Chop and Roll: Improving the Cutset Bound

Sudeep Kamath  
Department of Electrical Engineering  
Princeton University  
Email:sukamath@princeton.edu

Young-Han Kim  
Department of Electrical and Computer Engineering  
University of California, San Diego  
Email: yhk@ucsd.edu

Abstract—A new outer bound on the capacity region of a general noisy network with multiple messages is established. The bound considers an ordered partition of the nodes in the network, and has an intuitive interpretation as the directed information between inputs and outputs across these subsets of the partition. The standard cutset bound is recovered as a special case when the partition consists of two subsets. The new bound extends several existing bounds to the general network that were obtained for special classes of networks. Examples include the generalized network sharing (GNS) bound for graphical networks by Kamath, Tse, and Anantharam, the GNS bound for Gaussian networks by Kamath, Kannan, and Viswanath, and the generalized cutset bound for deterministic networks by Shomorony and Avestimehr. It is demonstrated by a few simple examples that the improvement over the cutset bound can be significant.

I. INTRODUCTION

Characterizing the capacity region of a general noisy network in a computable form is the holy grail of network information theory. This has come to be recognized as a very hard problem and characterizing the capacity region of even simple multiplexer networks such as the broadcast channel has been an open problem for several decades. However, one may attempt to approximately characterize the capacity region by presenting an inner bound (achieved by a coding scheme) and an outer bound (proved by information inequalities) that are close to each other.

So far, the cutset bound [1] has been the only general outer bound available for this problem that makes no assumptions on the underlying network model. The cutset bound considers an arbitrary bipartition of the network into two subsets of nodes and bounds the amount of information flow across from one subset to the other. The form of the bound is simple and intuitively easy to understand. Despite its simplicity, the cutset bound is exactly tight or approximately tight in numerous cases, ranging from the max-flow min-cut theorem [2], two-unicast in undirected graphs [3], and the network coding theorem for multicast [4] and two-level multicast [5] to multicast in linear deterministic and Gaussian networks [6], [7], multiple unicast in random wireless networks [8], [9] and bidirected Gaussian networks [10], broadcast in Gaussian networks [11], [12] and unicast in polylinking systems [13].

In this work, we provide an improvement over the cutset bound for a general memoryless network with multiple messages (flows). We obtain a bound that is at least as tight as the cutset bound but improves on it in general when there is more than one message in the network. We consider an ordered partition of the nodes in the network into multiple subsets (instead of just two as in the cutset bound). We derive from such a partition a new bound that has a simple form (in the spirit of the cutset bound), and provide an intuitive interpretation of the bound as the directed information between inputs and outputs across these subsets of the partition. The standard cutset bound is indeed a special case when the partition has two subsets. Numerous existing bounds in the literature derived for special classes of networks can be seen to be implied by this new bound. These include the generalized network sharing (GNS) bound for graphical networks [14], the GNS bound for Gaussian networks [15], and the generalized cutset bound for deterministic networks [16]. Our bound may be viewed as an extension of the GNS bound to general noisy networks. At the same time, our bound is inspired from the generalized cutset bound [16] and can be rewritten in a form that is compatible with it. Our proof of the new bound relies crucially on a peeling-off lemma that brings about the single-letter characterization, a style of argument that was used, for example, in [17].

Throughout the paper, we mostly follow the notation in [18]. In particular, a random variable is denoted by an uppercase letter (e.g., X, Y, Z). An unspecified constant is denoted by 0. We use $X^n_i$ to denote a sequence $(X_i, \ldots, X_n)$ for $0 \leq i \leq n$ (otherwise, $X^n_i = \emptyset$). We always drop the subscript $i = 1$. The distinction between $X^n_i = (X_i, \ldots, X_n)$ and $X^n_n = (X_{i+1}, \ldots, X_{n-1})$ will be always clear from the context. For a pair of integers $i \leq j$, $[i : j] = \{i, \ldots, j\}$. We will let $\epsilon_n$ denote any generic sequence that satisfies $\epsilon_n \to 0$ as $n \to \infty$.

The rest of the paper is organized as follows. Sec. II reviews the notion of directed information. Sec. III formally defines the network communication problem and establishes the new bound on the capacity region.
Sec. IV applies this bound to Gaussian networks. Sec. V presents a few examples to illustrate the strength of the new bound over the cutset and other existing bounds.

II. REVIEW OF DIRECTED INFORMATION

Directed information was introduced by Massey [19] to study the capacity of channels with feedback. The directed information from a sequence $A^n$ to another (synchronized) sequence $B^n$ is defined as

$$ I(A_1, A_2, \ldots, A_n \to B_1, B_2, \ldots, B_n) := \sum_{i=1}^{n} I(B_i; A_i^i | B_1^{i-1}) $$

$$ = \sum_{i=1}^{n} I(B_i^n; A_i^i | B_1^{i-1}) $$

(1)

(2)

We will refer to (1) and (2) as Form 1 and Form 2 of the directed information. Similarly, the directed information from the sequence $A^n$ to the sequence $B^n$ causally conditioned on a third sequence $C^n$ is defined as

$$ I(A_1, \ldots, A_n \to B_1, B_2, \ldots, B_n \mid C_1, \ldots, C_n) := I((C_1, A_1), \ldots, (C_n, A_n) \to B_1, \ldots, B_n) $$

$$ - I(C_1, \ldots, C_n \to B_1, \ldots, B_n) $$

$$ = \sum_{i=1}^{n} I(B_i^i; A_i^i | B_1^{i-1}, C_i) $$

(3)

The following simple fact about causally conditional directed information, which is proved in Appendix A, will become quite useful in our subsequent discussion.

**Lemma 1.**

$$ I(A_1, \ldots, A_n \to B_1, B_2, \ldots, B_n \mid C_1, \ldots, C_n) \leq \sum_{i=1}^{n} I(B_i^i, C_{i+1}^i; A_i^i | B_1^{i-1}, C_i) $$

(4)

We will refer to (3) and (4) as Form 1 and Form 2 of the causally conditioned directed information. Note that Form 2 of the causally conditioned directed information is not an equivalent quantity but an upper bound. The above definitions are summarized in Table I.

Directed information can be interpreted as the amount of information one sequence causally provides about another [20]. Based on this interpretation, we can intuitively understand the conservation law for directed information [21], which states that the total mutual information between two sequences is the sum of the causal information transfer from one side to the other and back:

$$ I(A_1, A_2, \ldots, A_n; B_1, B_2, \ldots, B_n) $$

$$ = I(A_1, A_2, \ldots, A_n \to B_1, B_2, \ldots, B_n) $$

$$ + I(\emptyset, B_1, \ldots, B_{n-1} \to A_1, A_2, \ldots, A_n) $$

(5)

where $\emptyset$ refers to an unspecified constant.

III. PROBLEM STATEMENT AND THE MAIN RESULT

Consider a noisy network communication system with $N$ nodes indexed by $1, \ldots, N$. One wishes to provide $K$ information flows over the network, where each flow $f \in [1 : K]$ is embodied by reliable communication of message $M_f$ from source node $s_f \in [1 : N]$ to a set of destination nodes $D_f \subseteq [1 : N]$. This system can be modeled as a multivariate discrete memoryless network (DMN) $N = (X_1 \times \cdots \times X_n, p(y_1, \ldots, y_N|x_1, \ldots, x_N), y_1 \times \cdots \times y_n)$ that consists of $N$ sender–receiver alphabet pairs $(X_v, Y_v), v \in [1 : N]$, and a collection of probability mass functions (pmfs) $p(y_1, \ldots, y_N|x_1, \ldots, x_N)$. The network is assumed to be memoryless. A $(2^{nR_1}, \ldots, 2^{nR_K}, n)$ code for the DMN consists of:

- $K$ message sets $[1 : 2^{nR_1}], \ldots, [1 : 2^{nR_K}]$,
- encoders $x_v: \Pi_{f \in F_v}[1 : 2^{nR_f}] \times Y_v \times \cdots \times Y_{v_1} \to X_v$, $i \in [1 : n]$, for each node $v \in [1 : N]$, where $F_v = \{ f : v = s_f \}$, and
- decoders $\hat{m}_f: \Pi_{v \in F_f}[1 : 2^{nR_v}] \times Y_{v_f} \to [1 : 2^{nR_f}]$ for each flow $f \in [1 : K]$ and each of its destinations $d \in D_f$.

Assume that $(M_1, \ldots, M_K)$ is uniformly distributed over $[1 : 2^{nR_1}] \times \cdots \times [1 : 2^{nR_K}]$. The average probability of error is defined as

$$ P_e^{(n)} = P(M_f \neq \hat{M}_f \text{ for some } f \in [1 : K], d \in D_f) $$

A rate tuple $(R_1, \ldots, R_K)$ is said to be achievable if there exists a sequence of $(2^{nR_1}, \ldots, 2^{nR_K}, n)$ codes such that $\lim_{n \to \infty} P_e^{(n)} = 0$. The capacity region $C(N)$ of the DMN is the closure of the set of achievable rates.

We now introduce the notion of cut that will be crucial in the subsequent discussion.

**Definition 1.** For any ordered $(L + 1)$-partition $\mathcal{P} = \{V_0, V_1, \ldots, V_L\}$ of $[1 : N]$, namely, $\cup_{j=0}^{L} V_j = [1 : N]$ and $V_j \cap V_k = \emptyset$ for $j \neq k$, we say that flow $f$ is cut by the ordered partition $\mathcal{P}$ if $s_f \in V_j$ and $D_f \cap V_k = \emptyset$ for some $k < j$ (see Fig. 1). Let

$$ \text{Cut}(\mathcal{P}) := \{ f : \text{flow } f \text{ is cut by } \mathcal{P} \}. $$

Fig. 1. A network with 5 flows, where $D_f = \{ d_f, d'_f \}$ for $f = 1, 2, 4$ and $D_f = \{ d_f \}$ for $f = 3, 5$. Flows 1, 2, 3, 5 are cut by the partition $(V_0, V_1, V_2, V_3, V_4)$ but flow 4 is not. Furthermore, fusing $V_1$ and $V_2$ will also cut the very same flows, and hence will lead to a tighter bound.
A few remarks are in order.

Let $L > 1$. Consider an ordered partition $\mathcal{P} = (\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_L)$ and another ordered partition $\mathcal{P}' = (\mathcal{V}_0', \mathcal{V}_1', \ldots, \mathcal{V}_{L-1}')$ defined by fusing two neighboring sets of $\mathcal{P}$ as

$$\mathcal{V}_j' = \begin{cases} \mathcal{V}_j & 0 \leq j < k - 1, \\ \mathcal{V}_{k-1} \cup \mathcal{V}_k & j = k - 1, \\ \mathcal{V}_{j+1} & k \leq j \leq L - 1 \end{cases}$$

for some $k \in [1:L]$. Then,

$$I(\emptyset, \mathcal{Y}_0, \mathcal{Y}_1, \ldots, \hat{\mathcal{Y}}_{L-1} = \mathcal{X}_0, \mathcal{X}_2, \ldots, \mathcal{X}_L)$$

for any $L \geq 1$ and any ordered $(L + 1)$-partition $\mathcal{P} = (\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_L)$. Here, $\hat{\mathcal{Y}}_j := (Y_v : v \in \mathcal{V}_j)$ and $\mathcal{X}_j := (X_v : v \in \mathcal{V}_j)$.

A few remarks are in order.

Remark 1. Let $L > 1$. Consider an ordered partition $\mathcal{P} = (\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_L)$ and another ordered partition $\mathcal{P}' = (\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_{L-1})$ defined by fusing two neighboring sets of $\mathcal{P}$ as

$$\mathcal{V}_j' = \begin{cases} \mathcal{V}_j & 0 \leq j < k - 1, \\ \mathcal{V}_{k-1} \cup \mathcal{V}_k & j = k - 1, \\ \mathcal{V}_{j+1} & k \leq j \leq L - 1 \end{cases}$$

for some $k \in [1:L]$. Then,

$$I(\emptyset, \mathcal{Y}_0, \mathcal{Y}_1, \ldots, \hat{\mathcal{Y}}_{L-1} = \mathcal{X}_0, \mathcal{X}_2, \ldots, \mathcal{X}_L)$$

for any $L \geq 1$ and any ordered $(L + 1)$-partition $\mathcal{P} = (\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_L)$. Here, $\hat{\mathcal{Y}}_j := (Y_v : v \in \mathcal{V}_j)$ and $\mathcal{X}_j := (X_v : v \in \mathcal{V}_j)$.

A few remarks are in order.
Remark 2. For deterministic networks, the inequalities in the new bound have the same form as those for the bound by Shomorony and Avestimehr [16]. Our theorem extends Theorem 1 therein, whose proof technique is limited to deterministic networks, to general noisy networks (their notation \( \Omega_j \) is equivalent to \( \cup_{k=1}^{j} \mathcal{V}_k \) in our notation). However, our notion of cut is more general than theirs, and hence our bound is a strict improvement over [16] even within the class of deterministic networks; see Sec. V-A.

Before we provide a proof of Theorem 1, we briefly discuss various alternate forms of the bound.

For \( L = 1 \), we have an ordered bipartition \( \mathcal{P} = (\mathcal{V}_0, \mathcal{V}_1) \), and flow \( f \) is cut by this partition (i.e., \( f \in \text{cut} (\mathcal{P}) \)) iff \( s_f \in \mathcal{V}_1 \) and \( D_f \cap \mathcal{V}_0 \neq \emptyset \). Thus, the corresponding inequality in Theorem 1 simplifies as

\[
\sum_{f \in \text{Cut}(\mathcal{P})} R_f \leq I(0, \tilde{Y}_0 \to \tilde{X}_0, \tilde{X}_1) = I(\tilde{X}_1; \tilde{Y}_0 | \tilde{X}_0),
\]

which is the standard cutset bound [1], [18, Sec. 18.4].

For \( L = 2 \), we have an ordered tripartition \( \mathcal{P} = (\mathcal{V}_0, \mathcal{V}_1, \mathcal{V}_2) \) and the corresponding inequality in Theorem 1 simplifies as

\[
\sum_{f \in \text{Cut}(\mathcal{P})} R_f \leq I(0, \tilde{Y}_0, \tilde{Y}_1 \to \tilde{X}_0, \tilde{X}_1, \tilde{X}_2)
\]

where (7), (8), and (9) are Form 1 of the directed information (1), Form 2 of the directed information (2), and the conservation law (5), respectively. The identity (7) splits the directed information with one input subset at a time, so that none of the mutual information terms is conditioned on output variables. The identity (8) splits the directed information with one output subset at a time, including all input variables on either the left side or under the conditioning.

For \( L > 2 \), we can similarly write these forms of the inequality as in (6)-(9).

Proof of Theorem 1: For each \( j \in [1 : L] \), let \( M_j = (M_j : s_j \in \mathcal{V}_j \cap \text{Cut}(\mathcal{P})) \) and let \( M_0 \) denote all the messages that are not cut by \( \mathcal{P} \), i.e., \( M_0 = (M_j : f \notin \text{Cut}(\mathcal{P})) \). Note that \( M_0 \) may contain messages from source nodes in any subset \( \mathcal{V}_j \), and hence \( \tilde{X}_{j,i} \) is a function of \( (M_0, M_j, Y_j^{i-1}) \) for \( j \in [0 : L] \). Let \( M = (M_1, \ldots, M_K) = (M_0, \ldots, M_L) \).

For \( L = 1 \), this is simply the cutset bound. We present the proof here for \( L = 2 \) and relegate the proof for general \( L \) to Appendix B.

By Fano’s inequality, we have

\[
\sum_{f \in \text{Cut}(\mathcal{P}), s_f \in \mathcal{V}_1} n(R_f - \epsilon_n) \leq I(\tilde{Y}_0^n; \tilde{M}_1 | \tilde{M}_0),
\]

where \( \epsilon_n \to 0 \) as \( n \to \infty \). Summing these inequalities, we have

\[
\sum_{f \in \text{Cut}(\mathcal{P})} n(R_f - \epsilon_n) \leq I(\tilde{Y}_0^n, \tilde{Y}_1^n; \tilde{M}_1, \tilde{M}_0),
\]

where \( (a) \) follows from Form 1 of the directed information and \( (b) \) follows by the peeling-off lemma (Lemma 2) stated below. Using a standard time-sharing random variable completes the proof.

Lemma 2 (Peeling-off lemma). For \( i \in [1 : n] \),

\[
I(\tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

\[
-I(\tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1} \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

\[
= I(\tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2) - I(\tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1} \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

\[
\leq I(\tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

\[
\leq I(\tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

\[
\leq I(\tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

\[
\leq I(\tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

\[
\leq I(\tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2)
\]

where \( (a) \) follows from Lemma 1, \( (b) \) follows since \( \tilde{X}_0^i \) is a function of \( (\tilde{M}_0, \tilde{Y}_0^{i-1}) \) and \( \tilde{X}_1^i \) is a function of \( (\tilde{M}_1, \tilde{Y}_1^{i-1}) \), and \( (c) \) follows from Markov conditions guaranteed by the memoryless property of the network.

Proof of Lemma 2: Consider
IV. GAUSSIAN NETWORKS

Consider a Gaussian network [18, Ch. 19] with channel gain matrix \( G \), power constraint \( P \) at each node, and i.i.d. \( N(0,1) \) noise components. The output of the network is

\[ Y^N = GX^N + Z^N. \]

Theorem 1 can be easily adapted to this setting.

**Corollary 1.** If \( (R_1, \ldots, R_K) \in \mathcal{E}(N) \), then there must exist a jointly Gaussian \( (X_1, \ldots, X_N) \) with \( E(X_i^2) \leq P \), \( i \in \{1 : N\} \), such that

\[
\sum_{f \in \text{Cut}(P)} R_f \leq I(\emptyset, \tilde{Y}_0, \ldots, \tilde{Y}_{L-1} \rightarrow \tilde{X}_0, \ldots, \tilde{X}_L) \tag{10}
= \frac{1}{2} \log_2 \left( \prod_{j=0}^{L-1} \text{Cov}(\tilde{Y}_j \mid \tilde{X}_j, \tilde{Y}_{j-1}) \right)
\]

for any \( L > 1 \) and any ordered \( (L + 1) \)-partition \( \mathcal{P} = (V_0, \ldots, V_L) \). Here \( \text{Cov}(\tilde{Y}_j \mid \tilde{X}_j, \tilde{Y}_{j-1}) \) denotes the conditional covariance matrix of the jointly Gaussian \( \tilde{Y}_j = (Y_v \mid v \in V_j) \) given \( (\tilde{X}_0, \tilde{Y}_{j-1}) \) and \( \cdot \mid \cdot \) denotes its determinant.

**Proof:** It suffices to show that Gaussian input distributions maximize the RHS of (10). This can be seen by using Form 2 of the directed information and the maximal differential entropy lemma [18, Ch. 2.2, Eq. (2.7)].

**Remark 3.** Corollary 1 implies the generalized network sharing (GNS) bound for Gaussian networks [15]. This implication can be proved by the network concatenation idea in [16], which is used to show that the GNS bound for graphical networks follows from the generalized cutset bound for deterministic networks. Furthermore, Corollary 1 strictly improves the GNS bound for Gaussian networks; see Sec. V-B.

V. EXAMPLES

A. Binary Symmetric Network

Consider the example network in Fig. 2. The network model is:

\[ Y_2 = X_1 \oplus Z_2, \]
\[ Y_3 = X_1 \oplus X_2 \oplus Z_3, \]

where \( Z_2 \) and \( Z_3 \) are independent \( \text{Bern}(\epsilon) \) noise components. The source-destination pairs are given by \( s_1 = 1, d_1 = 2, s_2 = 2, d_2 = 3, s_3 = 1, d_3 = 3 \). The cutset bound for this network is given by

\[ R_1 \leq 1 - H(\epsilon), \]
\[ R_2 + R_3 \leq 1 - H(\epsilon), \]
\[ R_1 + R_3 \leq 1 - 2H(\epsilon) + H(2\epsilon(1 - \epsilon)). \]

In comparison, Theorem 1 yields a tighter bound on the sum-rate:

\[ R_1 \leq 1 - H(\epsilon), \]
\[ R_2 + R_3 \leq 1 - H(\epsilon), \]
\[ R_1 + R_2 + R_3 \leq 1 - 2H(\epsilon) + H(2\epsilon(1 - \epsilon)). \]

When \( \epsilon = 0 \), this is a deterministic network and the generalized cutset bound in [16] leads to the following outer bound on the capacity region:

\[ R_1 + R_2 \leq 1, \quad R_1 + R_3 \leq 1, \quad R_2 + R_3 \leq 1. \]

It can be readily checked that our bound improves this bound with a new stronger inequality

\[ R_1 + R_2 + R_3 \leq 1. \]

Fig. 2. A binary symmetric network.

B. Gaussian Examples

1) A three-node network: Consider the example network in Fig. 3 with power constraint \( P \) on each node. The network model is

\[ Y_2 = X_1 + Z_2, \]
\[ Y_3 = X_1 + X_2 + Z_3, \]

where \( Z_2 \) and \( Z_3 \) are independent \( N(0,1) \) noise components. The source-destination pairs are given by \( s_1 = 1, d_1 = 2, s_2 = 2, d_2 = 3, s_3 = 1, d_3 = 3 \). The cutset bound as well as the GNS bound [15] for this network is

\[ R_f \leq \frac{1}{2} \log_2 (1 + P), \quad f = 1, 2, \]

In comparison, our new bound in Corollary 1 with the ordered partition \( \mathcal{P} = (\{3\}, \{2\}, \{1\}) \) yields

\[ R_1 + R_2 \leq \frac{1}{2} \log_2 (\text{Var}(Y_3) \text{Var}(Y_2|X_2, Y_3)) \]
\[ \leq \frac{1}{2} \log_2 ((1 + E(X_1 + X_2)^2) (1 + 1)) \]
\[ \leq \frac{1}{2} \log_2 (1 + 4P) + \frac{1}{2}. \]

Thus the upper bound of 2 on the sum degrees of freedom (DoF) provided by the cutset and GNS bounds is loose, and the actual sum DoF is 1, which is easily achievable by time division.
2) An N-node network: Consider an extension of the above network to a network with N nodes and N − 1 source–destination pairs \( s_f = f \) and \( d_f = f + 1, f \in [1 : N − 1] \). The network model is

\[
Y_v = \sum_{u=1}^{v-1} X_u + Z_v, \quad v \in [2 : N],
\]

where \( Z_2, \ldots, Z_N \) are independent N(0, 1) noise components. The cutset bound for this network (over all possible bipartition cuts) is

\[
R_f \leq \frac{1}{2} \log_2 (1 + P), \quad f \in [1 : N − 1].
\]

Thus, the cutset bound on the sum-capacity is

\[
C_{\text{sum}} := \max_{f \in [1 : N − 1]} \frac{N - 1}{2} \log_2 (1 + P).
\]

In comparison, our bound with the natural ordered partition \( \mathcal{P} = \{(N), \{N − 1\}, \ldots, \{1\}\} \) implies that

\[
C_{\text{sum}} \leq \frac{1}{2} \log_2 \left( \text{Var}(Y_N) \cdots \text{Var}(Y_2 | X_1, Y_3^N) \right).
\]

One can verify that \( \text{Var}(Y_N) \leq 1 + (N − 1)^2P \) and for \( j = 1, 2, \ldots, N − 2 \),

\[
\text{Var}(Y_{N-j} | X_{N-j}^N, Y_{N-j+1}^N) = 1 + \frac{\text{Var}(\sum_{v=1}^{N-j-1} X_v | X_{N-j}^N)}{j \text{Var}(\sum_{v=1}^{N-j-1} X_v | X_{N-j}^N)} + 1 \leq 1 + \frac{1}{j}.
\]

Hence, our new bound yields

\[
C_{\text{sum}} \leq \frac{1}{2} \log_2 \left( 1 + (N − 1)^2P \right) + \frac{1}{2} \sum_{j=1}^{N-1} \log_2 \left( 1 + \frac{1}{j} \right)
\]

\[
\leq \frac{1}{2} \log_2 \left( 1 + (N − 1)^2P \right) + \frac{\log_2 \left( \frac{1}{2} \sum_{j=1}^{N-1} \frac{1}{j} \right)}{2}.
\]

In summary, the cutset bound implies \( C_{\text{sum}} = O(N) \) while the new bound implies \( C_{\text{sum}} = O(\log N) \). Furthermore, a simple time-division scheme (with power control) yields a lower bound on the sum-capacity as

\[
C_{\text{sum}} \geq \frac{1}{2} \log_2 (1 + (N − 1)P).
\]

Therefore, the new bound captures the correct scaling behavior of sum-capacity in this Gaussian network.

ACKNOWLEDGEMENTS

SK would like to acknowledge support from the Information Theory and Applications Center at UCSD, and from the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant CCF-0939370. YHK would like to acknowledge support from the NSF under Grant CCF-1320895.

APPENDIX A

PROOF OF LEMMA 1

Consider

\[
I(A_1, \ldots, A_n \rightarrow B_1, \ldots, B_n \| C_1, \ldots, C_n)
\]

\[
= \sum_{j=1}^{n} I(B_j; A^j | B^{j-1} C^j)
\]

\[
= \sum_{j=1}^{n} \sum_{i=1}^{j} I(B_j; A_i | A^{i-1}, B^{j-1}, C^j)
\]

\[
\leq \sum_{i=1}^{n} \sum_{j=i}^{n} I(B_j; C_{j+1}; A_i | A^{i-1}, B^{j-1}, C^j)
\]

\[
= \sum_{i=1}^{n} I(B^n_i; C_{i+1}; A_i | A^{i-1}, B^{i-1}, C^j).
\]

APPENDIX B

PROOF OF THEOREM 1 FOR GENERAL L

By Fano’s inequality, we have for \( j \in [1 : L] \),

\[
\sum_{f \in \text{Cut}(\mathcal{P}), s_f \in V_j} n(R_f - \epsilon_n)
\]

\[
\leq I(M_j; \hat{Y}_0^n, \hat{Y}_1^n, \ldots, \hat{Y}_{L-1}^n | \tilde{M}_0, \tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_L).
\]

By summing these inequalities for \( j \in [1 : L] \),

\[
\sum_{f \in \text{Cut}(\mathcal{P})} n(R_f - \epsilon_n)
\]

\[
\leq I(\emptyset; \hat{Y}_0^n, \hat{Y}_1^n, \ldots, \hat{Y}_{L-1}^n | \tilde{M}_0, \tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_L).
\]

Finally, the following lemma will complete the proof.

Lemma 3 (Peeling-off lemma). For \( i \in [1 : n] \), the inequality in (11) on top of the next page holds.

Proof of Lemma 3: Consider the series of inequalities (12) on top of the next page, where where (a) follows by Lemma 1, (b) follows since \( X_k \) is a function of \( (M_0, M_k, \hat{Y}_k^{i-1}) \), and (c) follows from Markov conditions guaranteed by the memoryless property of the network. ■
\[
I(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_{L-1} \rightarrow \bar{M}_0, \bar{M}_1, \bar{M}_2, \ldots, \bar{M}_L) \\
\leq I(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_{L-1} \rightarrow \bar{M}_0, \bar{M}_1, \bar{M}_2, \ldots, \bar{M}_L) \\
+ I(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_{L-1} \rightarrow \bar{X}_0, \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_L). \\
\]

(11)

\[
I(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_{L-1} \rightarrow \bar{M}_0, \bar{M}_1, \ldots, \bar{M}_L) - I(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_{L-1} \rightarrow \bar{M}_0, \bar{M}_1, \ldots, \bar{M}_L) \\
= I(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_{L-1} \rightarrow \bar{M}_0, \bar{M}_1, \ldots, \bar{M}_L) \\
\leq \sum_{j=1}^{L} I(\bar{(M)_k^{j-1}} \mid \bar{Y}_j^{i-1}, \bar{Y}_{j-1}, \bar{X}_k^{i-1}) \\
\leq \sum_{j=1}^{L} I(\bar{M}, \bar{Y}_j^{i-1}, \bar{X}_k^{i-1}) \\
= I(\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_{L-1, i} \rightarrow \bar{X}_0, \bar{X}_1, \ldots, \bar{X}_{L,i}). \\
\]

(12)

REFERENCES