The Multicast Capacity of Deterministic Relay Networks With No Interference

Niranjan Ratnakar and Gerhard Kramer, *Member, IEEE*

***Abstract*—The multicast capacity is determined for networks that have deterministic channels with broadcasting at the transmitters and no interference at the receivers. The multicast capacity is shown to have a cut-set interpretation. It is further shown that one cannot always layer channel and network coding in such networks. The proof of the latter result partially generalizes to discrete memoryless broadcast channels and is used to bound the common rate for problems where one achieves a cut bound on throughput.**

***Index Terms*—Broadcasting, capacity, multicast, network coding, relaying.**

I. Introduction

CONSIDER a network represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ and $\mathcal{E}$ are the sets of vertices and directed edges, respectively (see Fig. 1). For example, the graph might represent a communication network where the vertices are terminals and the edges are channels. The usual network model lets every edge be activated independently (for instance, see [1]). In this paper, we are interested in a modification of this model that is known as a deterministic relay network with no interference. Such networks have one channel input x_u associated with every vertex u in $\mathcal{V}$, and one channel output $y_{u,v}$ associated with every edge (u, v) in $\mathcal{E}$. By *deterministic*, we mean that $y_{u,v}$ is a deterministic function $h_{u,v}(x_u)$ of x_u (see Fig. 2 for examples). By *no interference*, we mean that $y_{u,v}$ is a function of no inputs other than x_u. The former restriction clearly permits *broadcasting*, in the sense that the outgoing edges of a vertex share a common input. In fact, the set

$$\{h_{u,v}(\cdot) : u \text{ fixed}, v \in \mathcal{V}, (u, v) \in \mathcal{E}\}$$

is a deterministic broadcast channel (BC) [2, Sec. 1]. We restrict x_u and $y_{u,v}$ to have discrete and finite alphabets $\mathcal{X}_u$ and $\mathcal{Y}_{u,v}$, respectively.

The above model might at first glance seem fundamentally different from the usual network model since it takes both network and symbol-level, or physical-layer, effects into account.

Manuscript received March 14, 2005; revised January 23, 2006. The work of N. Ratnakar was supported in part by Bell Labs, Lucent Technologies and the Vodafone Foundation. The work of G. Kramer was supported in part by the Board of Trustees of the University of Illinois Subaward no. 04-217 under National Science Foundation Grant CCR-0325673. The material in this paper was presented in part at the First Workshop on Network Coding, Theory and Applications, Riva del Garda, Italy, April 2005, and the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

N. Ratnakar is with the Coordinated Science Laboratory, University of Illinois at Urbana Champaign, Urbana IL 61801 USA (e-mail: ratnakar@ifp.uiuc.edu).

G. Kramer is with Bell Laboratories, Lucent Technologies, Murray Hill NJ 07974 USA (e-mail: gkr@bell-labs.com).

Communicated by N. Cai, Guest Editor.

Digital Object Identifier 10.1109/TIT.2006.874431

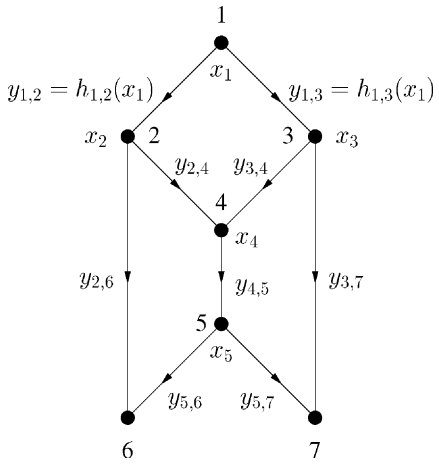


Fig. 1. Example of a deterministic relay network with no interference.

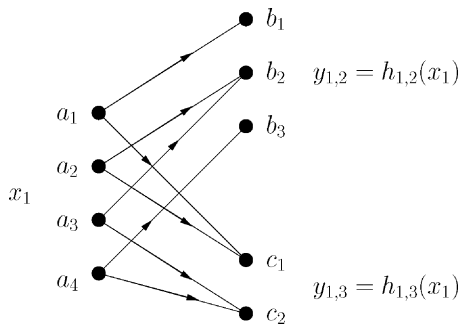


Fig. 2. A graph representing examples of $h_{1,2}(\cdot)$ and $h_{1,3}(\cdot)$ in Fig. 1.

However, consider Fig. 1 and suppose that the x_u, $u = 1, 2, 3, 5$, are vectors of length 2. Suppose further that the $h_{u,v}(\cdot)$, $(u, v) \in \mathcal{E}$, are chosen so that every $y_{u,v}$ is a different entry of x_u, e.g., let $x_2 = [x_{2,6}, x_{2,4}]$ and let $y_{2,6} = x_{2,6}$ and $y_{2,4} = x_{2,4}$. We then recover the usual network model where the capacity of edge (u, v) is $\log_2 |\mathcal{Y}_{u,v}|$ bits per edge use, where $|\mathcal{Y}_{u,v}|$ is the cardinality of $\mathcal{Y}_{u,v}$. Formally, the relations $y_{u,v} = h_{u,v}(x_u)$ do impose the extra constraint that the edges coming out of vertex u are all activated simultaneously. However, it turns out that this synchronization requirement does not impact the capacity of the network. It therefore poses no real constraint if one studies communication rates. We discuss this issue in more detail in Section II.

Historically, the above networks were introduced by Aref [3] and for brevity we refer to them as *Aref networks*. Aref determined the *unicast* capacity of his networks by showing that a protocol that uses *binning* [4, Sec. 14.4] at the physical layer

0018-9448/$20.00 © 2006 IEEE

and *routing* at the network layer achieves an information-theoretic cut-set bound [4, Sec. 14.10] that generalizes the usual cut-set bound of networking theory.

Our aim is to characterize the *multicast capacity* of Aref networks, i.e., the maximum rate at which one can simultaneously transmit one message from one vertex to several other vertices. We show that the multicast capacity is given by the aforementioned information-theoretic cut-set bound. We thereby generalize a result of Ahlswede *et al.* [1] to include deterministic broadcasting at every transmitter. A second result we show is that one cannot in general separate physical- and network-layer coding in Aref networks. The latter result is in contrast to the optimality of layering for multicasting in networks of discrete memoryless channels (DMCs) [5] and networks of two-way channels where the capacity region of each two-way channel is given by Shannon's outer bound for this channel [6], [7]. A third result we develop is a bound on the rate of a common message in discrete memoryless BCs.

This paper is organized as follows. In Section II, we define the network model and problem in more detail. In Section III, we derive the multicast capacity of Aref networks. In Section IV, we show that one cannot always layer channel and network coding. We further develop theory for discrete memoryless BCs.

II. Preliminaries

A. Notation and Clocking

Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ as described in Section I. We associate a discrete and finite random variable X_u with vertex u. We write $P_{X_u}(\cdot)$ for the distribution of X_u, and $P_{X_u}(x_u)$ or $P(x_u)$ for the probability that $X_u = x_u$. Similarly, we associate a discrete and finite random variable $Y_{u,v}$ with edge (u, v) where $Y_{u,v} = h_{u,v}(X_u)$. We use the usual notation $H(X)$, $H(X|Y)$, and $I(X;Y)$ for entropies and mutual information expressions [4, Ch. 2], [8, Ch. 2]. For $\mathcal{S} \subseteq \mathcal{V}$ and $\mathcal{S}' \subseteq \mathcal{V}$, we define $x_{\mathcal{S}} = [x_u : u \in \mathcal{S}]$ and

$$y_{\mathcal{S},\mathcal{S}'} = [y_{u,v} : u \in \mathcal{S}, v \in \mathcal{S}', (u,v) \in \mathcal{E}].$$

We further define $y_{u,\mathcal{S}'} = y_{\{u\},\mathcal{S}'}$ and $y_{\mathcal{S},u} = y_{\mathcal{S},\{u\}}$.

Every vertex u transmits and receives the N symbols

$$x_u^N = x_u^{(1)}, x_u^{(2)}, \ldots, x_u^{(N)}$$
$$y_{\mathcal{V},u}^N = y_{\mathcal{V},u}^{(1)}, y_{\mathcal{V},u}^{(2)}, \ldots, y_{\mathcal{V},u}^{(N)}.$$

We wish to define how $x_u^{(i)}$ is related to $y_{w,u}^{(j)}$. However, this depends on how and when the vertices are activated. For example, a simple approach is to activate every vertex at regular intervals according to a central *clock*. The clock ticks N times and at clock tick i vertex u transmits $x_u^{(i)}$ and receives $y_{\mathcal{V},u}^{(i)}$. The clocking assumption makes it easy to define how the vertices operate: the input $x_u^{(i)}$ must be a function of messages at vertex u and the past outputs

$$y_{\mathcal{V},u}^{(1)}, y_{\mathcal{V},u}^{(2)}, \ldots, y_{\mathcal{V},u}^{(i-1)}.$$

The clocking requirement might seem to be restrictive. However, as we proceed to show in Section III-B, the capacity of any nonclocked network is upper-bounded by a cut-set bound for clocked networks. We further show that clocked networks achieve this cut-set bound, and hence, the clocked network capacity is the largest possible. We therefore restrict attention to clocked networks for simplicity.

B. Problem and Coding

The multicasting problem has one source vertex (say vertex 1) and several destination vertices that we collect in a set $\mathcal{T}$. The source vertex has a message M that is uniformly distributed over $\{1, 2, \ldots, 2^{NR}\}$ where R is the *rate* and where we assume that NR is an integer for simplicity. A (clocked) communication strategy consists of encoding functions $f_u^{(i)}(\cdot)$, $u \in \mathcal{V}$, $i = 1, 2, \ldots, N$, and decoding functions $\hat{m}_t(\cdot)$, $t \in \mathcal{T}$.

- *Encoders*. Suppose $M = m$. At time i, vertex 1 transmits $x_1^{(i)} = f_1^{(i)}(m)$ and vertex u, $u \neq 1$, transmits
$$x_u^{(i)} = f_u^{(i)}\left(y_{\mathcal{V},u}^{(1)}, y_{\mathcal{V},u}^{(2)}, \ldots, y_{\mathcal{V},u}^{(i-1)}\right).$$
- *Decoders*. After time N, each destination vertex t puts out an estimate $\hat{m}_t(y_{\mathcal{V},t}^N)$.

The error probability is

$$P_e = \Pr\left[\bigcup_{t \in \mathcal{T}} \left\{\hat{m}_t\left(Y_{\mathcal{V},t}^N\right) \neq M\right\}\right]. \tag{1}$$

The rate R is said to be *achievable* if, for any $\epsilon > 0$, there exist encoders and decoders that make $P_e \leq \epsilon$ for some N. The *multicast capacity* C is the supremum of the achievable rates.

C. Cuts and Values

Consider a set $\mathcal{S}$, $\mathcal{S} \subseteq \mathcal{V}$, and let $\overline{\mathcal{S}}$ be its complement in $\mathcal{V}$. $\mathcal{S}$ is called a *cut* if $1 \in \mathcal{S}$ and $\overline{\mathcal{S}}$ contains one or more destination vertices, i.e., if $\overline{\mathcal{S}} \cap \mathcal{T} \neq \emptyset$. We denote the set of all cuts as Λ. The *boundary* of a cut $\mathcal{S}$ is defined as

$$\beta(\mathcal{S}) = \{u : (u,v) \in \mathcal{E}, u \in \mathcal{S}, v \in \overline{\mathcal{S}}\}. \tag{2}$$

For fixed input distributions $P_{X_1}(\cdot), P_{X_2}(\cdot), \ldots, P_{X_{|\mathcal{V}|}}(\cdot)$, we define the *value* of a cut $\mathcal{S}$ as

$$\text{Value}(\mathcal{S}) = \sum_{u \in \beta(\mathcal{S})} H(Y_{u,\overline{\mathcal{S}}}). \tag{3}$$

Note that the value of a cut depends on the input distributions but we do not explicitly include these as arguments in (3).

D. Typical Sequences and Sets

We use δ-typical sequences for our achievability proofs. Let $\mathcal{X}^n$ be the n-fold Cartesian product of $\mathcal{X}$. Let $\nu_{\underline{x}}(a)$ be the number of times the letter a occurs in the vector (or sequence) $\underline{x}$ of length n. The *empirical frequency* of a in $\underline{x}$ is

$$\pi_{\underline{x}}(a) = \frac{\nu_{\underline{x}}(a)}{n}. \tag{4}$$

Let $\mathcal{X}$ be the (discrete and finite) alphabet of the entries in $\underline{x}$, and let $\delta > 0$. Let X be a random variable with alphabet $\mathcal{X}$

and distribution $P_X(\cdot)$. The sequence $\underline{x}$ is said to be (robustly) δ-*typical* with respect to X if

$$\left|\pi_{\underline{x}}(a) - P_X(a)\right| \leq \delta \cdot P_X(a) \tag{5}$$

for all $a \in \mathcal{X}$ (see [9]). The set of δ-typical sequences of length n with respect to X is denoted $T_\delta^n(X)$. One can similarly define δ-typical sequences and sets with respect to pairs of random variables (see (57) in the Appendix). We further define the set of conditionally δ-typical sequences with respect to X, Y, and $\underline{x}$ as

$$T_\delta^n(Y|\underline{x}) = \{\underline{y} \in \mathcal{Y}^n : (\underline{x}, \underline{y}) \in T_\delta^n(X,Y)\}. \tag{6}$$

Some of the key properties of these objects are listed in the Appendix .

Remark 1: We underline symbols when considering sequences in the context of δ-typicality. For example, we write

$$\underline{x}_u = [x_u^{(1)}, x_u^{(2)}, \ldots, x_u^{(n)}]$$

and

$$\underline{x}_{\mathcal{S}} = [\underline{x}_u : u \in \mathcal{S}]$$

if these vectors are being tested for δ-typicality. The meaning of $\underline{y}_{\mathcal{S},\mathcal{S}'}$, $\underline{y}_{u,\mathcal{S}'}$, and $\underline{y}_{\mathcal{S},u}$ is similar.

III. Main Result

Consider an acyclic Aref network and number the vertices so that $(u,v) \in \mathcal{E}$ implies that $u < v$. We consider only the subgraph whose vertices and edges are on the paths from the source to the destinations. The following is our main result.

Theorem 1: The multicast capacity of an acyclic Aref network is

$$C = \max_{P_{X_1}(\cdot), P_{X_2}(\cdot), \ldots, P_{X_{|\mathcal{V}|}}(\cdot)} \min_{\mathcal{S} \in \Lambda} \text{Value}(\mathcal{S}). \tag{7}$$

For example, the multicast capacity of the network shown in Fig. 3 is

$$C = \max_{P_{X_1}(\cdot) P_{X_2}(\cdot)} \min\{\, H(Y_{1,2}Y_{1,3}),\, H(Y_{1,3}) + H(Y_{2,3})\,\}. \tag{8}$$

Suppose that the BC from vertex 1 to vertices 2 and 3 is given by Fig. 2. The best $P_{X_1}(\cdot)$ is uniform and (8) evaluates to

$$C = \min\{\, 2,\, 1 + C_{2,3}\,\} \tag{9}$$

where $C_{2,3} = \log_2 |\mathcal{Y}_{2,3}|$ is the capacity of edge $(2,3)$.

Remark 2: Theorem 1 has a cut-set interpretation as discussed further in Section II-C and Section III-B.

Remark 3: Theorem 1 includes the main result of [1] and [3, Ch. 3] as special cases. Our proof technique uses the same steps as in [1] with a few modifications.

Remark 4: We obtain an identical expression (7) for the multicast capacity of Aref networks with cycles by constructing a time-parameterized graph from $\mathcal{G}$ as described in [1, Sec. V.B].

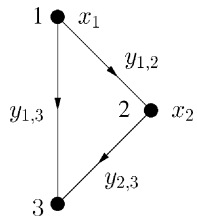


Fig. 3. An example of an Aref network.

A. Achievability

We divide the message m into B parts m_b, $b = 1, 2, \ldots, B$, that each take on values in $\{1, 2, \ldots, 2^{nR}\}$. We code in $B + |\mathcal{V}| - 2$ blocks of length n, i.e., we set $N = (B + |\mathcal{V}| - 2) \cdot n$ for $B \geq 1$. The overall rate is $R \cdot B/(B + |\mathcal{V}| - 2)$ which we make close to R by choosing B to be large.

Let $f_v^n(\cdot) = [f_v^{(i)}(\cdot) : i = 1, 2, \ldots, n]$ and set $f_v^{(i+b\cdot n)}(\cdot) = f_v^{(i)}(\cdot)$ for $i = 1, 2, \ldots, n$, all b, and all v, i.e., we use the same encoding function(s) for each block. We encode so that every m_b influences only the $x_u^{(i)}$ at clock ticks $i = (b+u-2)n+1$ to $i = (b+u-1)n$, i.e., every m_b influences exactly one block of length n at every vertex. We label this block as $\underline{x}_u(m_b)$ and its corresponding output blocks as $\underline{y}_{u,\mathcal{V}}(m_b)$.

Codebooks: Choose some $P_{X_1}(\cdot), P_{X_2}(\cdot), \ldots, P_{X_{|\mathcal{V}|}}(\cdot)$. At vertex 1, choose $f_1^n(\cdot)$ to map each of the indices in $\{1, 2, \ldots, 2^{nR}\}$ to a sequence $\underline{x}_1$ drawn uniformly from $T_\delta^n(X_1)$. At vertex u, choose $f_u^n(\cdot)$ to map each sequence in $T_\delta^n(Y_{\mathcal{V},u})$ to a sequence drawn uniformly from $T_\delta^n(X_u)$. Note that we have $\underline{y}_{u,v} \in T_\delta^n(Y_{u,v})$ for all (u,v) since $\underline{x}_u \in T_\delta^n(X_u)$ (see Lemma 4 in the Appendix).

Encoding: Consider block b.

- Vertex $u = 1$ transmits $\underline{x}_1(m_b) = f_1^n(m_b)$ for $b = 1, 2, \ldots, B$ and a default codeword otherwise.
- Vertex u transmits a default codeword if $b < u$ or $b \geq B + u$. Otherwise, it transmits

$$\underline{x}_u(m_{b-u+1}) = f_u^n\left(\underline{y}_{\mathcal{V},u}(m_{b-u+1})\right).$$

For instance, the coding strategy for the network shown in Fig. 3 is given in Table I. Observe how the transmissions are pipelined. Since vertex u receives blocks only from vertices $w < u$, by block b vertex u has received $\underline{y}_{\mathcal{V},u}(m_{b-u+1})$ and can thus transmit $\underline{x}_u(m_{b-u+1})$ as a function of these vectors.

Decoding: Destination vertex t decodes m_b after block $(t + b - 2)$ by using the function

$$\hat{m}_t\left(\underline{y}_{\mathcal{V},t}(m_b)\right) = \begin{cases} \text{error}, & \text{if } \underline{y}_{\mathcal{V},t}(m_b') = \underline{y}_{\mathcal{V},t}(m_b) \text{ for some } m_b' \neq m_b \\ m_b, & \text{otherwise.} \end{cases} \tag{10}$$

Note that we are abusing notation by using the same expression $\hat{m}_t(\cdot)$ as in (1).

Analysis: We bound the error probability for decoding m_b and later bound the error probability for decoding the overall message m. We briefly drop the index b for convenience since we consider only blocks pertaining to m_b. For example, we write m, $\underline{x}_u(m)$, and $\underline{y}_{u,v}(m)$ for m_b, $\underline{x}_u(m_b)$, and $\underline{y}_{u,v}(m_b)$, respectively.

TABLE I
CODING STRATEGY FOR THE NETWORK OF FIG. 3 FOR $B = 3$

Block b	Message	1 Transmits	2 Receives	2 Transmits	3 Receives	Decoder output
1	m_1	$\underline{x}_1(m_1)$	$\underline{y}_{1,2}(m_1)$	$\cdot$	$\underline{y}_{1,3}(m_1)$	$\cdot$
2	m_2	$\underline{x}_1(m_2)$	$\underline{y}_{1,2}(m_2)$	$\underline{x}_2(m_1)$	$\underline{y}_{1,3}(m_2), \underline{y}_{2,3}(m_1)$	$\hat{m}_t(\underline{y}_{1,3}(m_1), \underline{y}_{2,3}(m_1))$
3	m_3	$\underline{x}_1(m_3)$	$\underline{y}_{1,2}(m_3)$	$\underline{x}_2(m_2)$	$\underline{y}_{1,3}(m_3), \underline{y}_{2,3}(m_2)$	$\hat{m}_t(\underline{y}_{1,3}(m_2), \underline{y}_{2,3}(m_2))$
4	$\cdot$	$\cdot$	$\cdot$	$\underline{x}_2(m_3)$	$\underline{y}_{2,3}(m_3)$	$\hat{m}_t(\underline{y}_{1,3}(m_3), \underline{y}_{2,3}(m_3))$

Consider a destination vertex t. Let $\overline{P}_e(t, m, m')$ be the average probability that vertex t cannot distinguish between m and m', where the average is over the ensemble of encoding functions. Let $\mathcal{S}(m, m')$ be the set of vertices u for which

$$\underline{y}_{\mathcal{V},u}(m) \neq \underline{y}_{\mathcal{V},u}(m') \tag{11}$$

i.e., $\mathcal{S}(m, m')$ is the set of the vertices that *can* distinguish between m and m'. We view $\mathcal{S}(m, m')$ as a random variable that is a function of the encoding functions. We may as well consider $1 \in \mathcal{S}(m, m')$. Suppose vertex t cannot distinguish between m and m', so that $t \in \overline{\mathcal{S}}(m, m')$. $\mathcal{S}(m, m')$ is thus a cut between vertices 1 and t. Let Λ_t be the set of such cuts, i.e., we define $\Lambda_t = \{\mathcal{S} \subset \mathcal{V} : 1 \in \mathcal{S}, t \in \overline{\mathcal{S}}\}$. We can write

$$\begin{aligned}\overline{P}_e(t, m, m') &= \Pr\left[\bigcup_{\mathcal{S}\in\Lambda_t} \{\mathcal{S}(m, m') = \mathcal{S}\}\right] \\ &= \sum_{\mathcal{S}\in\Lambda_t} \Pr[\mathcal{S}(m, m') = \mathcal{S}]. \end{aligned} \tag{12}$$

Furthermore, we claim that a necessary condition for the event $\mathcal{S}(m, m') = \mathcal{S}$ is that one must have

$$\underline{y}_{u,\overline{\mathcal{S}}}(m) = \underline{y}_{u,\overline{\mathcal{S}}}(m') \tag{13}$$

for *all* $u \in \beta(\mathcal{S}(m, m'))$. To see this, note that if (13) was not true for some $u \in \beta(\mathcal{S}(m, m'))$, then there is a vertex $v \in \overline{\mathcal{S}}(m, m')$ that can distinguish between m and m', contradicting our definition of $\mathcal{S}(m, m')$. We can thus write

$$\begin{aligned}\Pr[\mathcal{S}(m, m') = \mathcal{S}] &\leq \Pr\left[\bigcap_{u\in\beta(\mathcal{S})} \left\{\underline{Y}_{u,\overline{\mathcal{S}}}(m) = \underline{Y}_{u,\overline{\mathcal{S}}}(m')\right\}\right] \\ &= \prod_{u\in\beta(\mathcal{S})} \Pr\left[\underline{Y}_{u,\overline{\mathcal{S}}}(m) = \underline{Y}_{u,\overline{\mathcal{S}}}(m')\right] \end{aligned} \tag{14}$$

where the equality follows because the $\underline{X}_u(m)$, $u \in \mathcal{V}$, $m = 1, 2, \ldots, 2^{nR}$ are statistically independent.

We proceed to bound (14). We have

$$(\underline{x}_u(m'), \underline{y}_{u,\overline{\mathcal{S}}}(m')) \in T_\delta^n(X_u, Y_{u,\overline{\mathcal{S}}})$$

by Lemma 4 in the Appendix. The event (13) thus implies

$$\left(\underline{x}_u(m'), \underline{y}_{u,\overline{\mathcal{S}}}(m)\right) \in T_\delta^n(X_u, Y_{u,\overline{\mathcal{S}}}). \tag{15}$$

But note that $\underline{X}_u(m')$ is independent of $\underline{Y}_{u,\overline{\mathcal{S}}}(m)$. The probability of (15) occurring is thus

$$\left|T_\delta^n(X_u|\, \underline{y}_{u,\overline{\mathcal{S}}}(m))\right| \Big/ \left|T_\delta^n(X_u)\right|. \tag{16}$$

We use Lemmas 2 and 3 in the Appendix to bound

$$|T_\delta^n(X_u)| \geq (1 - \epsilon_\delta(n)) \cdot 2^{n(1-\delta)H(X_u)} \tag{17}$$

$$|T_\delta^n(X_u|\underline{y}_{u,\overline{\mathcal{S}}}(m))| \leq 2^{n(1+\delta_2)H\left(X_u|Y_{u,\overline{\mathcal{S}}}\right)} \tag{18}$$

where $\epsilon_\delta(n) \to 0$ as $n \to \infty$. Inserting (17) and (18) into (16), we have

$$\begin{aligned}&\Pr\left[\underline{Y}_{u,\overline{\mathcal{S}}}(m) = \underline{Y}_{u,\overline{\mathcal{S}}}(m')\right] \\ &\quad\leq (1-\epsilon_\delta(n))^{-1} \cdot 2^{n(\delta H(X_u)+\delta_2 H(X_u|Y_{u,\overline{\mathcal{S}}}))} \cdot 2^{-nH\left(Y_{u,\overline{\mathcal{S}}}\right)} \end{aligned} \tag{19}$$

where we have used $H(Y_{u,\overline{\mathcal{S}}}|X_u) = 0$. Inserting (19) into (14), using $|\beta(\mathcal{S})| \leq |\mathcal{E}|$, and defining

$$\delta' = \max_{u,\mathcal{S}} \delta H(X_u) + \delta_2 H(X_u|Y_{u,\overline{\mathcal{S}}})$$

we have

$$\begin{aligned}&\Pr[\mathcal{S}(m, m') = \mathcal{S}] \\ &\qquad\leq (1 - \epsilon_\delta(n))^{-|\mathcal{E}|} \cdot 2^{n|\mathcal{E}|\delta'} \cdot 2^{-n\,\mathrm{Value}(\mathcal{S})}. \end{aligned} \tag{20}$$

Inserting (20) into (12), and using the fact that the number of cuts is less than $2^{|\mathcal{V}|}$, we have

$$\begin{aligned}&\overline{P}_e(t, m, m') \\ &\quad\leq (1 - \epsilon_\delta(n))^{-|\mathcal{E}|} \cdot 2^{|\mathcal{V}|+n|\mathcal{E}|\delta'} \cdot 2^{-n \min_{\mathcal{S}\in\Lambda_t} \mathrm{Value}(\mathcal{S})}. \end{aligned} \tag{21}$$

We now add the index b back to m_b. Let $\overline{P}_e(m)$ be the average error probability when the (overall) message m was transmitted. We use the union bound over all B blocks, all destinations t, and all $m'_b \neq m_b$ to write

$$\begin{aligned}\overline{P}_e(m) &\leq \sum_{b=1}^{B} \sum_{t\in\mathcal{T}} \sum_{m'_b\neq m_b} \overline{P}_e(t, m_b, m'_b) \\ &\leq B \cdot |\mathcal{T}| \cdot (2^{nR} - 1) \cdot (1 - \epsilon_\delta(n))^{-|\mathcal{E}|} \\ &\qquad \cdot 2^{|\mathcal{V}|+n|\mathcal{E}|\delta'} \cdot 2^{-n \min_{\mathcal{S}\in\Lambda} \mathrm{Value}(\mathcal{S})}. \end{aligned} \tag{22}$$

The average error probability (22) can be made small if n is large and

$$R < -|\mathcal{E}|\delta' + \min_{\mathcal{S}\in\Lambda} \mathrm{Value}(\mathcal{S}). \tag{23}$$

Finally, we optimize over all input distributions, choose δ and δ_2 small, and choose n and B large. The result is that we can make the overall rate approach C in (7) while at the same time ensuring that $P_e \leq \epsilon$ for any positive ϵ.

B. Converse

A clocked Aref network is a special case of the networks described in [4, Sec. 14.10]. We can thus simply apply [4, Theorem 14.10.1]. However, we supply a shortened proof of this theorem for Aref networks to show that no step requires the clocking assumption.

Proposition 1: The multicasting capacity of an Aref network is bounded by

$$C \leq \max_{P_{X_1X_2\cdots X_{|\mathcal{V}|}}(\cdot)} \min_{\mathcal{S}\in\Lambda} H(Y_{\beta(\mathcal{S}),\overline{\mathcal{S}}}) \tag{24}$$

irrespective of whether the network is clocked or not.

Proof: Fano's inequality [4, p. 39] requires that

$$NR \leq I(M; Y^N_{\mathcal{V},t}) \tag{25}$$

for reliable communication for any t with $t \in \mathcal{T}$. Let $\mathcal{S} \in \Lambda_t$ so that $Y^N_{\mathcal{V},t}$ is a function of $Y^N_{\beta(\mathcal{S}),\overline{\mathcal{S}}}$. We use standard steps to bound

$$\begin{aligned} I(M; Y^N_{\mathcal{V},t}) &\leq H(Y^N_{\mathcal{V},t}) \\ &\leq H\left(Y^N_{\beta(\mathcal{S}),\overline{\mathcal{S}}}\right) \\ &\leq \sum_{i=1}^{N} H\left(Y^{(i)}_{\beta(\mathcal{S}),\overline{\mathcal{S}}}\right) \qquad (26) \\ &= N \cdot H\left(Y^{(Q)}_{\beta(\mathcal{S}),\overline{\mathcal{S}}} | Q\right) \qquad (27) \\ &\leq N \cdot H\left(Y^{(Q)}_{\beta(\mathcal{S}),\overline{\mathcal{S}}}\right) \qquad (28) \end{aligned}$$

where (27) follows by defining Q to be a random variable that is uniform over $\{1, 2, \ldots, N\}$. We now argue as in [4, p. 447] that there exist $Y_{u,v}$, $(u, v) \in \mathcal{E}$, that satisfy (24). Observe that no step requires the clocking assumption. □

We next optimize the joint distribution in (24).

Lemma 1: The right-hand side of (24) is maximized by independent inputs.

Proof: For any fixed $P_{X_1X_2\cdots X_{|\mathcal{V}|}}(\cdot)$, we have

$$H(Y_{\beta(\mathcal{S}),\overline{\mathcal{S}}}) \leq \sum_{u\in\beta(\mathcal{S})} H(Y_{u,\overline{\mathcal{S}}}) \tag{29}$$

because conditioning cannot increase entropy. Moreover, the right-hand side of (29) depends only on the marginals of $P_{X_1X_2\cdots X_{|\mathcal{V}|}}(\cdot)$. Thus, $H(Y_{\beta(\mathcal{S}),\overline{\mathcal{S}}})$ cannot decrease by replacing $P_{X_1X_2\cdots X_{|\mathcal{V}|}}(\cdot)$ by the product of its marginals. □

We remark that both Proposition 1 and Lemma 1 apply to cyclic Aref networks as well as acyclic ones. Finally, note that the right-hand side of (29) is $\mathrm{Value}(\mathcal{S})$. Hence, we have the result that (7) is an upper bound on C.

C. Relation to the Unicast Capacity

The multicasting capacity of the usual networks without broadcasting is

$$C = \min_{t\in\mathcal{T}} C_{1,t} \tag{30}$$

where $C_{1,t}$ is the unicast capacity from vertex 1 to vertex t. However, for Aref networks such a relationship is not necessarily true. Consider, e.g., the BC in Fig. 4 where the graph on the right in Fig. 4 specifies $h_{1,2}(\cdot)$ and $h_{1,3}(\cdot)$. It is easy to check that $C_{1,2} = C_{1,3} = \log_2(3)$ bits/use by choosing three out of four of the input letters to have probability $1/3$. The multicast capacity is, however, only $C = 1.5$ bits/use, and is achieved only if all four input letters have probability $1/4$.

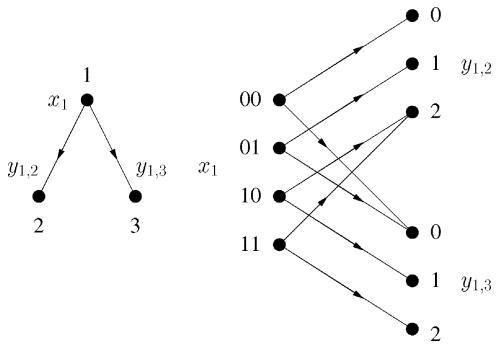


Fig. 4. An Aref network for which $C < \min_{t\in\mathcal{T}} C_{1,t}$.

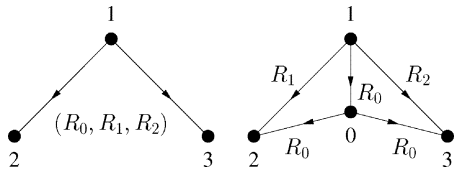


Fig. 5. An Aref network operating at rates (R_0, R_1, R_2) and its point-to-point link (PPL) network.

IV. Layering Channel and Network Coding

A practically interesting question is whether one can *layer* channel and network coding. It is not so obvious how to define such layering in general networks but for Aref networks a natural definition is as follows. First, on the *physical* layer one performs (local) channel coding for each BC to obtain a network comprised of point-to-point links (PPLs). Second, on the *network* layer one performs network coding for the resulting noise-free network. We consider more details of these layers.

Physical Layer: Suppose that vertex u has m outgoing edges and therefore m output vertices. Each of the $2^m - 1$ nonempty subsets of these vertices can be transmitted a different message. We assume that each message is treated as being statistically independent. An alternative approach with statistically dependent messages could, e.g., apply results of [10].

A BC with m outputs has a $(2^m - 1)$-dimensional capacity region $\mathcal{C}_u$ that is the set of rate-tuples at which one can transmit reliably to the m output vertices. For example, consider the BC with $m = 2$ shown on the left in Fig. 5; its capacity region $\mathcal{C}_1$ is a three-dimensional region. Suppose that $\underline{R}_1 = (R_0, R_1, R_2)$ is in $\mathcal{C}_1$, where R_0 is the common rate, R_1 the rate to vertex 2, and R_2 the rate to vertex 3. Given $\underline{R}_1$, the BC can effectively be replaced by a network $\mathrm{PPL}(\underline{R}_1)$ of PPLs shown on the right in Fig. 5. Observe that we have introduced one auxiliary vertex to the original network. More generally, we need to introduce $2^m - 1 - m$ auxiliary vertices to convert a BC to a PPL network.

We apply the above conversion to Aref networks, i.e., at every vertex u we replace the BC by a network $\mathrm{PPL}(\underline{R}_u)$. We call the overall network thus obtained a PPL network. We denote its graph as $\mathcal{G}_{\mathrm{PPL}}$.

Network Layer: We perform network coding on $\mathcal{G}_{\mathrm{PPL}}$. Interesting code constructions are known and have been investigated in detail in [11], [12].

A. Properties of Deterministic Broadcast Channels

A natural question is whether layering is optimal. Unfortunately, as we show by example, the answer is negative in general. We begin by developing several properties of deterministic BCs. We write $A - B - C$ to mean $I(A; C|B) = 0$, i.e., $A - B - C$ forms a Markov chain.

Claim 1: The capacity region of a deterministic BC with $Y_{1,2} = h_{1,2}(X_1)$ and $Y_{1,3} = h_{1,3}(X_1)$ is the set of nonnegative (R_0, R_1, R_2) satisfying

$$R_0 \le \min\left[I(T; Y_{1,2}), I(T; Y_{1,3})\right] \tag{31}$$
$$R_0 + R_1 \le H(Y_{1,2}) \tag{32}$$
$$R_0 + R_2 \le H(Y_{1,3}) \tag{33}$$
$$R_0 + R_1 + R_2 \le \min\left[I(T; Y_{1,2}), I(T; Y_{1,3})\right] + H(Y_{1,2}Y_{1,3}|T) \tag{34}$$

for some T alphabet $\mathcal{T}$ satisfying $|\mathcal{T}| \le |\mathcal{X}_1| + 2$.

Proof: See Problem 4.11 in [13, p. 391] and [14], [15]. □

Claim 2: In order to achieve

$$R_0 + R_1 + R_2 = H(Y_{1,2}Y_{1,3}) \tag{35}$$

in (34) the triple $(Y_{1,2}, Y_{1,3}, T)$ must form the Markov chains

$$Y_{1,3} - Y_{1,2} - T \tag{36}$$
$$Y_{1,2} - Y_{1,3} - T. \tag{37}$$

Proof: Expressions (34) and (35) give

$$H(Y_{1,2}Y_{1,3}) \le I(T; Y_{1,2}) + H(Y_{1,2}Y_{1,3}|T) \tag{38}$$
$$H(Y_{1,2}Y_{1,3}) \le I(T; Y_{1,3}) + H(Y_{1,2}Y_{1,3}|T). \tag{39}$$

Rearranging terms, we obtain

$$I(T; Y_{1,3}|Y_{1,2}) \le 0 \tag{40}$$
$$I(T; Y_{1,2}|Y_{1,3}) \le 0. \tag{41}$$

which gives (36) and (37) since mutual information is nonnegative. □

Claim 3: The triple $(Y_{1,2}, Y_{1,3}, T)$ satisfies the double Markov relations (36) and (37) if and only if there exist functions $f(\cdot)$ of $Y_{1,2}$ and $g(\cdot)$ of $Y_{1,3}$ such that

$$\text{i)} \quad f(Y_{1,2}) = g(Y_{1,3}) \quad \text{with probability 1} \tag{42}$$
$$\text{ii)} \quad [Y_{1,2}, Y_{1,3}] - [f(Y_{1,2}), g(Y_{1,3})] - T. \tag{43}$$

Note that $[Y_{1,2}, Y_{1,3}]$ is independent of T if $f(Y_{1,2}) = g(Y_{1,3})$ is a constant.

Proof: See Problem 4.25 in [13, p. 402]. □

Definition 1: (See Problem 3.12 in [13, p. 350].) The joint distribution $P_{XY}(\cdot)$ is *indecomposable* if there are no functions $f(\cdot)$ and $g(\cdot)$ with respective domains $\mathcal{X}$ and $\mathcal{Y}$ so that

- $\Pr[f(X) = g(Y)] = 1$ and
- $f(X)$ takes on at least two values with positive probability.

Claim 4: If (35) is valid, then the maximum common rate is

$$R_0 = \max H\left(f(Y_{1,2})\right) \tag{44}$$

where the maximization is over all $f(\cdot)$ and $g(\cdot)$ satisfying (42) and (43). In particular, we have $R_0 = 0$ if (35) is valid and $P_{Y_{1,2}Y_{1,3}}(\cdot)$ is indecomposable.

Proof: The validity of (35) implies that there are T, $f(\cdot)$, and $g(\cdot)$ satisfying (42) and (43). We find that, for such T, $f(\cdot)$, and $g(\cdot)$, we have

$$\begin{aligned} R_0 &\le I(T; Y_{1,2}) \\ &\le I(f(Y_{1,2}); Y_{1,2}) \\ &= H(f(Y_{1,2})) \end{aligned} \tag{45}$$

where the first step follows by (31) and the second by (42) and (43). Furthermore, the choice $T = f(Y_{1,2}) = g(Y_{1,3})$ permits $R_0 = H(f(Y_{1,2}))$ in Claim 1. □

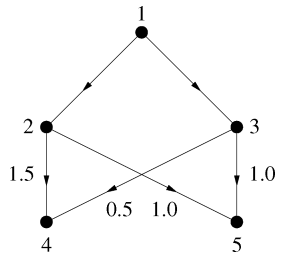


Fig. 6. An Aref network for which separating channel and network coding is suboptimal. The BC of vertex 1 is shown in Fig. 2.

B. A Counterexample

Consider the Aref network shown in Fig. 6 for which the BC at node 1 is shown in Fig. 2. The remaining edges $(2, 4), (2, 5), (3, 4)$, and $(3, 5)$ in Fig. 6 represent PPLs with the labeled capacities.

Claim 5: The multicast capacity of the network shown in Fig. 6 is $C = 2$. Furthermore, one achieves this rate only if the input X_1 is uniform.

Proof: Suppose X_1 is uniform. We compute

$$H(Y_{1,2}Y_{1,3}) = H(X_1) = 2, H(Y_{1,2}) = 1.5$$

and $H(Y_{1,3}) = 1$. We further find that we can achieve $\min_{S \in \Lambda} \text{Value}(S) = 2$. However, we also have

$$C \le H(Y_{1,2}Y_{1,3}) \le H(X_1) \tag{46}$$

where the first step follows by the cut-set bound and the second because the BCs are deterministic. Moreover, $H(X_1) \le 2$ with equality if and only if X_1 is uniform. □

We continue by restricting attention to uniform X_1.

Claim 6: If the input distribution of the BC in Fig. 2 is uniform then $P_{Y_{1,2}Y_{1,3}}(\cdot)$ is indecomposable.

Proof: Consider $f(\cdot)$ and $g(\cdot)$ with

$$\Pr[f(Y_{1,2}) = g(Y_{1,3})] = 1.$$

By inspecting Fig. 2, we find that $f(b_1) = f(b_2) = g(c_1)$ and $f(b_2) = f(b_3) = g(c_2)$. The functions $f(\cdot)$ and $g(\cdot)$ thus evaluate to a constant. □

Let the point (R_0, R_1, R_2) be a point in the capacity region of the BC in Fig. 2. We perform channel coding for this BC and obtain the PPL network shown in Fig. 7.

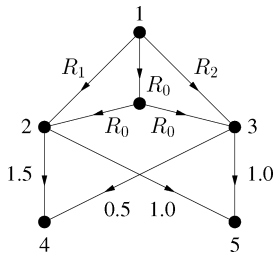


Fig. 7. The network $\mathcal{G}_{\mathrm{PPL}}$ for the Aref network in Fig. 6.

Claim 7: The PPL networks generated from the Aref network in Fig. 2 (and depicted in Fig. 7) all have multicast capacity less than 2.

Proof: The cut-set bound permits a multicast rate of 2 only if (35) is valid. Claims 4-6 thus imply that $R_0 = 0$ for uniform X_1. Applying the cut-set bound for the cuts $\mathcal{S} = \{1, 3, 5\}$ and $\mathcal{S} = \{1, 2, 4\}$ with $R_0 = 0$, we require

$$R_1 + 0.5 \geq 2$$
$$R_2 + 1 \geq 2.$$

We thus require $R_1 + R_2 \geq 2.5$ which is impossible by the cut-set bound. □

C. Discrete Memoryless BCs

The above theory for deterministic BCs leads to an apparently new result on the common rate for discrete memoryless BCs (see [2], [4, Sec. 14.6]). We generalize Claims 1, 2, and 4 where the first claim is a capacity region outer bound.

Claim 8: The capacity region of a discrete, memoryless BC $P_{Y_1Y_2|X}(\cdot)$ is a subset of the nonnegative (R_0, R_1, R_2) satisfying

$$R_0 \leq \min\left[I(T;Y_1), I(T;Y_2)\right] \tag{47}$$
$$R_0 + R_1 \leq I(X;Y_1) \tag{48}$$
$$R_0 + R_2 \leq I(X;Y_2) \tag{49}$$
$$R_0 + R_1 + R_2 \leq \min\left[I(T;Y_1), I(T;Y_2)\right] + I(X;Y_1Y_2|T) \tag{50}$$

where $T - X - [Y_1, Y_2]$ and $|\mathcal{T}| \leq |\mathcal{X}| + 2$.

Proof: The proof is similar to the proof of the converse of Claim 1 and is omitted. □

Remark 5: Claim 8 can be strengthened by taking advantage of the fact that the capacity region of a memoryless BC depends only on the marginals $P_{Y_1|X}(\cdot)$ and $P_{Y_2|X}(\cdot)$ [16].

Claim 9: In order to achieve

$$R_0 + R_1 + R_2 = I(X;Y_1Y_2) \tag{51}$$

in (50), the triple (Y_1, Y_2, T) must form the Markov chains

$$Y_1 - Y_2 - T \tag{52}$$
$$Y_2 - Y_1 - T. \tag{53}$$

Proof: The proof is similar to the proof of Claim 2 and is omitted. □

Claim 10: If (51) is valid, then the maximum common rate is bounded by

$$R_0 \leq \max H(f(Y_1)) \tag{54}$$

where the maximization is over all $f(\cdot)$ and $g(\cdot)$ satisfying (42) and (43) with $Y_{1,2}$ and $Y_{1,3}$ replaced by Y_1 and Y_2, respectively. In particular, we have $R_0 = 0$ if (51) is valid and $P_{Y_1Y_2}(\cdot)$ is indecomposable.

Proof: The proof is similar to the proof of Claim 4 and is omitted. Note that, as opposed to (44), the expression (54) is only a bound because Claim 8 is only a bound. □

Claim 10 applies to several classes of BCs. We consider one such class next.

Claim 11: BCs for which $P_{Y_1Y_2|X}(y_1, y_2|x) > 0$ for all y_1, y_2, and x have indecomposable $P_{Y_1Y_2}(\cdot)$ irrespective of the choice of input distribution $P_X(\cdot)$ (we assume that $P_{Y_1Y_2|X}(\cdot|\, x)$ is well-defined even if $P_X(x) = 0$).

Proof: Consider functions $f(\cdot)$ and $g(\cdot)$ on $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively. Let $f^{-1}(a) = \{y_1 : f(y_1) = a\}$. The condition

$$\Pr[f(Y_1) = g(Y_2)] = 1$$

requires that, for $P_{Y_2}(y_2) > 0$ and $g(y_2) = a$, we have

$$\sum_{y_1 \in f^{-1}(a)} P_{Y_1|Y_2}(y_1|y_2) = 1. \tag{55}$$

However, since $P_{Y_1Y_2|X}(y_1, y_2|x) > 0$ for all y_1, y_2 and x, our channels have $P_{Y_1Y_2}(y_1, y_2) > 0$ for all y_1 and y_2. We thus have $f^{-1}(a) = \mathcal{Y}_1$ so that $f(\cdot)$ evaluates to the constant a. □

For instance, consider an additive white Gaussian noise (AWGN) multiantenna BC. It is known that the the maximum throughput is given by (51) if one uses Sato's outer bound [16] (see Remark 5 and [17]–[20]). Moreover, by Claim 11, any discretized version of such a channel will have indecomposable $P_{Y_1Y_2}(\cdot)$. We thus find that $R_0 = 0$ is required to achieve the maximum throughput for the continuous channel (see the discussion in [8, Sec. 2.5]).

V. Conclusion

We have shown that the multicast capacity of Aref networks has a cut-set interpretation. We have further shown that one cannot always layer channel and network coding for multicasting in Aref networks. Some of the theory for the layering result extends to discrete memoryless BCs. Finally, we remark that the network model studied here might be considered an intermediate step toward modeling wireless networks because it includes broadcasting. Two recent works take a related approach and study packet erasure networks [21], [22].

Appendix
Robustly Typical Sequences

Let X be a random variable with discrete and finite alphabet $\mathcal{X}$ and distribution $P_X(\cdot)$. Recall that the set of δ-typical sequences of length n with respect to X is

$$T_\delta^n(X) = \left\{\underline{x} \in \mathcal{X}^n : \left|\pi_{\underline{x}}(a) - P_X(a)\right| \leq \delta \cdot P_X(a) \text{ for all } a \in \mathcal{X}\right\}. \tag{56}$$

For the technical lemmas below, we define the *support set* of X to be $\mathcal{S}_X = \{a \in \mathcal{X} : P_X(a) > 0\}$. Let μ_X be the smallest positive probability of $P_X(\cdot)$.

Suppose we have two random variables X and Y with respective discrete and finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ and joint distribution $P_{XY}(\cdot)$. The set of δ-typical sequences with respect to X and Y is defined as

$$T_\delta^n(X,Y) = \{(\underline{x},\underline{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\pi_{\underline{x},\underline{y}}(a,b) - P_{XY}(a,b)\right| \le \delta \cdot P_{XY}(a,b) \text{ for all } (a,b) \in \mathcal{X} \times \mathcal{Y}\} \quad (57)$$

where $\pi_{\underline{x},\underline{y}}(a,b)$ is the empirical frequency defined in the obvious way. We define the support set of X and Y to be

$$\mathcal{S}_{X,Y} = \{(a,b) \in \mathcal{X} \times \mathcal{Y} : P_{XY}(a,b) > 0\}.$$

Let $\mu_{X,Y}$ be the smallest positive probability of $P_{XY}(\cdot)$.

Recall from Section II-D that the set of conditionally δ-typical sequences with respect to X, Y, and $\underline{x}$ is

$$T_\delta^n(Y|\underline{x}) = \{\underline{y} \in \mathcal{Y} : (\underline{x},\underline{y}) \in T_\delta^n(X,Y)\}. \quad (58)$$

The following properties of typical sequences and sets are proved in the [9, Appendix].

Lemma 2: Let $0 < \delta \le 1$ and

$$\epsilon_\delta(n) = 2|\mathcal{S}_X| e^{-\delta^2 \cdot \mu_X \cdot n/3}. \quad (59)$$

We have

$$(1 - \epsilon_\delta(n)) \cdot 2^{n(1-\delta)H(X)} \le |T_\delta^n(X)| \le 2^{n(1+\delta)H(X)}. \quad (60)$$

Lemma 3: Let $0 < \delta_1 < \delta_2 \le 1$ and

$$\epsilon_{\delta_1,\delta_2}(n) = 2|\mathcal{S}_{X,Y}| e^{-\frac{(\delta_2-\delta_1)^2}{1+\delta_1} \cdot \mu_{X,Y} \cdot n/3}. \quad (61)$$

We have

$$(1 - \epsilon_{\delta_1,\delta_2}(n)) \cdot 2^{n(1-\delta_2)H(Y|X)} \le |T_{\delta_2}^n(Y|\underline{x})| \quad (62)$$

$$|T_{\delta_2}^n(Y|\underline{x})| \le 2^{n(1+\delta_2)H(Y|X)} \quad (63)$$

where (62) holds for every $\underline{x} \in T_{\delta_1}^n(X)$ and (63) holds for every $\underline{x} \in \mathcal{X}^n$.

Note that $\epsilon_\delta(n)$ and $\epsilon_{\delta_1,\delta_2}(n)$ approach zero exponentially with n. We need the following technical result in Section III-A.

Lemma 4: Suppose we have $Y = f(X)$, $\underline{x} \in T_\delta^n(X)$ and

$$\underline{y} = (f(x_1), f(x_2), \ldots, f(x_n)).$$

We then have $(\underline{x},\underline{y}) \in T_\delta^n(X,Y)$ and $\underline{y} \in T_\delta^n(Y)$.

Proof: We have $|\pi_{\underline{x}}(a) - P_X(a)| \le \delta \cdot P_X(a)$ for all $a \in \mathcal{X}$. Consider a pair $(a,b) \in \mathcal{X} \times \mathcal{Y}$. We clearly have $\pi_{\underline{x},\underline{y}}(a,b) = \pi_{\underline{x}}(a)$, $P_{XY}(a,b) = P_X(a)$, and therefore, $(\underline{x},\underline{y}) \in T_\delta^n(X,Y)$. This further implies that $\underline{y} \in T_\delta^n(Y)$ by [9, Lemma 21]. □

ACKNOWLEDGMENT

The authors wish to thank T. Ho and R. Koetter for helpful discussions and suggestions. The authors also wish to thank the Guest Editor N. Cai and the anonymous reviewers for useful remarks. N. Ratnakar wishes to thank Bell Labs, Lucent Technologies, for their support during a summer internship in 2004.

REFERENCES

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

[2] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.

[3] M. R. Aref, "Information Flow in Relay Networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1980.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[5] S. P. Borade, "Network information flow: Limits and achievability," in *Proc. 2002 IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 139.

[6] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Mathematics and Statistical Probability*, vol. 1, 1961, pp. 611–644.

[7] G. Kramer and S. A. Savari, "On networks of two-way channels," in *Proc. DIMACS Workshop on Algebraic Coding Theory and Information Theory*, vol. 68, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, A. Ashikhmin and A. Barg, Eds., Rutgers University, Piscataway, NJ, Dec. 2003, pp. 133–143.

[8] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[9] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[10] T. S. Han and M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 641–650, Sep. 1987.

[11] S. Jaggi, P. Sanders, P. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. M. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1973–1982, Jun. 2005.

[12] T. Ho, M. Médard, J. Shi, M. Effros, and D. Karger, "On randomized network coding," in *Proc. 41st Annu. Allerton Conf. Communication, Control, CompUTING*, Monticello, IL, Oct. 2003, pp. 11–20.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theory for Discrete Memoryless Systems*. Budapest, Hungary: Akadémiai Kiadó, 1981.

[14] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Pered. Inform.*, vol. 16, no. 1, pp. 24–34, Jan./Mar. 1980.

[15] T. S. Han, "The capacity region for the deterministic broadcast channel with a common message," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 122–125, Jan. 1980.

[16] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 374–377, May 1978.

[17] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[18] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[19] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[20] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[21] R. Gowaikar, A. F. Dana, R. Palanki, B. Hassibi, and M. Effros, "On the capacity of wireless erasure networks," in *Proc. 2004 IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 401.

[22] D. S. Lun, M. Médard, and M. Effros, "On coding for reliable communication over packet networks," in *Proc. 42nd Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 20–29.