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 multiuser 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 $\emptyset$. We use X_i^n to denote a sequence $(X_i, \ldots, X_n)$ for $0 \leq i \leq n$ (otherwise, $X_i^n = \emptyset$). We always drop the subscript $i = 1$. The distinction between $X_i^n = (X_i, \ldots, X_n)$ and $X_v^n = (X_{v1}, \ldots, X_{vn})$ will be always clear from the context. For a pair of integers $i \leq j$, $[i : j] = \{i, i+1, \ldots, j\}$. We will let ϵ_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

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

$$:= \sum_{i=1}^{n} I(B_i; A^i | B^{i-1}) \qquad (1)$$

$$= \sum_{i=1}^{n} I(B_i^n; A_i | A^{i-1}, B^{i-1}). \qquad (2)$$

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

$$I(A_1, \ldots, A_n \to B_1, \ldots, B_n \,\|\, C_1, \ldots, C_n)$$

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

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

$$= \sum_{i=1}^{n} I(B_i; A^i | B^{i-1}, C^i) \qquad (3)$$

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.

$$I(A_1, \ldots, A_n \to B_1, \ldots, B_n \,\|\, C_1, \ldots, C_n)$$

$$\leq \sum_{i=1}^{n} I(B_i^n, C_{i+1}^n; A_i | A^{i-1}, B^{i-1}, C^i). \qquad (4)$$

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:

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

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

$$\quad + I(\emptyset, B_1, \ldots, B_{n-1} \to A_1, A_2, \ldots, A_n), \qquad (5)$$

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 $\mathcal{D}_f \subseteq [1 : N]$. This system can be modeled as a multimessage discrete memoryless network (DMN) $\mathcal{N} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_n, p(y_1, \ldots, y_N | x_1, \ldots, x_N), \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n)$ that consists of N sender–receiver alphabet pairs $(\mathcal{X}_v, \mathcal{Y}_v), v \in [1 : N]$, 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 $[1 : 2^{\lceil nR_1 \rceil}], \ldots, [1 : 2^{\lceil nR_K \rceil}]$,
- encoders $x_{vi} : \Pi_{f \in \mathcal{F}_v} [1 : 2^{\lceil nR_f \rceil}] \times \mathcal{Y}_v^{i-1} \to \mathcal{X}_{vi}$, $i \in [1 : n]$, for each node $v \in [1 : N]$, where $\mathcal{F}_v := \{f : v = s_f\}$, and
- decoders $\hat{m}_{fd} : \Pi_{g \in \mathcal{F}_d} [1 : 2^{\lceil nR_g \rceil}] \times \mathcal{Y}_d^n \to [1 : 2^{\lceil nR_f \rceil}]$ for each flow $f \in [1 : K]$ and each of its destinations $d \in \mathcal{D}_f$.

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

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

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 $\mathscr{C}(\mathcal{N})$ 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 $\mathcal{P} = (\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_L)$ of $[1 : N]$, namely, $\cup_{j=0}^{L} \mathcal{V}_j = [1 : N]$ and $\mathcal{V}_j \cap \mathcal{V}_k = \emptyset$ for $j \neq k$, we say that flow f is *cut* by the ordered partition $\mathcal{P}$ if $s_f \in \mathcal{V}_j$ and $\mathcal{D}_f \cap \mathcal{V}_k \neq \emptyset$ for some $k < j$ (see Fig. 1). Let

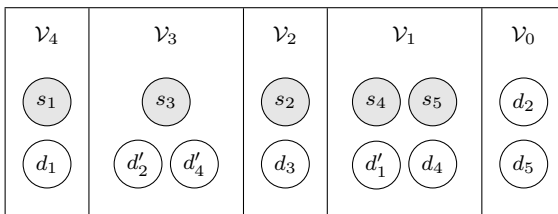$$\mathrm{Cut}(\mathcal{P}) := \{f : \text{flow } f \text{ is cut by } \mathcal{P}\}.$$

Fig. 1. A network with 5 flows, where $\mathcal{D}_f = \{d_f, d'_f\}$ for $f = 1, 2, 4$ and $\mathcal{D}_f = \{d_f\}$ for $f = 3, 5$. Flows $1, 2, 3, 5$ are cut by the partition $(\mathcal{V}_0, \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4)$ but flow 4 is not. Furthermore, fusing $\mathcal{V}_1$ and $\mathcal{V}_2$ will also cut the very same flows, and hence will lead to a tighter bound.

$I\left(A_1, A_2, \ldots, A_n \to B_1, B_2, \ldots, B_n\right)$	$= \sum_{i=1}^{n} I\left(B_i; A^i \mid B^{i-1}\right)$ (*Form 1*)	$\boxed{A_1}\ \boxed{A_2}\ \boxed{A_3}\ \boxed{A_4}\ \boxed{A_5}\ \boxed{A_6}$ $\boxed{B_1}\ \boxed{B_2}\ \boxed{B_3}\ \boxed{B_4}\ \boxed{B_5}\ \boxed{B_6}$
$I\left(A_1, A_2, \ldots, A_n \to B_1, B_2, \ldots, B_n\right)$	$= \sum_{i=1}^{n} I\left(B_i^n; A_i \mid A^{i-1}, B^{i-1}\right)$ (*Form 2*)	$\boxed{A_1}\ \boxed{A_2}\ \boxed{A_3}\ \boxed{A_4}\ \boxed{A_5}\ \boxed{A_6}$ $\boxed{B_1}\ \boxed{B_2}\ \boxed{B_3}\ \boxed{B_4}\ \boxed{B_5}\ \boxed{B_6}$
$I\left(A_1, \ldots, A_n \to B_1, \ldots, B_n \| C_1, \ldots, C_n\right)$	$= \sum_{i=1}^{n} I\left(B_i; A^i \mid B^{i-1}, C^i\right)$ (*Form 1*)	$\boxed{A_1}\ \boxed{A_2}\ \boxed{A_3}\ \boxed{A_4}\ \boxed{A_5}\ \boxed{A_6}$ $\boxed{B_1}\ \boxed{B_2}\ \boxed{B_3}\ \boxed{B_4}\ \boxed{B_5}\ \boxed{B_6}$ $\boxed{C_1}\ \boxed{C_2}\ \boxed{C_3}\ \boxed{C_4}\ \boxed{C_5}\ \boxed{C_6}$
$I\left(A_1, \ldots, A_n \to B_1, \ldots, B_n \| C_1, \ldots, C_n\right)$	$\leq \sum_{i=1}^{n} I\left(B_i^n, C_{i+1}^n; A_i \mid A^{i-1}, B^{i-1}, C^i\right)$ (*Form 2*)	$\boxed{A_1}\ \boxed{A_2}\ \boxed{A_3}\ \boxed{A_4}\ \boxed{A_5}\ \boxed{A_6}$ $\boxed{B_1}\ \boxed{B_2}\ \boxed{B_3}\ \boxed{B_4}\ \boxed{B_5}\ \boxed{B_6}$ $\boxed{C_1}\ \boxed{C_2}\ \boxed{C_3}\ \boxed{C_4}\ \boxed{C_5}\ \boxed{C_6}$

TABLE I

TWO FORMS FOR DIRECTED INFORMATION AND CAUSALLY CONDITIONAL DIRECTED INFORMATION.

If $L = 1$, we have an ordered bipartition $(\mathcal{V}_0, \mathcal{V}_1)$, and our notion of flow being cut by the partition is identical to the standard notion of cut $(\mathcal{V}, \mathcal{V}^c) = (\mathcal{V}_1, \mathcal{V}_0)$ considered in the cutset bound.

We are ready to state our main result.

Theorem 1. *If the rate tuple* $(R_1, \ldots, R_K) \in \mathscr{C}(\mathcal{N})$, *then there must exist a pmf* $p(x_1, \ldots, x_N)$ *such that*

$$\sum_{f \in \mathrm{Cut}(\mathcal{P})} R_f$$
$$\leq I(\emptyset, \tilde{Y}_0, \tilde{Y}_1, \ldots, \tilde{Y}_{L-1} \to \tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_L)$$

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

A few remarks are in order.

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

$$I(\emptyset, \tilde{Y}_0, \tilde{Y}_1, \ldots, \tilde{Y}_{L-1} \to \tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_L)$$
$$- I(\emptyset, \tilde{Y}'_0, \tilde{Y}'_1, \ldots, \tilde{Y}'_{L-2} \to \tilde{X}'_0, \tilde{X}'_1, \ldots, \tilde{X}'_{L-1})$$
$$= I(\tilde{X}_k; \tilde{Y}_{k-1} \mid \tilde{X}_0^{k-1}, \tilde{Y}_0^{k-2}) \geq 0.$$

Thus, if $\mathrm{cut}(\mathcal{P}) = \mathrm{cut}(\mathcal{P}')$, then $\mathcal{P}'$ yields a tighter bound than $\mathcal{P}$ does (see Fig. 1). This implies that if there are K information flows in the network, it suffices to consider ordered partitions of at most $K+1$ subsets (i.e., $L \leq K$). In particular, if all the information sources are located on the same node or all the destinations are located on the same node (either of which happens if there is only one information flow), then it suffices to consider only bipartitions ($L = 1$) and the new bound reduces to the cutset bound without any improvement.

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 Ω_j is equivalent to $\cup_{k=j}^{L} \mathcal{V}_k$ 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 $\mathcal{D}_f \cap \mathcal{V}_0 \neq \emptyset$. Thus, the corresponding inequality in Theorem 1 simplifies as

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

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

$$
\begin{aligned}
\sum_{f \in \text{Cut}(\mathcal{P})} & R_f \\
\leq \ & I(\emptyset, \tilde{Y}_0, \tilde{Y}_1 \to \tilde{X}_0, \tilde{X}_1, \tilde{X}_2) \\
= \ & I(\tilde{Y}_0, \tilde{Y}_1 \to \tilde{X}_1, \tilde{X}_2 | \tilde{X}_0) \quad (6) \\
= \ & I(\tilde{X}_1; \tilde{Y}_0 | \tilde{X}_0) + I(\tilde{X}_2; \tilde{Y}_0, \tilde{Y}_1 | \tilde{X}_0, \tilde{X}_1) \quad (7) \\
= \ & I(\tilde{X}_1, \tilde{X}_2; \tilde{Y}_0 | \tilde{X}_0) + I(\tilde{X}_2; \tilde{Y}_1 | \tilde{X}_0, \tilde{X}_1, \tilde{Y}_0) \quad (8) \\
= \ & I(\tilde{X}_0^2; \tilde{Y}_0^2) - I(\tilde{X}_0, \tilde{X}_1, \tilde{X}_2 \to \tilde{Y}_0, \tilde{Y}_1, \tilde{Y}_2), \quad (9)
\end{aligned}
$$

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 $\tilde{M}_j = (M_f : s_f \in \mathcal{V}_j \cap \text{Cut}(\mathcal{P}))$ and let $\tilde{M}_0$ denote all the messages that are not cut by $\mathcal{P}$, i.e., $\tilde{M}_0 = (M_f : f \notin \text{Cut}(\mathcal{P}))$. Note that $\tilde{M}_0$ may contain messages from source nodes in any subset $\mathcal{V}_j$, and hence $\tilde{X}_{j,i}$ is a function of $(\tilde{M}_0, \tilde{M}_j, \tilde{Y}_j^{i-1})$ for $j \in [0 : L]$. Let $M = (M_1, \ldots, M_K) = (\tilde{M}_0, \ldots, \tilde{M}_L)$.

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

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

$$\sum_{f \in \text{Cut}(\mathcal{P}), s_f \in \mathcal{V}_2} n(R_f - \epsilon_n) \leq I(\tilde{M}_2; \tilde{Y}_0^n, \tilde{Y}_1^n | \tilde{M}_0, \tilde{M}_1),$$

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

$$
\begin{aligned}
\sum_{f \in \text{Cut}(\mathcal{P})} & n(R_f - \epsilon_n) \\
& \overset{(a)}{\leq} I(\emptyset, \tilde{Y}_0^n, \tilde{Y}_1^n \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2) \\
& \overset{(b)}{\leq} \sum_{i=1}^{n} I(\emptyset, \tilde{Y}_{0i}, \tilde{Y}_{1i} \to \tilde{X}_{0i}, \tilde{X}_{1i}, \tilde{X}_{2i}),
\end{aligned}
$$

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 $i \in [1 : n]$,*

$$
\begin{aligned}
I(\emptyset, & \tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2) \\
& \leq I(\emptyset, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1} \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2) \\
& \quad + I(\emptyset, \tilde{Y}_{0i}, \tilde{Y}_{1i} \to \tilde{X}_{0i}, \tilde{X}_{1i}, \tilde{X}_{2i}).
\end{aligned}
$$

Proof of Lemma 2: Consider

$$
\begin{aligned}
& I(\emptyset, \tilde{Y}_0^i, \tilde{Y}_1^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2) \\
& \quad - I(\emptyset, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1} \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2) \\
& = I(\emptyset, \tilde{Y}_{0i}, \tilde{Y}_{1i} \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2 \,\|\, \emptyset, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1}) \\
& \overset{(a)}{\leq} I(\tilde{M}_1, \tilde{M}_2, \tilde{Y}_1^{i-1}; \tilde{Y}_{0i} | \tilde{M}_0, \tilde{Y}_0^{i-1}) \\
& \quad + I(\tilde{M}_2; \tilde{Y}_{1i} | \tilde{Y}_{0i}, \tilde{M}_0, \tilde{M}_1, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1}) \\
& \overset{(b)}{=} I(\tilde{M}_1, \tilde{M}_2, \tilde{Y}_1^{i-1}; \tilde{Y}_{0i} | \tilde{M}_0, \tilde{Y}_0^{i-1}, \tilde{X}_{0i}) \\
& \quad + I(\tilde{M}_2; \tilde{Y}_{1i} | \tilde{Y}_{0i}, \tilde{M}_0, \tilde{M}_1, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1}, \tilde{X}_{0i}, \tilde{X}_{1i}) \\
& \leq I(M, \tilde{Y}_1^{i-1}, \tilde{Y}_0^{i-1}; \tilde{Y}_{0i} | \tilde{X}_{0i}) \\
& \quad + I(M, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1}; \tilde{Y}_{1i} | \tilde{Y}_{0i}, \tilde{X}_{0i}, \tilde{X}_{1i}) \\
& \leq I(M, \tilde{Y}_1^{i-1}, \tilde{Y}_0^{i-1}, \tilde{X}_{1i}, \tilde{X}_{2i}; \tilde{Y}_{0i} | \tilde{X}_{0i}) \\
& \quad + I(M, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1}, \tilde{X}_{2i}; \tilde{Y}_{1i} | \tilde{Y}_{0i}, \tilde{X}_{0i}, \tilde{X}_{1i}) \\
& \overset{(c)}{=} I(\tilde{X}_{1i}, \tilde{X}_{2i}; \tilde{Y}_{0i} | \tilde{X}_{0i}) + I(\tilde{X}_{2i}; \tilde{Y}_{1i} | \tilde{Y}_{0i}, \tilde{X}_{0i}, \tilde{X}_{1i}) \\
& = I(\emptyset, \tilde{Y}_{0i}, \tilde{Y}_{1i} \to \tilde{X}_{0i}, \tilde{X}_{1i}, \tilde{X}_{2i}),
\end{aligned}
$$

where (a) follows by Lemma 1, (b) follows since $\tilde{X}_{0i}$ is a function of $(\tilde{M}_0, \tilde{Y}_0^{i-1})$ and $\tilde{X}_{1i}$ is a function of $(\tilde{M}_0, \tilde{M}_1, \tilde{Y}_1^{i-1})$, and (c) follows from Markov conditions guaranteed by the memoryless property of the network. ∎

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* $(R_1, \ldots, R_K) \in \mathscr{C}(\mathcal{N})$, *then there must exist a jointly Gaussian* $(X_1, \ldots, X_N)$ *with* $\mathsf{E}(X_v^2) \leq P$, $v \in [1:N]$, *such that*

$$\sum_{f \in \mathrm{Cut}(\mathcal{P})} R_f$$
$$\leq I(\emptyset, \tilde{Y}_0, \ldots, \tilde{Y}_{L-1} \to \tilde{X}_0, \ldots, \tilde{X}_L) \quad (10)$$
$$= \frac{1}{2} \log_2 \left(\prod_{j=0}^{L-1} |\mathrm{Cov}(\tilde{Y}_j | \tilde{X}_0^j, \tilde{Y}_0^{j-1})| \right)$$

for any $L > 1$ *and any ordered* $(L+1)$-*partition* $\mathcal{P} = (\mathcal{V}_0, \ldots, \mathcal{V}_L)$. *Here* $\mathrm{Cov}(\tilde{Y}_j | \tilde{X}_0^j, \tilde{Y}_0^{j-1})$ *denotes the conditional covariance matrix of the jointly Gaussian* $\tilde{Y}_j = (Y_v : v \in \mathcal{V}_j)$ *given* $(\tilde{X}_0^j, \tilde{Y}_0^{j-1})$ *and* $|\cdot|$ *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 $\mathrm{Bern}(\epsilon)$ 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
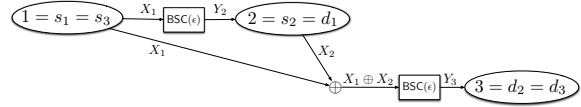
$$R_1 + R_2 + R_3 \leq 1.$$

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 $s_1 = 1, d_1 = 2, s_2 = 2, d_2 = 3$. 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

$$R_1 + R_2 \leq \frac{1}{2} \log_2 \left(\mathrm{Var}(Y_3) \, \mathrm{Var}(Y_2 | X_2, Y_3) \right)$$
$$\leq \frac{1}{2} \log_2 \left((1 + \mathsf{E}(X_1 + X_2)^2)(1+1) \right)$$
$$\leq \frac{1}{2} \log_2(1 + 4P) + \frac{1}{2}.$$

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.
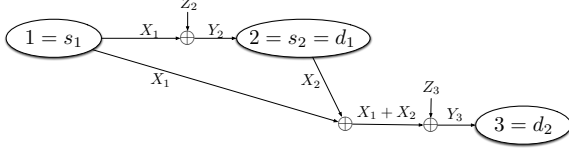
Fig. 3. A 3-node Gaussian network.

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

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

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

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

Thus, the cutset bound on the sum-capacity is

$$C_{\text{sum}} := \max \sum_{f=1}^{N-1} R_f \le \frac{N-1}{2} \log_2(1 + P).$$

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

$$C_{\text{sum}} \le \frac{1}{2} \log_2\left(\text{Var}(Y_N) \cdots \text{Var}(Y_2 | X_2^{N-1}, Y_3^N)\right).$$

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

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

Hence, our new bound yields

$$C_{\text{sum}} \le \frac{1}{2} \log_2\left(1 + (N-1)^2 P\right) + \frac{1}{2} \sum_{j=1}^{N-1} \log_2\left(1 + \frac{1}{j}\right)$$
$$\le \frac{1}{2} \log_2\left(1 + (N-1)^2 P\right) + \frac{\log_2 e}{2} \sum_{j=1}^{N-1} \frac{1}{j}.$$

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

$$C_{\text{sum}} \ge \frac{1}{2} \log_2(1 + (N-1)P).$$

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

APPENDIX A
PROOF OF LEMMA 1

Consider

$$I(A_1, \ldots, A_n \to B_1, \ldots, B_n \| C_1, \ldots, C_n)$$
$$= \sum_{j=1}^n I(B_j; A^j | B^{j-1} C^j)$$
$$= \sum_{j=1}^n \sum_{i=1}^j I(B_j; A_i | A^{i-1}, B^{j-1}, C^j)$$
$$= \sum_{i=1}^n \sum_{j=i}^n I(B_j; A_i | A^{i-1}, B^{j-1}, C^j)$$
$$\le \sum_{i=1}^n \sum_{j=i}^n I(B_j, C_{j+1}; A_i | A^{i-1}, B^{j-1}, C^j)$$
$$= \sum_{i=1}^n I(B_i^n, C_{i+1}^n; A_i | A^{i-1}, B^{i-1}, C^i).$$

APPENDIX B
PROOF OF THEOREM 1 FOR GENERAL L

By Fano's inequality, we have for $j \in [1 : L]$,

$$\sum_{f \in \text{Cut}(\mathcal{P}), s_f \in \mathcal{V}_j} n(R_f - \epsilon_n)$$
$$\le I(\tilde{M}_j; \tilde{Y}_0^n, \tilde{Y}_1^n, \ldots, \tilde{Y}_{j-1}^n | \tilde{M}_0, \ldots, \tilde{M}_{j-1}).$$

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

$$\sum_{f \in \text{Cut}(\mathcal{P})} n(R_f - \epsilon_n)$$
$$\le I(\emptyset, \tilde{Y}_0^n, \tilde{Y}_1^n, \ldots, \tilde{Y}_{L-1}^n \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_L).$$

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 $\tilde{X}_{ki}$ is a function of $(\tilde{M}_0, \tilde{M}_k, \tilde{Y}_k^{i-1})$, and (c) follows from Markov conditions guaranteed by the memoryless property of the network. ∎

$$I(\emptyset, \tilde{Y}_0^i, \tilde{Y}_1^i, \ldots, \tilde{Y}_{L-1}^i \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_L)$$
$$\leq\ I(\emptyset, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1}, \ldots, \tilde{Y}_{L-1}^{i-1} \to \tilde{M}_0, \tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_L)$$
$$+ I(\emptyset, \tilde{Y}_{0i}, \tilde{Y}_{1i}, \ldots, \tilde{Y}_{L-1,i} \to \tilde{X}_{0i}, \tilde{X}_{1i}, \tilde{X}_{2i} \ldots, \tilde{X}_{Li}) \qquad (11)$$

$$I(\emptyset, \tilde{Y}_0^i, \tilde{Y}_1^i, \ldots, \tilde{Y}_{L-1}^i \to \tilde{M}_0, \tilde{M}_1, \ldots, \tilde{M}_L) - I(\emptyset, \tilde{Y}_0^{i-1}, \tilde{Y}_1^{i-1}, \ldots, \tilde{Y}_{L-1}^{i-1} \to \tilde{M}_0, \tilde{M}_1, \ldots, \tilde{M}_L)$$
$$= I(\emptyset, \tilde{Y}_{0i}, \ldots, \tilde{Y}_{L-1,i} \to \tilde{M}_0, \ldots, \tilde{M}_L \,\|\, \emptyset, \tilde{Y}_0^{i-1}, \ldots, \tilde{Y}_{L-1}^{i-1})$$
$$\overset{(a)}{\leq} \sum_{j=1}^{L} I\big((\tilde{M}_k)_{k=j}^{L}, (\tilde{Y}_k^{i-1})_{k=j}^{L-1}; \tilde{Y}_{j-1,i} \,\big|\, (\tilde{Y}_{ki})_{k=0}^{j-2}, (\tilde{M}_k)_{k=0}^{j-1}, (\tilde{Y}_k^{i-1})_{k=0}^{j-1}\big)$$
$$\overset{(b)}{=} \sum_{j=1}^{L} I\big((\tilde{M}_k)_{k=j}^{L}, (\tilde{Y}_k^{i-1})_{k=j}^{L-1}; \tilde{Y}_{j-1,i} \,\big|\, (\tilde{Y}_{ki})_{k=0}^{j-2}, (\tilde{M}_k)_{k=0}^{j-1}, (\tilde{Y}_k^{i-1})_{k=0}^{j-1}, (\tilde{X}_{ki})_{k=0}^{j-1}\big)$$
$$\leq \sum_{j=1}^{L} I\big(M, (\tilde{Y}_k^{i-1})_{k=0}^{L-1}; \tilde{Y}_{j-1,i} \,\big|\, (\tilde{Y}_{ki})_{k=0}^{j-2}, (\tilde{X}_{ki})_{k=0}^{j-1}\big)$$
$$\leq \sum_{j=1}^{L} I\big(M, (\tilde{Y}_k^{i-1})_{k=0}^{L-1}, (\tilde{X}_{ki})_{k=j}^{L}; \tilde{Y}_{j-1,i} \,\big|\, (\tilde{Y}_{ki})_{k=0}^{j-2}, (\tilde{X}_{ki})_{k=0}^{j-1}\big)$$
$$\overset{(c)}{=} \sum_{j=1}^{L} I\big((\tilde{X}_{ki})_{k=j}^{L}; \tilde{Y}_{j-1,i} \,\big|\, (\tilde{Y}_{ki})_{k=0}^{j-2}, (\tilde{X}_{ki})_{k=0}^{j-1}\big)$$
$$= I(\emptyset, \tilde{Y}_{0i}, \tilde{Y}_{1i}, \ldots, \tilde{Y}_{L-1,i} \to \tilde{X}_{0i}, \tilde{X}_{1i}, \ldots, \tilde{X}_{Li}). \qquad (12)$$

REFERENCES

[1] A. El Gamal, "On information flow in relay networks," in *Proc. IEEE National Telecom Conf.*, Nov. 1981.

[2] L. R. Ford, Jr. and D. R. Fulkerson, "Maximal flow through a network," *Canad. J. Math.*, vol. 8, pp. 399–404, 1956.

[3] T. C. Hu, "Multi-commodity network flows," *Oper. Res.*, vol. 11, no. 3, pp. 344–360, May–June 1963.

[4] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[5] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, October 2003.

[6] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

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

[8] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[9] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, October 2007.

[10] S. Kannan, A. Raja, and P. Viswanath, "Local PHY + global flow: A layering principle for wireless networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, August 2011.

[11] S. Kannan and A. Raja and P. Viswanath, "Approximately optimal wireless broadcasting," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7154–7167, 2012.

[12] S. H. Lim, K. T. Kim, and Y.-H. Kim, "Distributed decode–forward for broadcast," in *Proc. IEEE Inf. Theory Workshop*, Hobart, Australia, Nov. 2014.

[13] M. X. Goemans, S. Iwata, and R. Zenklusen, "An algorithmic framework for wireless information flow," in *Proc. 47th Ann. Allerton Conf. Comm. Control Comput.*, Monticello, Illinois, October 2009.

[14] S. Kamath, D. Tse, and V. Anantharam, "Generalized network sharing outer bound and the two-unicast problem," in *Proc. Int. Symp. Netw. Coding*, Beijing, China, July 2011.

[15] S. Kamath, S. Kannan, and P. Viswanath, "Network capacity under traffic symmetry: Wireline and wireless networks," *IEEE Transactions On Information Theory*, vol. 60, no. 9, pp. 5457–5469, September 2014.

[16] I. Shomorony and S. Avestimehr, "A generalized cut-set bound for deterministic multi-flow networks and its applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, Hawaii, July 2014.

[17] E. Ardestanizadeh, M. Franceschetti, T. Javidi, and Y.-H. Kim, "Wiretap channel with secure rate-limited feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5353–5361, December 2009.

[18] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[19] J. Massey, "Causality, feedback and directed information," in *Proc. IEEE Int. Symp. Inf. Theory Appl.*, Nov. 1990.

[20] H. H. Permuter, Y.-H. Kim, and T. Weissman, "Interpretations of directed information in portfolio theory, data compression, and hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 3248–3259, Jun. 2011.

[21] J. Massey and P. Massey, "Conservation of mutual and directed information," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, 2005, pp. 157–158.