that (143) is equal to (141) for the range of C given in (68). We start with (143). Fix C and let λ be the optimizing correlation coefficient. The functions $f_0(C,\rho)$ and $f_2(C,\rho)$ are decreasing in ρ and $f_3(\rho)$ is increasing in ρ. So depending on how $f_3(0)$ compares with $\min(f_0(C,0), f_2(C,0))$ we have the following cases for λ:

- If $2C \le \frac{1}{2}\log(1+2P)$, then we have $\lambda = 0$ and $R = 2C$ is achievable using independent Gaussian random variables X_1, X_2 with zero mean and variance P.
- If $2C > \frac{1}{2}\log(1+2P)$, then λ is such that either $R_{\max}^{(l)} = f_3(\lambda) = f_2(C,\lambda)$ (where the cut-set bound is achievable) or $R_{\max}^{(l)} = f_3(\lambda) = f_0(C,\lambda)$. We show that in the latter case $R_{\max}^{(l)}$ and $R_{\max}^{(u)}$ are equal if (68) is satisfied.

Suppose λ is such that

$$R_{\max}^{(l)} = f_3(\lambda) = f_0(C,\lambda). \tag{144}$$

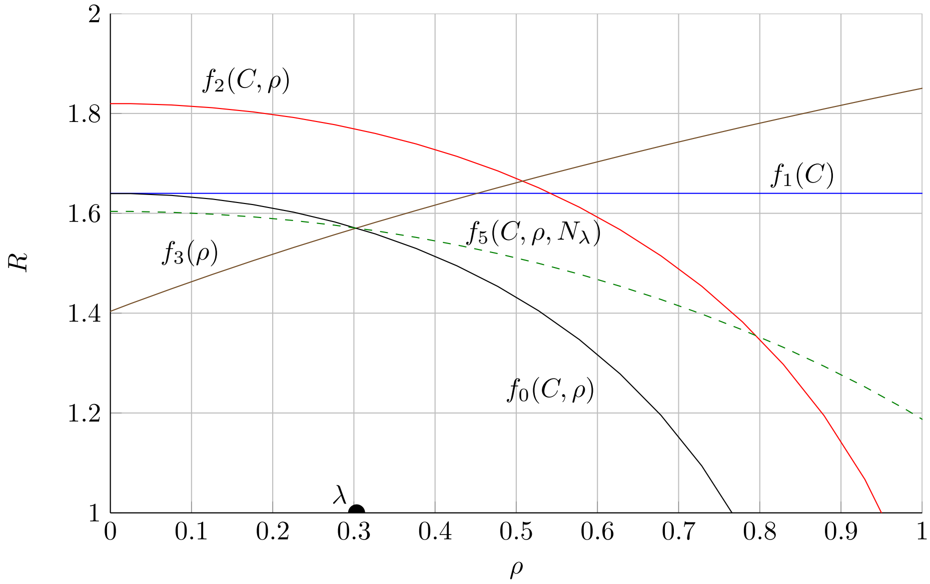


Fig. 8: Several functions of ρ. At $\rho = \lambda$, the curves $f_0(C,\rho)$, $f_3(\rho)$, and $f_5(C,\rho,N_\lambda)$ all have the same value.

Here λ is defined by the crossing point of $f_3(\rho)$ and $f_0(C,\rho)$ and is such that

$$C = \frac{1}{4}\log\left(\frac{1+2P(1+\lambda)}{1-\lambda^2}\right). \tag{145}$$

The RHS of (145) is an increasing function of λ and thus we have $\lambda \le \rho^{(2)}$ if and only if (68) is satisfied.

Next consider (141) in the regime of C given by (68). First note that for a fixed N, $f_5(C,\rho,N)$ is decreasing in ρ. Let

$$N_\lambda = P\left(\frac{1}{\lambda} - \lambda\right) - 1. \tag{146}$$

Since $\lambda \le \rho^{(2)}$, we have $N_\lambda \ge 0$ and the upper bound may be written as follows:

$$\begin{aligned} R \le& f_5(C,\rho,N_\lambda) \\ =& 2C - \frac{1}{2}\log\left(N_\lambda + 2^{2R_{\max}^{(l)}}\right) - \frac{1}{2}\log\left(1+N_\lambda\right) + \log\left(1+N_\lambda+P\left(1-\rho^2\right)\right) \\ \overset{(a)}{=}& 2C - \frac{1}{2}\log\left(1+N_\lambda+2P(1+\lambda)\right) - \frac{1}{2}\log\left(1+N_\lambda\right) + \log\left(1+N_\lambda+P\left(1-\rho^2\right)\right) \end{aligned} \tag{147}$$

where (a) holds by (144). The RHS of (147), evaluated for $\rho = \lambda$, is given by

$$f_5(C,\lambda,N_\lambda) = 2C - \frac{1}{2}\log\left(\frac{1}{1-\lambda^2}\right) = f_0(C,\lambda).$$

This follows by the argument in (42)-(45). We conclude that $f_5(C,\rho,N_\lambda)$ is equal to $f_0(C,\lambda) = f_3(\lambda)$ at $\rho = \lambda$. Since $f_5(C,\rho,N_\lambda)$ is decreasing in ρ, λ is the optimal $\rho^\star$ in (141) too. This is illustrated in Fig. 8. So in the regime characterized by (68) the upper bound is equal to $R_{\max}^{(l)}$ and is thus achievable.

Remark 26. *A similar result can be established for asymmetric networks. Let $f_0(C_1,C_2,\rho)$ and $f_3(\rho)$ be the RHSs of* (32) *and* (35)*, respectively. Define λ to be the optimizing correlation coefficient in* (32)-(35) *and $R_{\max}^{(l)}$ as the maximum achievable rate. One can check that $R_{\max}^{(l)}$ is equal to the cut-set bound if C_1, C_2 satisfy $C_1 + C_2 \le \frac{1}{2}\log\left(1+P_1+P_2\right)$, or if* (72) *or* (73) *are satisfied. The cut-set bound may not be achievable by* (32)-(35) *when we have*

$$R_{\max}^{(l)} = f_3(\lambda) = f_0(C_1,C_2,\lambda). \tag{148}$$

For C_1, C_2 where (71) *is satisfied, we have*

$$\lambda \le \rho^{(2)}. \tag{149}$$

Therefore, following the steps in (146)-(147)*, we have a matching upper bound based on Theorem 3 (in the form of* (141) *but written for general C_1, C_2). Finally, for C_1, C_2 that satisfy* (74) *the cut-set bound is achievable using Theorem 1 with $\frac{U}{\sqrt{P_1}} = \frac{X_1}{\sqrt{P_1}} = \frac{X_2}{\sqrt{P_2}} \sim \mathcal{N}(0,1)$.*

Appendix G
Proof of Lemma 2

Consider the following optimization problem:

$$\max_{p(x_1,x_2)} \min_{p(u|y)} \min \left\{ \begin{array}{l} 2C \\ C + I(X_2;Y|X_1) \\ C + I(X_1;Y|X_2) \\ I(X_1X_2;Y) \\ \frac{1}{2}\left(2C + I(X_1X_2;Y|U) + I(X_1;U|X_2) + I(X_2;U|X_1)\right) \end{array} \right\}. \tag{150}$$

We first note that the objective function in (150) is symmetric in $p(x_1, x_2)$ (when $p(u|y)$ is given by the channel in Fig. 5). More precisely, for every pmf $p(x_1, x_2)$, the pmf $\bar{p}(x_1, x_2)$ with $\bar{p}(0, 0) = p(1, 1)$, $\bar{p}(0, 1) = p(1, 0)$, $\bar{p}(1, 0) = p(0, 1)$, $\bar{p}(1, 1) = p(0, 0)$ gives the same objective function. Let $p^{(1)}(x_1, x_2)$ be the pmf that attains the optimal value of (150). Take the pmf $p^{(1)}(x_1, x_2)$ and form the doubly symmetric pmf

$$p^{\star}(x_1, x_2) = \frac{p^{(1)}(x_1, x_2) + \bar{p}^{(1)}(x_1, x_2)}{2}, \quad x_1 = 0, 1, \ x_2 = 0, 1.$$

For a fixed $p(u|y)$, all terms of the min expression in (150) are concave functions of $p(x_1, x_2)$ (see Remark 3). Therefore, at any point $(p^{\star}(x_1, x_2), p(u|y))$ they take on values larger than or equal to their respective values at $(p^{(1)}(x_1, x_2), p(u|y))$ (or $(\bar{p}^{(1)}(x_1, x_2), p(u|y))$). This proves that there exists at least one optimizing doubly symmetric pmf $p(x_1, x_2)$ in (150).

Appendix H
Proof of Theorem 7

The proof is similar to the proof of Theorem 4. The lower and upper bounds match for $C \leq 0.75$ and $C \geq 1$. In the former regime, the cut-set bound is achievable using no cooperation among the relays with $p = \frac{1}{2}$ in (75)-(77). In the latter case, the cut-set bound is achievable using full cooperation among the relays with $p = \frac{1}{3}$ in (75)-(77). We prove the theorem for C's satisfying (85) and we assume $C \leq 1$.

Define

$$\begin{array}{l} g_1(C) = 2C \\ g_3(p) = h_2(p) + 1 - p \\ g_4(C, p) = C + h_2(p) - \frac{p}{2} \end{array} \tag{151}$$

so that the upper bound of Corollay 2 is given by

$$R_{\max} = \max_{0 \leq p \leq \frac{1}{2}} \min\{g_1(C), g_3(p), g_4(C, p)\}. \tag{152}$$

The functions $g_1(C)$, $g_3(p)$, and $g_4(C, p)$ are drawn in Fig. 9 as functions of $p \leq \frac{1}{2}$ for different values of C, where C increases from Fig. 9a to Fig. 9d. We consider two probabilities:

$$p^{(1)} = \frac{1}{1 + \sqrt{2}} \tag{153}$$

$$p^{(3)} = 2(1 - C). \tag{154}$$

Remark 27. *$g_4(C, p)$ is concave in p and it attains its maximum at $p = p^{(1)}$. $g_3(p)$ is also concave and it attains its maximum at $p = \frac{1}{3}$. We have $\frac{1}{3} \leq p^{(1)}$. $g_3(p)$ and $g_4(C, p)$ cross at $p = p^{(3)}$.*

Remark 28. *For the regime*

$$2C \leq g_4(C, p^{(1)}) \tag{155}$$

we have

$$C \leq h_2(p^{(1)}) - \frac{p^{(1)}}{2} < 1 - \frac{p^{(1)}}{2}. \tag{156}$$

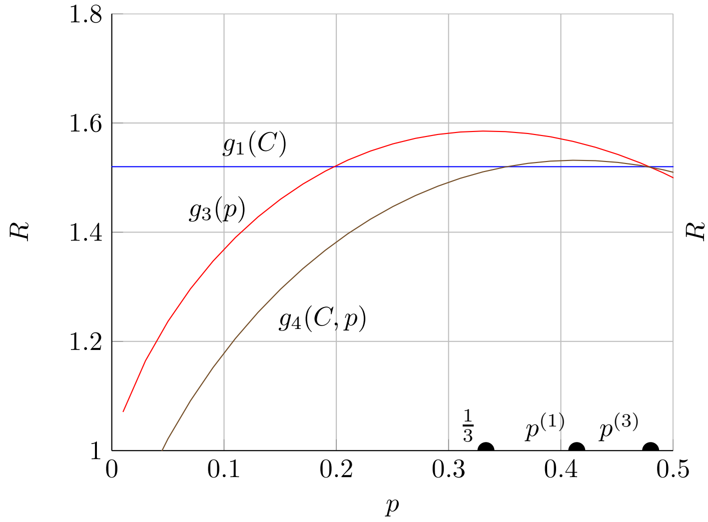
As a result, we have

$$g_4(C, p^{(1)}) < g_3(p^{(1)}).$$

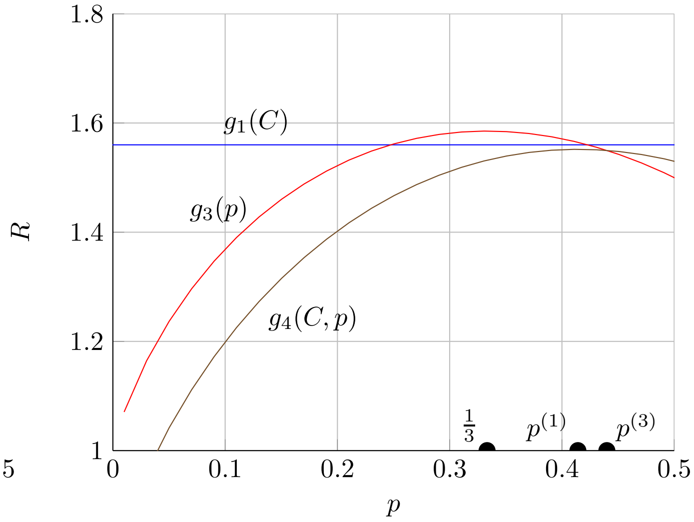
The implication is that $g_3(p)$ is not "limiting" for the regime given by (155).

Fix the value of C and denote the maximizing p in (152) by $p^{\star}$. We have one of the following cases:

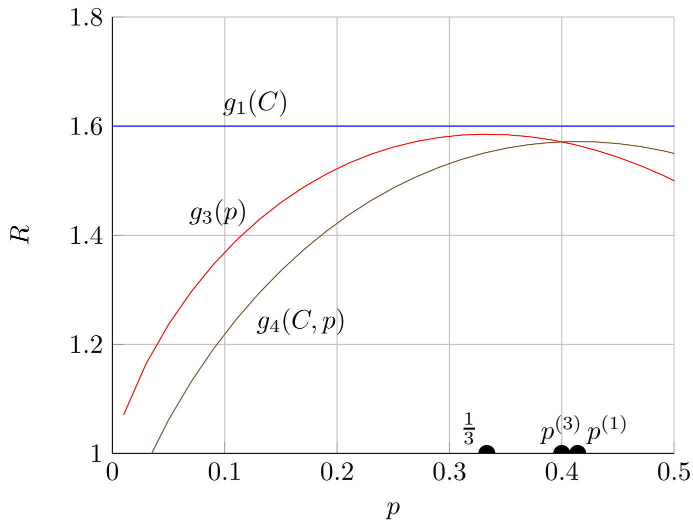
(a) $p^{\star}$ is such that $R_{\max} = g_1(C)$.



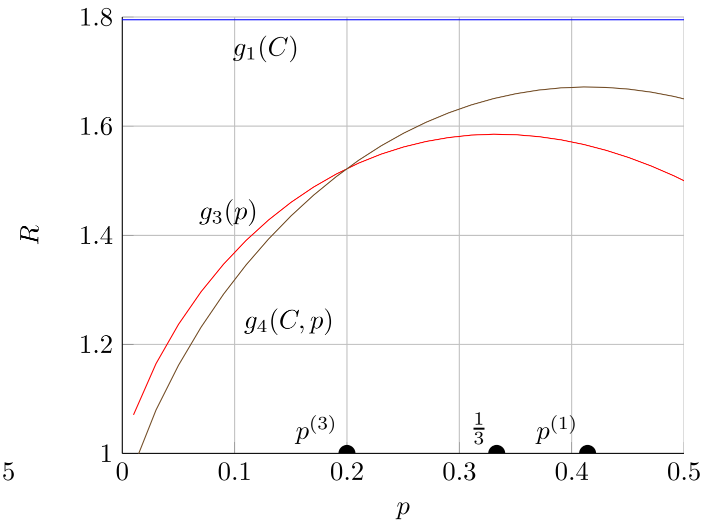
(a)



(b)



(c)



(d)

Fig. 9: The functions in (151) for different values of C.

(b) $p^\star = p^{(1)}$ and $R_{\max} = g_4(C, p^{(1)})$.
(c) $p^\star = p^{(3)}$ and $R_{\max} = g_3(p^\star) = g_4(C, p^\star)$.
(d) $p^\star = \frac{1}{3}$ and $R_{\max} = g_3(\frac{1}{3})$.

When $C = 0$, we are in case (a). We remain in this case as long as $2C \leq g_4(C, \rho^{(1)})$. When $2C = g_4(C, p^{(1)})$ we have $g_4(C, p^{(1)}) < g_3(p^{(1)})$, see Remark 28. So as C increases we have $p^\star = p^{(1)}$ and we move into case (b), see Fig. 9b. In this regime, $g_3(p)$ and $g_4(C, p)$ cross at the point $p^{(3)}$ which is larger than or equal to $p^{(1)}$. As C further increases, the crossing point $p^{(3)}$ of $g_3(p)$ and $g_4(C, p)$ decreases towards $p^{(1)}$, and as soon as $g_3(p^{(1)}) = g_4(C, p^{(1)})$ we move into case (c) where $p^\star = p^{(3)}$, see Fig. 9c. In this case, we have

$$R_{\max} = g_3(p^\star) = g_4(C, p^\star). \tag{157}$$

Using (157), we obtain

$$\begin{aligned} R_{\max} &= 2g_4(C, p^\star) - g_3(p^\star) \\ &= 2C + h_2(p^\star) - 1. \end{aligned} \tag{158}$$

So the upper bound reduces to

$$R \leq C + h_2(p^\star) \tag{159}$$

$$R \leq h_2(p^\star) + 1 - p^\star \tag{160}$$
$$R \leq 2C + h_2(p^\star) - 1 \tag{161}$$

which matches the lower bound in (75)-(77). Finally, when $p^{(3)} \leq \frac{1}{3}$ we have $p^\star = \frac{1}{3}$ and full cooperation is achieved. We thus meet the cut-set bound. This concludes the proof.

APPENDIX I
AN UPPER BOUND ON $H(U|X_1, Q = i)$ AND $H(U|X_2, Q = i)$

$H(U|X_1, Q = i)$ may be expanded as follows:

$$\begin{aligned}
&H(U|X_1, Q = i)\\
&= p_{X_1|Q}(0|i) h_2\left(\frac{p_{X_1X_2|Q}(0,0|i)(1-\alpha) + \frac{p_{X_1X_2}(0,1)}{2}}{p_{X_1|Q}(0|i)}\right) + p_{X_1|Q}(1|i) h_2\left(\frac{p_{X_1X_2|Q}(1,1|i)(1-\alpha) + \frac{p_{X_1X_2}(1,0)}{2}}{p_{X_1|Q}(1|i)}\right)\\
&\leq h_2\left((1-q_i)(1-\alpha) + \frac{q_i}{2}\right)\\
&= h_2(\alpha \star \frac{q_i}{2})
\end{aligned} \tag{162}$$

where the inequality is by the concavity of $h_2(.)$ in its argument. $H(U|X_2Q)$ may be bounded similarly.

APPENDIX J
PROOF OF LEMMA 3

First, note that for $x \leq h_2(y)$, we have $g(x, y) = 0$. For $x \geq h_2(y)$, $g(x, y)$ is non-negative. Since $g(x, y)$ is continuous at $x = h_2(y)$, it suffices to prove convexity of $g(x, y)$ in the regime $x \geq h_2(y)$. Recall that $y \leq 1$ and $\alpha \leq \frac{1}{2}$. We prove that the Hessian matrix

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 g(x,y)}{\partial x^2} & \frac{\partial^2 g(x,y)}{\partial x \partial y} \\ \frac{\partial^2 g(x,y)}{\partial y \partial x} & \frac{\partial^2 g(x,y)}{\partial y^2} \end{bmatrix}. \tag{163}$$

is positive semi-definite. Using Sylvester's criterion [32, Theorem 7.2.5], $\mathbf{H}$ is positive semi-definite if and only if its leading principal minors are non-negative; i.e., if

$$\frac{\partial^2 g(x,y)}{\partial x^2} \geq 0 \tag{164}$$
$$\frac{\partial^2 g(x,y)}{\partial x^2}\frac{\partial^2 g(x,y)}{\partial y^2} - \frac{\partial^2 g(x,y)}{\partial x \partial y}\frac{\partial^2 g(x,y)}{\partial y \partial x} \geq 0. \tag{165}$$

We use the following notation:

$$z = h_2^{-1}\left(\frac{x - h_2(y)}{1-y}\right), \quad y \leq 1 \tag{166}$$
$$s = \alpha \star \left(\frac{y}{2} + (1-y)z\right) \tag{167}$$
$$r(x) = -\frac{h_2'(x)}{h_2''(x)} = x(1-x)\ln\frac{1-x}{x} \qquad 0 \leq x \leq \frac{1}{2} \tag{168}$$
$$h_2'(x) = \frac{1}{\ln 2}\ln\frac{1-x}{x} \qquad 0 \leq x \leq \frac{1}{2} \tag{169}$$
$$h_2''(x) = -\frac{1}{\ln 2}\frac{1}{x(1-x)} \qquad 0 \leq x \leq \frac{1}{2}. \tag{170}$$

Taking the partial derivatives, we have

$$\begin{aligned}
\frac{\partial^2 g(x,y)}{\partial x^2} &= \frac{(1-2\alpha)}{(1-y)\left(h_2'(z)\right)^3}\left(-h_2''(z)h_2'(s) + (1-2\alpha)(1-y)h_2''(s)h_2'(z)\right)\\
&= \frac{(1-2\alpha)h_2''(z)h_2''(s)}{(1-y)\left(h_2'(z)\right)^3}\left(r(s) - (1-2\alpha)(1-y)r(z)\right)
\end{aligned} \tag{171}$$

$$\begin{aligned}
&\frac{\partial^2 g(x,y)}{\partial x^2}\frac{\partial^2 g(x,y)}{\partial y^2} - \frac{\partial^2 g(x,y)}{\partial x \partial y}\frac{\partial^2 g(x,y)}{\partial y \partial x}\\
&= \frac{(1-2\alpha)^2}{\left(h_2'(z)\right)^3(1-y)}\left(\begin{array}{l} -\frac{(1-2\alpha)}{4}\left(-h_2''(z)h_2'(s) + (1-2\alpha)(1-y)h_2''(s)h_2'(z)\right)h_2''(\alpha \star \frac{y}{2}) \\ +\frac{h_2''(z)}{h_2'(z)}h_2''(y)\left(h_2'(s)\right)^2 \\ +(1-2\alpha)\left(-(\frac{1}{2}-z)^2 h_2''(z) - (1-y)h_2''(y)\right)h_2'(s)h_2''(s) \end{array}\right).
\end{aligned} \tag{172}$$

We shall prove that (171) and (172) are non-negative for all non-negative parameters $\alpha \leq \frac{1}{2}$, $y \leq 1$, $z \leq \frac{1}{2}$. We start with (171). We claim that

$$r(s) - (1-2\alpha)(1-y)r(z) \tag{173}$$

is non-negative because it is equal to 0 at $z = \frac{1}{2}$ and is a non-increasing function of z, $0 \leq z \leq \frac{1}{2}$. To see this, consider

$$\frac{\partial}{\partial z}\left(r(s) - (1-2\alpha)(1-y)r(z)\right) = (1-2\alpha)(1-y)\left(r'(s) - r'(z)\right) \tag{174}$$

where

$$r'(x) = (1-2x)\ln\frac{1-x}{x} - 1. \tag{175}$$

Since $r'(x)$ is non-increasing in x and $s \geq z$, (174) is non-positive. This line of argument is very similar to [20, Theorem 2] and [29, Theorem 2].

To prove the non-negativity of (172) we proceed as follows. We have $h_2''(\alpha \star \frac{y}{2}) \leq h_2''(s)$ because $h_2''(x)$ is increasing in x, $0 \leq x \leq \frac{1}{2}$, and $\alpha \star \frac{y}{2} \leq s$. We thus have

$$\begin{aligned}
&\frac{\partial^2 g(x,y)}{\partial x^2}\frac{\partial^2 g(x,y)}{\partial y^2} - \frac{\partial^2 g(x,y)}{\partial x \partial y}\frac{\partial^2 g(x,y)}{\partial y \partial x} \\
&\geq \frac{(1-2\alpha)^2}{\left(h_2'(z)\right)^3(1-y)}\left(\begin{array}{l} -\frac{(1-2\alpha)}{4}\left(-h_2''(z)h_2'(s) + (1-2\alpha)(1-y)h_2''(s)h_2'(z)\right)h_2''(s) \\ +\frac{h_2''(z)}{h_2'(z)}h_2''(y)\left(h_2'(s)\right)^2 \\ +(1-2\alpha)\left(-(\frac{1}{2}-z)^2 h_2''(z) - (1-y)h_2''(y)\right)h_2'(s)h_2''(s) \end{array}\right) \\
&= \frac{(1-2\alpha)^4\left(h_2''(s)\right)^2 h_2''(z)h_2''(y)}{\left(h_2'(z)\right)^4(1-y)}\left(-\frac{1}{4}y(1-y)^2\ln\left(\frac{1-z}{z}\right)r(z) + \frac{(r(s))^2}{(1-2\alpha)^2} - (1-y)^2\frac{r(s)r(z)}{1-2\alpha}\right).
\end{aligned} \tag{176}$$

It suffices to prove the non-negativity of

$$t(\alpha,y,z) = -\frac{1}{4}y(1-y)^2\ln\left(\frac{1-z}{z}\right)r(z) + \frac{(r(s))^2}{(1-2\alpha)^2} - (1-y)^2\frac{r(s)r(z)}{1-2\alpha} \tag{177}$$

for every non-negative parameter $\alpha \leq \frac{1}{2}$, $y \leq 1$, $z \leq \frac{1}{2}$. First we show that $t(\alpha,y,z)$ is non-decreasing in α, and conclude that $t(\alpha,y,z)$ is non-negative if and only if $t(0,y,z)$ is non-negative. By taking the derivative of $t(\alpha,y,z)$ with respect to α we obtain

$$\frac{\partial t(\alpha,y,z)}{\partial \alpha} = \frac{\left(2r(s) - (1-2\alpha)(1-y)^2 r(z)\right)\left(2r(s) + (1-2\alpha)(1-y)(1-2z)r'(s)\right)}{(1-2\alpha)^3}. \tag{178}$$

Both terms in the numerator are non-negative. The first term is non-negative because it is larger than (173) and the second term is non-negative because it is a non-increasing function of z, $0 \leq z \leq \frac{1}{2}$, and is equal to 0 at $z = \frac{1}{2}$.

Finally, we plot

$$f(y,z) = \frac{t(0,y,z)}{(1-y)^2(1-2z)^4}$$

in Fig. 10 and demonstrate that the function is non-negative over $0 \leq y \leq 1$ and $0 \leq z \leq \frac{1}{2}$. The terms in the denominator of $f(y,z)$ capture the behaviour of $t(0,y,z)$ around $z = \frac{1}{2}$ and $y = 1$. $f(y,z)$ is zero at $y = 0$ with a strictly positive slope at $y = 0$ for all z, $0 < z < \frac{1}{2}$.

Appendix K
Proof of Theorem 8

Consider (75)-(77) and let $R_{\max}^{(l)}$ be the maximum rate admissible. We have

$$R_{\max}^{(l)} = \max_{0 \leq p \leq 1} \min\{g_0(C_1, C_2, p), g_2(C_1, p), g_3(p)\} \tag{179}$$

where

$$g_0(C_1, C_2, p) = C_1 + C_2 - 1 + h_2(p) \tag{180}$$

$$g_2(C_1, p) = C_1 + h_2(p) \tag{181}$$

$$g_3(p) = h_2(p) + 1 - p. \tag{182}$$

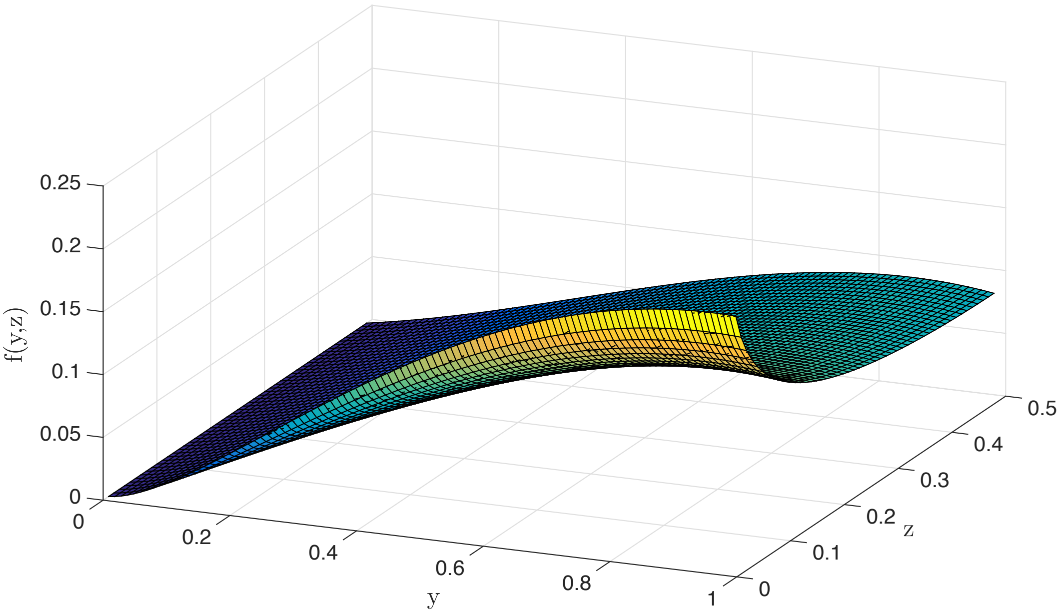


Fig. 10: The function $f(y,z)$ is non-negative for y,z, $0 \le y \le 1$, $0 \le z \le \frac{1}{2}$.

Consider next (91)-(93), (99) with $q = P_Y(1)$ as in (90). Let $R_{\max}^{(u)}$ be the maximum admissible rate. Since the function $h_2(\alpha \star \left(\frac{q}{2} + (1-q)h_2^{-1}(x)\right))$ is non-decreasing in x, we have

$$R_{\max}^{(u)} \le C_1 + C_2 - h_2\left(\alpha \star \left(\frac{q}{2} + (1-q)h_2^{-1}\left(\min\left(1, \frac{\left(R_{\max}^{(l)} - h_2(q)\right)^+}{1-q}\right)\right)\right)\right) - (1-q)h_2(\alpha) - q + 2h_2\left(\alpha \star \frac{q}{2}\right). \tag{183}$$

So the capacity is upper bounded by

$$R_{\max}^{(u)} \le \max_{0\le q\le 1} \min_{0\le\alpha\le\frac{1}{2}} \min\{g_1(C_1,C_2), g_2(C_1,q), g_3(q), g_5(C_1,C_2,q,\alpha)\} \tag{184}$$

where $g_1(C_1,C_2) = C_1 + C_2$ and $g_5(C_1,C_2,q,\alpha)$ is the RHS of (183). Since (184) depends only on $q = p_{X_1X_2}(0,1) + p_{X_1X_2}(1,0)$, we may assume, without loss of generality, that $p(x_1,x_2)$ is a doubly symmetric binary pmf with parameter q. Using (79) the bound in (183) may be re-written as

$$g_5(C_1,C_2,q,\alpha) = g_0(C_1,C_2,q) + I(X_1;X_2|U) + 1 - h_2\left(\alpha \star \left(\frac{q}{2} + (1-q)h_2^{-1}\left(\min\left(1, \frac{\left(R_{\max}^{(l)} - h_2(q)\right)^+}{1-q}\right)\right)\right)\right) \tag{185}$$

where $I(X_1;X_2|U)$ is a function of α and q. We denote this conditional mutual information by $I_{\alpha,q}(X_1;X_2|U)$.

Consider (179) and let η be the optimizing p. We have $\eta \le \frac{1}{2}$. Note that when $C_2 \ge 1$, $g_0(C_1,C_2,p)$ is redundant in (179) and the cut-set bound is achievable. Otherwise, $g_2(C_1,p)$ is redundant and we have one of the following cases:

(a) $\eta = \frac{1}{3}$ and $R_{\max}^{(l)} = g_3(\frac{1}{3})$. In this case, $R_{\max}^{(l)} = \log_2(3)$ and the cut-set bound is achievable.
(b) $\eta = \frac{1}{2}$ and $R_{\max}^{(l)} = g_0(C_1,C_2,\frac{1}{2})$. In this case, $R_{\max}^{(l)} = C_1 + C_2$ and the cut-set bound is achievable.
(c) η is such that $R_{\max}^{(l)} = g_0(C_1,C_2,\eta) = g_3(\eta)$. We show that $R_{\max}^{(l)}$ and $R_{\max}^{(u)}$ match in this case. Here, we have

$$\eta = 2 - C_1 - C_2 \tag{186}$$
$$R_{\max}^{(l)} = g_0(C_1,C_2,\eta) = g_3(\eta). \tag{187}$$

Consider (184) in regime (c) and let α_η be the solution of

$$\alpha_\eta(1-\alpha_\eta) = \left(\frac{\eta}{2(1-\eta)}\right)^2 \tag{188}$$

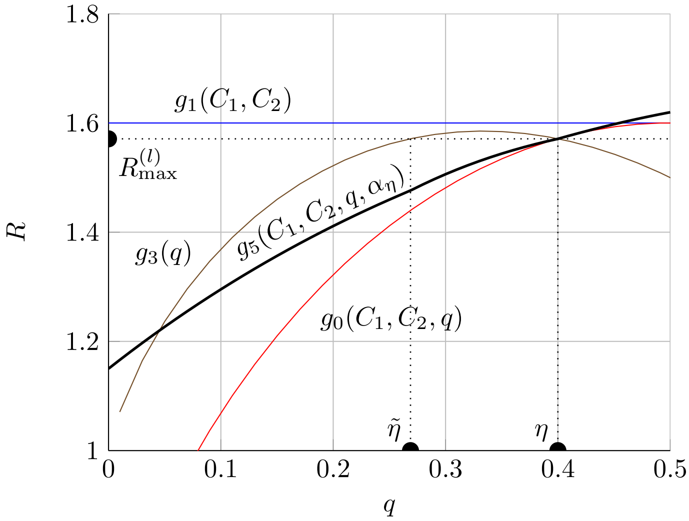


Fig. 11: Several functions of q. At $q = \eta$, the curves $g_0(C_1, C_2, q)$, $g_3(q)$, and $g_5(C_1, C_2, q, \alpha_\eta)$ all have the same value.

that is less than or equal to $\frac{1}{2}$ where η is given by (186). With this choice of α, we study $g_5(C_1, C_2, q, \alpha_\eta)$. At $q = \eta$, $X_1 - U - X_2$ forms a Markov chain and using (187) we have

$$g_5(C_1, C_2, \eta, \alpha_\eta) = g_0(C_1, C_2, \eta). \tag{189}$$

In regime (c) given by (187), $g_3(q)$ and $g_0(C_1, C_2, q)$ cross at $q = \eta$. At this point, $g_3(q)$ is decreasing and $g_0(C_1, C_2, q)$ is increasing in q. Also, $g_5(C_1, C_2, q, \alpha_\eta)$ crosses the two curves at $q = \eta$ as shown in (189). Therefore, if $g_5(C_1, C_2, q, \alpha_\eta)$ is non-decreasing in q, $q \leq \eta$, then η maximizes (184) and $R^{(u)}_{\max} = R^{(l)}_{\max}$, see Fig. 11.

It remains to show that $g_5(C_1, C_2, q, \alpha_\eta)$ is non-decreasing in q, $q \leq \eta$. $g_5(C_1, C_2, q, \alpha_\eta)$ is continuous and piece-wise concave. We thus look at the following two regimes and prove that $g_5(C_1, C_2, q, \alpha_\eta)$ is non-decreasing in both regimes: $q \leq \tilde{\eta}$ and $\tilde{\eta} \leq q \leq \eta$ where $\tilde{\eta}$ and η are the two solutions of $R^{(l)}_{\max} = 1 + h_2(q) - q$.

- $q \leq \tilde{\eta}$: Here, we have $R^{(l)}_{\max} \geq 1 + h_2(q) - q$ and

$$\begin{aligned} g_5(C_1, C_2, q, \alpha_\eta) &= g_0(C_1, C_2, q) + I_{\alpha_\eta, q}(X_1; X_2|U) \\ &= C_1 + C_2 - (1-q)h_2(\alpha_\eta) - q + 2h_2\left(\alpha_\eta \star \frac{q}{2}\right) - 1. \end{aligned} \tag{190}$$

The RHS of (190) is concave in q. By showing that this function is non-decreasing at $q = \eta$ we prove that it is non-decreasing in q, $q \leq \eta$. We first show that $I_{\alpha_\eta, q}(X_1; X_2|U)$ has a zero derivative at $q = \eta$:

$$\begin{aligned} \frac{\partial I_{\alpha_\eta, q}(X_1; X_2|U)}{\partial q} &= h_2(\alpha_\eta) - 1 - \log\left(\frac{1-q}{q}\right) - (1 - 2\alpha_\eta)\log\left(\frac{\alpha_\eta + \frac{q}{2(1-q)}}{(1-\alpha_\eta) + \frac{q}{2(1-q)}}\right) \\ &= h_2(\alpha_\eta) - 1 - \log\left(\frac{1-q}{q}\right) - (1 - 2\alpha_\eta)\log\left(\frac{\alpha_\eta + \frac{q}{2(1-q)}}{(1-\alpha_\eta) + \frac{q}{2(1-q)}} \times \frac{\frac{q}{2(1-q)} - \alpha_\eta}{\frac{q}{2(1-q)} - \alpha_\eta}\right) \\ &= h_2(\alpha_\eta) - 1 - \log\left(\frac{1-q}{q}\right) - (1 - 2\alpha_\eta)\log\left(\frac{(\frac{q}{2(1-q)})^2 - \alpha_\eta^2}{(\frac{q}{2(1-q)})^2 - \alpha_\eta(1-\alpha_\eta) + \frac{q}{2(1-q)}(1-2\alpha_\eta)}\right). \end{aligned} \tag{191}$$

We use (188) to write

$$\begin{aligned} \left.\frac{\partial I_{\alpha_\eta, q}(X_1; X_2|U)}{\partial q}\right|_{q=\eta} &= h_2(\alpha_\eta) - \log\left(\frac{2(1-\eta)}{\eta}\right) - (1 - 2\alpha_\eta)\log\left(\frac{\alpha_\eta(1-2\alpha_\eta)}{\frac{\eta}{2(1-\eta)}(1-2\alpha_\eta)}\right) \\ &= h_2(\alpha_\eta) + \frac{1}{2}\log\left(\alpha_\eta(1-\alpha_\eta)\right) - (1-2\alpha_\eta)\log\left(\alpha_\eta\right) + \left(\frac{1}{2} - \alpha_\eta\right)\log(\alpha_\eta(1-\alpha_\eta)) \\ &= 0. \end{aligned} \tag{192}$$

At $q = \eta$, $I_{\alpha_\eta, q}(X_1; X_2|U)$ has a zero derivative and $g_0(C_1, C_2, q)$ is non-decreasing. So the RHS of (190) is non-decreasing at $q = \eta$, and since it is concave, it is also non-decreasing at all q, $q \leq \eta$, and in particular at all q, $q \leq \tilde{\eta}$.

- $\tilde{\eta} \le q \le \eta$: In this regime, we have

$$g_5(C_1, C_2, q, \alpha_\eta) = g_0(C_1, C_2, q) + I_{\alpha_\eta, q}(X_1; X_2|U) + 1 - h_2\left(\alpha_\eta \star \left(\frac{q}{2} + (1-q)h_2^{-1}\left(\frac{\left(R_{\max}^{(l)} - h_2(q)\right)^+}{1-q}\right)\right)\right) \tag{193}$$

$$= C_1 + C_2 - (1-q)h_2(\alpha) - q + 2h_2\left(\alpha \star \frac{q}{2}\right) - h_2\left(\alpha \star \left(\frac{q}{2} + (1-q)h_2^{-1}\left(\frac{\left(R_{\max}^{(l)} - h_2(q)\right)^+}{1-q}\right)\right)\right). \tag{194}$$

The RHS of (194) is concave in q (see Appendix J). Furthermore, (194) is non-decreasing at $q = \eta$. To see this, one can take its derivative with respect to q and evaluate it at $q = \eta - \epsilon$, $\epsilon \to 0$:

$$\begin{aligned}
\lim_{\epsilon \to 0} \left.\frac{\partial g_5(C_1, C_2, q, \alpha_\eta)}{\partial q}\right|_{q=\eta-\epsilon} &\overset{(a)}{=} \log\left(\frac{1-\eta}{\eta}\right) + \left.\frac{\partial I_{\alpha_\eta, q}(X_1; X_2|U)}{\partial q}\right|_{q=\eta} - (1-2\alpha_\eta)^2\left(1 - h_2'(\eta)\right)(1-\eta) \\
&\overset{(b)}{=} \log\left(\frac{1-\eta}{\eta}\right) - \left(1 - \left(\frac{\eta}{1-\eta}\right)^2\right)\left(1 - \log\left(\frac{1-\eta}{\eta}\right)\right)(1-\eta) \\
&= \log\left(\frac{1-\eta}{\eta}\right) - \left(1 - \frac{\eta}{1-\eta}\right)\left(1 - \log\left(\frac{1-\eta}{\eta}\right)\right)
\end{aligned} \tag{195}$$

where (a) follows by differentiating (193) with respect to q and evaluating it at $q = \eta$ and (b) follows by (192) and (188). The function $\log(x) - (1 - \frac{1}{x})(1 - \log(x))$ is equal to 0 at $x = 1$ and is non-decreasing for $x \ge 1$; therefore, (195) is non-negative and $g_5(C_1, C_2, q, \alpha_\eta)$ is non-decreasing in q, $\tilde{\eta} \le q \le \eta$.

REFERENCES

[1] B. E. Schein, *Distributed coordination in network information theory*. PhD Dissertation, Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science., 2001.
[2] D. Traskov and G. Kramer, "Reliable communication in networks with multi-access interference," in *Proc. Inf. Theory Workshop*, Tahoe City, CA, Sept. 2007, pp. 343–348.
[3] W. Kang and N. Liu, "The Gaussian multiple access diamond channel," in *Proc. Int. Symp. Inf. Theory*, St. Petersburg, July 2011, pp. 1499–1503.
[4] W. Kang and S. Ulukus, "Capacity of a class of diamond channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4955–4960, Aug. 2011.
[5] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3008–3023, July 2008.
[6] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
[7] U. Niesen and S. Diggavi, "The approximate capacity of the Gaussian n-relay diamond network," in *IEEE Int. Symp. Inf. Theory*, St. Petersburg, July 2011, pp. 259–263.
[8] A. Sengupta, I.-H. Wang, and C. Fragouli, "Optimizing quantize-map-and-forward relaying for Gaussian diamond networks," in *Proc. Inf. Theory Workshop*, Lausanne, Sept. 2012, pp. 381–385.
[9] B. Chern and A. Ozgur, "Achieving the capacity of the n-relay Gaussian diamond network within $\log(n)$ bits," in *Proc. Inf. Theory Workshop*, Lausanne, Sept. 2012, pp. 377–380.
[10] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Sys. Techn. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.
[11] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, UK: Cambridge University Press, 2011.
[12] S. Saeedi Bidokhti and G. Kramer, "Capacity bounds for a class of diamond networks," in *Proc. Int. Symp. Inf. Theory*, Honolulu, HI, July 2014, pp. 1196–1200.
[13] W. Kang, N. Liu, and W. Chong, "The Gaussian multiple access diamond channel," *CoRR*, vol. abs/1104.3300, 2015. [Online]. Available: http://arxiv.org/abs/1104.3300
[14] F. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 2006.
[16] A. Hekstra and F. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 44–53, Jan. 1989.
[17] J. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 711–716, Sept. 1987.
[18] R. Tandon and S. Ulukus, "Dependence balance based outer bounds for Gaussian networks with cooperation and feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4063–4086, July 2011.
[19] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–i," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.
[20] H. Witsenhausen, "Entropy inequalities for discrete channels," *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 610–616, Sept. 1974.
[21] R. Ahlswede and J. Körner, "On the connection between the entropies of input and output distributions of discrete memoryless channels," in *Proc. 5th Conf. Probability Theory*, Brasov, Sept. 1974, pp. 13–23, (Academy Rep. Soc. Romania, Bucharest, 1977).
[22] N. Chayat and S. Shamai, "Extension of an entropy property for binary input memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1077–1079, Sept. 1989.
[23] S. Shamai and A. Wyner, "A binary analog to the entropy-power inequality," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1428–1430, Nov. 1990.
[24] P. Harremoës and C. Vignat, "An Entropy Power Inequality for the binomial family," *J. Inequal. Pure Appl. Math*, vol. 4, no. 5, p. 10pp., 2003, article 93 (electronic). [Online]. Available: https://hal-upec-upem.archives-ouvertes.fr/hal-00621718
[25] N. Sharma, S. Das, and S. Muthukrishnan, "Entropy power inequality for a family of discrete random variables," in *Proc. Int. Symp. Inf. Theory*, July 2011, pp. 1945–1949.

[26] O. Johnson and Y. Yu, "Monotonicity, thinning, and discrete versions of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5387–5395, Nov. 2010.
[27] S. Haghighatshoar, E. Abbe, and E. Telatar, "A new entropy power inequality for integer-valued random variables," in *Proc. Int. Symp. Inf. Theory*, July 2013, pp. 589–593.
[28] V. Jog and V. Anantharam, "The entropy power inequality and Mrs. Gerber's lemma for groups of order 2^n," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3773–3786, July 2014.
[29] F. Cheng, "Generalization of Mrs. Gerber's lemma," *CoRR*, vol. abs/1409.1668, 2014. [Online]. Available: http://arxiv.org/abs/1409.1668
[30] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.
[31] C. Shannon, "A mathematical theory of communication," *Bell Sys. Techn. J.*, vol. 27, no. 3, pp. 379–423, July 1948, reprinted in Claude Elwood Shannon: Collected Papers, N. J. A. Sloane and A. D. Wyner, Eds. New York: IEEE Press, 1993, pp. 5-83.
[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1985.