Capacity Results for the Discrete Memoryless Network

Gerhard Kramer, *Member, IEEE*

***Abstract*—A discrete memoryless network (DMN) is a memoryless multiterminal channel with discrete inputs and outputs. A sequence of inner bounds to the DMN capacity region is derived by using code trees. Capacity expressions are given for three classes of DMNs: 1) a single-letter expression for a class with a common output, 2) a two-letter expression for a binary-symmetric broadcast channel (BC) with partial feedback, and 3) a finite-letter expression for push-to-talk DMNs. The first result is a consequence of a new capacity outer bound for common output DMNs. The third result demonstrates that the common practice of using a time-sharing random variable does not include all time-sharing possibilities, namely, time sharing of channels. Several techniques for improving the bounds are developed: 1) causally conditioned entropy and directed information simplify the inner bounds, 2) code trellises serve as simple code trees, 3) superposition coding and binning with code trees improves rates. Numerical computations show that the last technique enlarges the best known rate regions for a multiple-access channel (MAC) and a BC, both with feedback. In addition to the rate bounds, a sequence of inner bounds to the DMN reliability function is derived. A numerical example for a two-way channel illustrates the behavior of the error exponents.**

***Index Terms*—Capacity, causality, feedback, multiuser channels, random coding.**

I. Introduction

SHANNON created the area of network information theory by introducing the two-way channel. In a sense, Shannon solved the two-way channel problem by giving a sequence $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \ldots$ of inner bound regions that becomes the capacity region in the limit [1, Secs. 1 and 15]. However, the sequence is often considered to have little value because the boundary of its Lth term can usually not be computed—see, e.g., the discussion in [2, p. 259]. One is usually satisfied only with a *single-letter* capacity expression, i.e., one that includes only those channel input and output random variables involved in *one* use of the channel, plus perhaps a few auxiliary random variables.

Shannon describes the shortcomings of his *limiting expression* by calling its evaluation "impractical" [1, Sec. 16]. However, his *sequence of inner bounds* can be useful. For instance, we find examples where $\mathcal{R}_2$ has capacity points that are not in $\mathcal{R}_1$ (see Sections V-F and VI-B). The sequence further gives a general approach for improving communication systems. Codes designed to attain rate points in $\mathcal{R}_2$ will be better than codes designed for $\mathcal{R}_1$, and they will give hints on the structure of the best codes.

Such reasons motivate following Shannon's example. Several authors have done just that, e.g., for the relay channel [3], multiple-access channels (MACs) without memory [4], [5] and with memory [6]–[8], the interference channel [9], [10], and the broadcast channel (BC) [11] (see also [12] and [13]). Rather than dealing with these cases separately, we will treat the most general memoryless channel directly. We call this channel and its associated system of random variables the *Discrete Memoryless Network* (DMN). The DMN seems to have been considered first in [14, Sec. 1.6] (see also [12, Sec. X]). Special cases of this model are discussed in [2, Ch. 3] and [15, Sec. 14.10]. Some network models for source coding are described in [64] and [65]. However, we do not consider multiterminal source coding.

The DMN subsumes a wide variety of network models including, e.g., networks of discrete memoryless channels (DMCs), MACs, BCs, relay channels, and so forth. As in [1] we derive a sequence of inner bounds to the capacity region. Much of the derivation is a straightforward extension of [1] and [16, Ch. 5] but there are subtle issues involving feedback, broadcasting, and interference that require changes. Furthermore, along the way we present several new concepts and results such as causal conditioning, code trellises, a new capacity outer bound, and new capacity regions.

This paper is organized as follows. We begin by introducing the concept of causal conditioning in Section II. Section III discusses the DMN model and code trees, and gives the DMN capacity in terms of a limiting process. Section IV shows how to simplify the capacity expression and gives a new capacity outer bound for common-output DMNs. Section V discusses examples and presents new single-letter and finite-letter capacity regions. Section VI describes how to adapt superposition coding [17] and binning [18] techniques to include feedback. This section contains numerical examples showing that code trees enlarge some of the best known rate regions. Section VII gives proofs and a numerical example showing the behavior of error exponents for a two-way channel. Finally, Section VIII concludes the paper.

II. Preliminaries

A. Notation

Throughout the paper, random variables are written with upper case letters and values they take on with the corre-

Manuscript received April 3, 2000; revised August 18, 2002. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Massachusetts Institute of Technology, Cambridge, Aug. 1998; the IEEE Information Theory and Communications Workshop, Kruger National Park, South Africa, June 1999; and the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002.

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

Communicated by P. Narayan, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2002.806135

sponding lower case letters. Probability distributions are often denoted by only P when the arguments of P specify the distribution, e.g., $P(x|y) = P_{X|Y}(x|y)$. We consider only discrete and finite random variables.

Subscripts on a symbol are used to denote the symbol's source and/or to denote the symbol's position in a sequence. For example, X_{22} could mean "the output sequence of the *22nd* encoder," "the *22nd* random variable in the sequence $X_1, X_2, X_3, \ldots$," or the "the *2nd* output of the *2nd* encoder." The context will make clear which of these interpretations is in use. We also sometimes add commas to separate subscripts, e.g., $X_{2,2}$ for the last case above. Superscripts denote finite-length sequences of symbols, e.g., $x^N = x_1, x_2, \ldots, x_N$.

A DMN can have many terminals and we will need to manipulate the inputs and outputs of several of them simultaneously. To do this, we choose a similar notation to [15, Ch. 14] and denote sets of random variables with subscripts in brackets. For instance, if $\mathcal{S} = \{1, 3\}$ then $X_{(\mathcal{S})}$ denotes X_1, X_3; $U_{(\mathcal{S})}^{B(\mathcal{S})}$ denotes $U_1^{B_1}, U_3^{B_3}$; and $\boldsymbol{A}_{(\mathcal{S})}^L$ denotes $\boldsymbol{A}_1^L, \boldsymbol{A}_3^L$. Set subscripts without brackets often denote sums, e.g., $R_{\mathcal{S}} = \sum_{s \in \mathcal{S}} R_s$. The cardinality of a set $\mathcal{X}$ is denoted by $|\mathcal{X}|$.

The notation of [16, Ch. 2] for entropy and mutual information is used. All logarithms are to the base 2 so that our units are bits.

B. Functional Dependence Graphs (FDGs) and d-Separation

The random variables of most DMNs are related to each other in a complicated manner. We use graphs to ease the understanding of these relationships and to prove conditional independence results.

A graphical technique for establishing conditional independence in so-called Bayesian networks was introduced in [19]. These results were generalized to other types of graphs by various authors (see, e.g., [20], [21]) and we wish to consider *functional dependence graphs* (FDGs). Suppose we have N random variables that are defined by M *independent* random variables by N functions. An FDG is a directed graph (a set of vertices and a set of ordered pairs of these vertices called edges) having $M+N$ vertices representing the random variables, and in which edges are drawn from one vertex to another if the random variable of the former vertex is an argument of the function defining the random variable of the latter vertex. For example, Fig. 1 depicts the FDG for the first three uses of a channel with feedback. In this graph, the channel input symbol X_n is a function of the message U^B and the past channel outputs Y^{n-1}. We have drawn the feedback links using dashed lines to emphasize the role that feedback plays. The output Y_n is a function of X_n and the *noise* random variable Z_n. The graph has $N = 6$ random variables defined by $M = 4$ independent random variables. The M vertices representing the independent U^B, Z_1, Z_2, and Z_3 are distinguished by drawing them with a hollow circle.

We are interested in establishing conditional independence results by using FDGs. For example, for the above channel one can show that $I(U^B; Y_2|X^2) = 0$ by using the functional *equations*. Alternatively, a *graphical* criterion called d*-separation* proves the same result. By d-separation we mean the following.

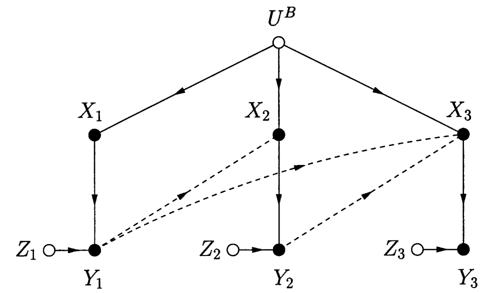

Fig. 1. The FDG for the first three uses of a memoryless channel with feedback.

Definition 1: Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be disjoint subsets of the vertices of an FDG $\boldsymbol{G}$. $\mathcal{Z}$ is said to d*-separate* $\mathcal{X}$ from $\mathcal{Y}$ if there is no path between a vertex in $\mathcal{X}$ and a vertex in $\mathcal{Y}$ after the following manipulations of the graph have been performed.

1) Consider the subgraph $\boldsymbol{G}_{\mathcal{X}\mathcal{Y}\mathcal{Z}}$ of $\boldsymbol{G}$ consisting of the vertices in $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, as well as the edges and vertices encountered when moving *backward* one or more edges starting from any of the vertices in $\mathcal{X}$ or $\mathcal{Y}$ or $\mathcal{Z}$.

2) In $\boldsymbol{G}_{\mathcal{X}\mathcal{Y}\mathcal{Z}}$ delete all edges coming *out* of the vertices in $\mathcal{Z}$. Call the resulting graph $\boldsymbol{G}_{\mathcal{X}\mathcal{Y}|\mathcal{Z}}$.

3) Remove the arrows on the remaining edges of $\boldsymbol{G}_{\mathcal{X}\mathcal{Y}|\mathcal{Z}}$ to obtain the undirected graph $\boldsymbol{G}_{\mathcal{X}\mathcal{Y}|\mathcal{Z}}^u$.

The above is a reformulation of a definition in [22, p. 117], and in Appendix A we prove that the definitions are equivalent. The motivation for the reformulation is that it clearly distinguishes between the independence due to causality (step 1) and due to conditioning (step 2). A fundamental result of [22, Sec. 3.3] is that d-separation establishes conditional independence in FDGs having no directed cycles. That is, if $\mathcal{Z}$ d-separates $\mathcal{X}$ from $\mathcal{Y}$ in $\boldsymbol{G}$ and we collect the random variables of the vertices in $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ in the respective vectors $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ then $I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{Z}) = 0$.

C. Causal Conditioning and Directed Information

Coding for the DMN is restricted by *causality*, i.e., the transmitting terminals cannot use their nth-channel outputs to code until time $n+1$ and later. We introduce the concept of *causal conditioning* that captures the essential aspects of such coding. Our approach is an extension of Marko's [23] and Massey's [24]. Several properties of the defined quantities are developed in Appendix B.

Definition 2: The probability distribution of the sequence X^N *causally conditioned* on the sequence Y^N is

$$P\left(x^N \| y^N\right) := \prod_{n=1}^{N} P\left(x_n \left| x^{n-1} y^n\right.\right). \tag{1}$$

This definition differs from $P(x^N|y^N)$ only in that y^n replaces y^N in each term on the right-hand side of (1). It differs from Marko's $p(x|x_n y_n)$ only in that y_n is included in the conditioning [23, Sec. IV]. The name "causal" refers to the conditioning on *past and present* values of Y^N only.

Definition 3: The entropy of the sequence X^N *causally conditioned* on the sequence Y^N is

$$H\left(X^N \| Y^N\right) := \sum_{n=1}^{N} H\left(X_n \left| X^{n-1} Y^n\right.\right). \tag{2}$$

This definition differs from $H(X^N|Y^N)$ only in that Y^n replaces Y^N in each term on the right-hand side of (2). $H(X^N\|Y^N)$ is similar to Marko's *free information* [23, eq. (8)].

The *directed information* flowing from a sequence X^N to a sequence Y^N was introduced by Massey [24] and can be written as

$$\begin{aligned} I\left(X^N \rightarrow Y^N\right) &= H\left(Y^N\right) - H\left(Y^N \| X^N\right) \\ &= \sum_{n=1}^{N} I\left(X^n; Y_n \left| Y^{n-1}\right.\right). \end{aligned} \tag{3}$$

Directed information is similar to Marko's *directed transinformation* [23, eqs. (15) and (16)]. Note that whereas

$$I(X^N; Y^N) = I(Y^N; X^N)$$

in general we have

$$I(X^N \rightarrow Y^N) \neq I(Y^N \rightarrow X^N).$$

Definition 4: The directed information flowing from X^N to Y^N when *causally conditioned* on the sequence Z^N is

$$\begin{aligned} I\left(X^N \rightarrow Y^N \| Z^N\right) &:= H\left(Y^N \| Z^N\right) - H\left(Y^N \| X^N Z^N\right) \\ &= \sum_{n=1}^{N} I\left(X^n; Y_n \left| Y^{n-1} Z^n\right.\right). \end{aligned} \tag{4}$$

This definition naturally extends directed information to include causal conditioning.

For most of the paper we will be interested in information *rates*. We use the following notation for per-letter entropies and informations:

$$\begin{aligned} H_N(X) &= H\left(X^N\right)/N \\ I_N(X; Y) &= I\left(X^N; Y^N\right)/N \\ I_N(X; Y|Z) &= I\left(X^N; Y^N \left| Z^N\right.\right)/N. \end{aligned} \tag{5}$$

Similarly, for causally conditioned entropy and informations we write

$$\begin{aligned} H_N(X\|Y) &= H\left(X^N \| Y^N\right)/N \\ I_N(X \rightarrow Y) &= I\left(X^N \rightarrow Y^N\right)/N \\ I_N(X \rightarrow Y\|Z) &= I\left(X^N \rightarrow Y^N \| Z^N\right)/N. \end{aligned} \tag{6}$$

In some cases of interest the terms in (5) and (6) have limits as N tends to infinity. We denote these limits by $H_\infty(X)$, $I_\infty(X; Y)$, and so on. The first of these limits, $H_\infty(X)$, is the usual information rate or *entropy rate* of the source producing the sequence $X_1, X_2, X_3, \ldots$. If the source is *stationary*, then we have (see [15, p. 64])

$$H_\infty(X) = \lim_{n\rightarrow\infty} H\left(X_n \left| X^{n-1}\right.\right). \tag{7}$$

Similar results hold for the other limits of (5) and (6), and are developed in Appendix B.

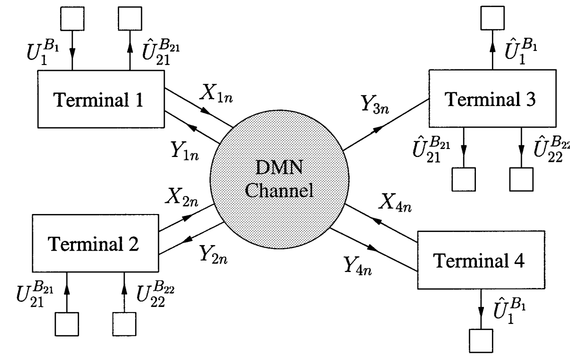

Fig. 2. A four-terminal DMN.

III. Model and Capacity

A. Model

We consider the setup of [14, Sec. 1.6] with a small change to the DMN channel (defined below). This setup is somewhat more general than the model treated in [15, Sec. 14.10] because there are more messages at each terminal. An example of a DMN is shown in Fig. 2, and we use this example to clarify our notation.

A T-terminal DMN is defined by four types of random variables: messages $U_s^{B_s}$, channel inputs X_{tn}, channel outputs Y_{tn}, and message estimates $\hat{U}_s^{B_s}$. We impose four restrictions on the random variables. First, we require that the $U_s^{B_s}$ are *statistically independent*. Without loss of essential generality, we assume that $U_s^{B_s}$ is a string of B_s bits whose rate is $R_s = H(U_s^{B_s})/N$ bits per use, where N is the number of times we use the DMN channel.

Next, suppose that terminal t transmits M_t messages $U_{t1}^{B_{t1}}, \ldots, U_{tM_t}^{B_{tM_t}}$. In general, M_t ranges from 0 to $2^{T-1} - 1$ since terminal t might want to send a different message to each nonempty subset of the other $T-1$ terminals. We can simplify notation by writing that terminal t transmits $U_{(\mathcal{E}_t)}^{B(\mathcal{E}_t)}$, where $\mathcal{E}_t = \{t1, \ldots, tM_t\}$ is terminal t's *encoding index set*. For example, in Fig. 2 we have $\mathcal{E}_1 = \{1\}$ and $\mathcal{E}_2 = \{21, 22\}$. As we have done here, we will continue to write U_t rather than U_{t1} when $M_t = 1$. We assume that every message is encoded and decoded as a whole, so we often drop the superscripts and write that terminal t encodes $U_{(\mathcal{E}_t)}$. Our second restriction on the DMN random variables is that at time n, terminal t encodes by using only its messages and the available feedback Y_t^{n-1}, i.e.,

$$X_{tn} = \boldsymbol{a}_{tn}\left(U_{(\mathcal{E}_t)}, Y_t^{n-1}\right) \tag{8}$$

for some functions $\boldsymbol{a}_{tn}$, $n = 1, 2, \ldots, N$. The random variables X_{tn} and Y_{tn} take on values in the respective finite alphabets $\mathcal{X}_t$ and $\mathcal{Y}_t$.

For the third restriction, we define the *DMN channel* by the input alphabets $\mathcal{X}_t$, the output alphabets $\mathcal{Y}_t$, $t = 1, 2, \ldots, T$,

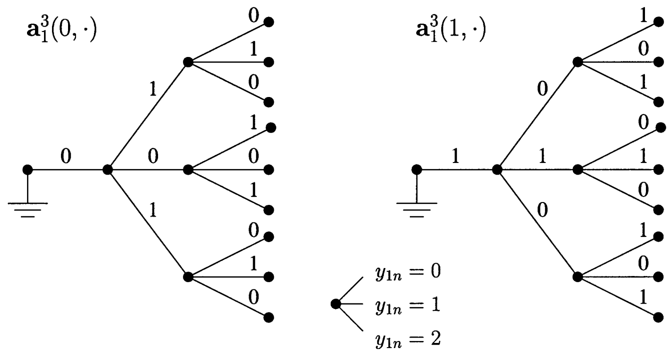

Fig. 3. A code with two code trees.

and the conditional probability distribution $P(y_1, y_2, \ldots, y_T | x_1, x_2, \ldots, x_T)$. We use the DMN channel N times so that we must consider properties of the joint distribution $P(y_1^N, \ldots, y_T^N | x_1^N, \ldots, x_T^N)$. By a *memoryless* channel we mean that the following is true for all n (see also [24]):

$$P\left(y_{1n}, \ldots, y_{Tn} \left| x_1^n, \ldots, x_T^n, y_1^{n-1}, \ldots, y_T^{n-1}\right.\right) = P_{Y_1 \cdots Y_T | X_1 \cdots X_T}\left(y_{1n}, \ldots, y_{Tn} \left| x_{1n}, \ldots, x_{Tn}\right.\right). \quad (9)$$

This condition is a little different than that given in [14, eq. (1.6.1)]. Observe that a DMN channel is the same as a DMC with vector-valued inputs and outputs.

For the fourth restriction, let the *decoding index set* $\mathcal{D}_t$ be the set of subscripts of the messages to be decoded at terminal t. For example, in Fig. 2 we have $\mathcal{D}_1 = \{21\}$, $\mathcal{D}_3 = \{1, 21, 22\}$, and $\mathcal{D}_4 = \{1\}$. We require that

$$\hat{U}_{(\mathcal{D}_t)} = d_t\left(U_{(\mathcal{E}_t)}, Y_t^N\right) \quad (10)$$

for some functions d_t, $t = 1, 2, \ldots, T$. In summary, the four restrictions on the DMN random variables are: the messages are statistically independent, property (8), property (9), and property (10). *These restrictions define the problem under consideration.*

B. Code Trees

We borrow terminology from [25, p. 349] and interpret the functions $\boldsymbol{a}_t^N = \boldsymbol{a}_{t1}, \ldots, \boldsymbol{a}_{tN}$ in (8) as a *code tree* (also called a *strategy* [1]). By a code tree we mean a rooted tree of depth N having one branch leaving the root vertex and $|\mathcal{Y}_t|$ branches leaving the other vertices, and whose branches are labeled with a symbol from $\mathcal{X}_t$. Each message combination $u_{(\mathcal{E}_t)}$ is assigned one code tree $\boldsymbol{a}_t^N(u_{(\mathcal{E}_t)}, \cdot)$ and each path through the tree corresponds to one of the channel output sequences y_t^{N-1}. If there is no feedback we may write $\boldsymbol{a}_t^N(u_{(\mathcal{E}_t)}, \cdot)$ simply as $x_t^N(u_{(\mathcal{E}_t)})$.

A DMN *code* is a list of code trees. For example, a code with two code trees is shown in Fig. 3. The encoder of Terminal 1 maps $U_1 = 0$ into the code tree $\boldsymbol{a}_1^3(0, \cdot)$ and $U_1 = 1$ into $\boldsymbol{a}_1^3(1, \cdot)$. Thus, $\boldsymbol{A}_1^3$ is a random code tree that is a function of U_1. To explain how the code works, suppose that $U_1 = 0$ so that $\boldsymbol{a}_1^3(0, \cdot)$ is chosen. Terminal 1 then sends $x_{11} = 0$, and if $y_{11} = 2$ it would next send $x_{12} = 1$, and so on. We will sometimes denote code trees in the manner

$$\boldsymbol{a}_1^3 = [0,\ 101,\ 010101010]. \quad (11)$$

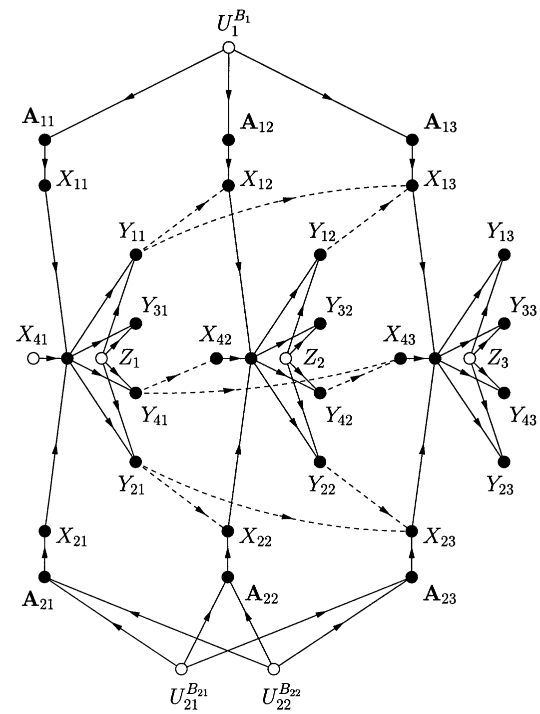

Fig. 4. The FDG for three uses of the DMN of Fig. 2. To simplify the graph, all channel input symbols at time n go through a common unlabeled vertex.

The meaning is that $\boldsymbol{a}_{11} = 0$ is the symbol on the branch leaving the root of the tree $\boldsymbol{a}_1^3$, that $\boldsymbol{a}_{12} = [1, 0, 1]$ is the ordered list of symbols leaving the vertex at depth 1, and that $\boldsymbol{a}_{13} = [0, 1, 0, 1, 0, 1, 0, 1, 0]$ is the ordered list of symbols leaving the vertices at depth 2. The DMN code trees are selected *before* the channel is used and they are known by all terminals.

The FDG of the DMN of Fig. 2 is depicted in Fig. 4, and it demonstrates the complexity of the system of random variables under consideration. The code trees $\boldsymbol{A}_t^N$ and code tree lists $\boldsymbol{A}_{tn}$ are random variables that are functions of the random variables $U_{(\mathcal{E}_t)}$. The random variables $Z_n, n = 1, \ldots, N$, represent noise, i.e.,

$$Y_{tn} = f_t\left(X_{1n}, \ldots, X_{Tn}, Z_n\right) \quad (12)$$

for some functions f_t, $t = 1, \ldots, T$ (see [26, Sec. 11]).

C. Merging Functions

Terminal 2 in Fig. 2 broadcasts U_{21} to terminals 1 and 3 and U_{22} to Terminal 3 *simultaneously*. We treat such situations by using *merging functions* (or merging channels) as described in [11, Sec. II-C] (see also [27]). By a merging function we mean the function g_{tn} obtained by rewriting (8) as

$$X_{tn} = g_{tn}\left(\boldsymbol{a}_{t1n}\left(U_{t1}, Y_t^{n-1}\right), \ldots, \boldsymbol{a}_{tM_tn}\left(U_{tM_t}, Y_t^{n-1}\right)\right) \quad (13)$$

where the $\boldsymbol{a}_{tmn}$ are appropriate code trees. Terminal 2 now has the form given in Fig. 5, where the alphabets $\mathcal{X}_{21}$ and $\mathcal{X}_{22}$ of X_{21} and X_{22} are chosen to satisfy (13). Note that this type of "decomposition" of the encoders is always possible because the alphabets $\mathcal{X}_{21}$ and $\mathcal{X}_{22}$ can be *chosen freely*. For instance, we can choose $X_{21n} = [U_{21}, Y_2^{n-1}]$ and $X_{22n} = U_{22}$. The effect

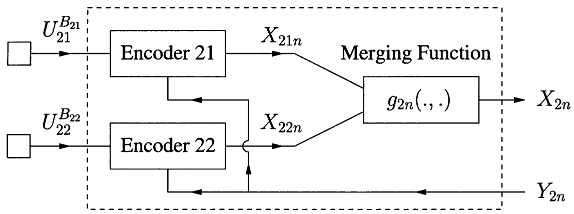

Fig. 5. The encoder of terminal 2 with merging functions.

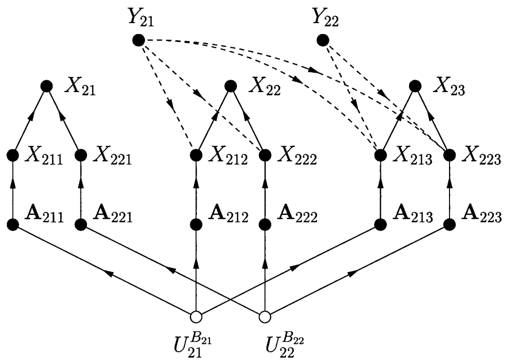

Fig. 6. The FDG for terminal 2's encoder with merging functions.

of merging functions on the FDG is depicted in Fig. 6, where $\boldsymbol{A}_{2n}$ has been split into $\boldsymbol{A}_{21n}$ and $\boldsymbol{A}_{22n}$, and new vertices have been introduced to represent the merging function inputs X_{21n} and X_{22n}. The code trees $\boldsymbol{A}_{21}^N$ and $\boldsymbol{A}_{22}^N$ have the same tree structure as the code tree $\boldsymbol{A}_2^N$ but might have different alphabets for the branch labels.

The point of introducing merging functions is essentially to simplify notation. We have transformed the DMN into a new network where every terminal has at most *one* message to transmit. This network will be memoryless but time varying because the g_{tn} can be chosen freely and will normally change with n. The time-varying nature of the g_{tn} does not play an important role, however, because they are known ahead of time to all terminals in the network.

We emphasize that terminal t's encoding functions remain unchanged by choosing appropriate code trees $\boldsymbol{A}_{tm}^N$ and merging functions $g_{t1}, g_{t2}, \ldots, g_{tN}$. Moreover, the code trees will be independent because they are functions of independent messages. This observation is important for deriving our *outer* bound to the DMN capacity region. For the *inner* bound, the independence of the $\boldsymbol{A}_{tm}^N$ motivates random coding with independent code trees. This restriction is not necessary, though, and will be dropped when we consider more sophisticated coding techniques in Section VI-B.

D. Capacity

Let $\hat{U}_{t(\mathcal{D}_t)}^{B(\mathcal{D}_t)}$ be terminal t's estimate of $U_{(\mathcal{D}_t)}^{B(\mathcal{D}_t)}$. The average *bit* error probability of terminal t is

$$P_t = \frac{1}{B_{\mathcal{D}_t}} \sum_{d \in \mathcal{D}_t} \sum_{b=1}^{B_d} \Pr\left(\hat{U}_{t,\,db} \neq U_{db}\right) \tag{14}$$

where $B_{\mathcal{D}_t} = \sum_{d \in \mathcal{D}_t} B_d$. The *capacity region* $\mathcal{C}_{\mathrm{DMN}}$ is the set of rate-tuples $(R_{(\mathcal{E}_1)}, R_{(\mathcal{E}_2)}, \ldots, R_{(\mathcal{E}_T)})$ which one can approach with arbitrarily small positive P_t for all t. Alternatively, $\mathcal{C}_{\mathrm{DMN}}$ is the set of rate points which one can approach while making the following sum of *block* error probabilities small for all t:

$$P_{B,\,t} = \sum_{\mathcal{S} \subseteq \mathcal{D}_t,\, \mathcal{S} \neq \emptyset} \Pr\left(\hat{U}_{t(\mathcal{S})}^{B_{t(\mathcal{S})}} \neq U_{(\mathcal{S})}^{B(\mathcal{S})},\, \hat{U}_{t(\mathcal{S}^C)}^{B_{t(\mathcal{S}^C)}} = U_{(\mathcal{S}^C)}^{B(\mathcal{S}^C)}\right) \tag{15}$$

where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{D}_t$ and $\emptyset$ is the empty set. Note that (15) is a sum of probabilities of mutually exclusive events.

The equivalence of the *bit* error and *block* error capacity regions can be demonstrated as in [16, p. 119]. If a block error occurs at terminal t then at least one bit error has occurred there, but not more than $B_{\mathcal{D}_t}$ bit errors, so that $P_t \leq P_{B,\,t} \leq B_{\mathcal{D}_t} \cdot P_t$. This implies that if the block error probability is small so is the bit error probability, and hence the block error capacity region is inside the bit error capacity region. It, thus, suffices to find an outer bound to the bit error capacity region and an inner bound to the block error capacity region that are the same, and this we proceed to do.

We prove the following two theorems in Section VII. Both theorems have $2^{|\mathcal{D}_t|} - 1$ bounds for each t, and each bound corresponds to one of the terms in the sum of (15).

Theorem 1 (Inner Bounds to $\mathcal{C}_{\mathrm{DMN}}$): The convex hull $\mathcal{R}_L$ of the rate-tuples $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$ satisfying

$$0 \leq \sum_{s \in \mathcal{S}} R_s \leq I_L\left(\boldsymbol{A}_{(\mathcal{S})}; Y_t \left| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\mathcal{S}^C)}\right.\right) \tag{16}$$

for all $t = 1, \ldots, T$ and all $\mathcal{S} \subseteq \mathcal{D}_t$ is contained within $\mathcal{C}_{\mathrm{DMN}}$, where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{D}_t$ and the $\boldsymbol{A}_s^L$ are statistically independent.

For example, the two-way channel is a $T = 2$ terminal channel with $\mathcal{D}_1 = \{2\}$ and $\mathcal{D}_2 = \{1\}$ so that $\mathcal{R}_L$ is the convex hull of the rate regions

$$\begin{aligned} 0 \leq R_1 &\leq I_L(\boldsymbol{A}_1; Y_2|\boldsymbol{A}_2) \\ 0 \leq R_2 &\leq I_L(\boldsymbol{A}_2; Y_1|\boldsymbol{A}_1) \end{aligned} \tag{17}$$

where $\boldsymbol{A}_1^L$ and $\boldsymbol{A}_2^L$ are independent. We emphasize that one *cannot* in general replace the code trees $\boldsymbol{A}_t^L$ with the channel input sequences X_t^L, as might be expected. The bounds (17) are an explicit characterization of a rate region described by Shannon in [1, Sec. 15].

Theorem 2 (Outer Bound to $\mathcal{C}_{\mathrm{DMN}}$): $\mathcal{C}_{\mathrm{DMN}}$ is contained within the closure $\mathcal{C}_{\mathrm{DMN}}^{\mathrm{OUT}}$ of the set of rate points $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$ satisfying

$$0 \leq \sum_{s \in \mathcal{S}} R_s \leq I_L\left(\boldsymbol{A}_{(\mathcal{S})}; Y_t \left| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\mathcal{S}^C)}\right.\right) \tag{18}$$

where $\mathcal{S} \subseteq \mathcal{D}_t$, $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{D}_t$, L is any positive integer, the $\boldsymbol{A}_s^L$ are statistically independent, and $t = 1, 2, \ldots, T$.

We denote by $\lim_{L\to\infty} \mathcal{R}_L$ the set of limit points of the convergent sequences whose Lth term is in $\mathcal{R}_L$. The following theorem gives the capacity region.

Theorem 3: $\mathcal{C}_{\mathrm{DMN}} = \lim_{L\to\infty} \mathcal{R}_L$.

Proof: The region $\mathcal{C}_{\mathrm{DMN}}^{\mathrm{OUT}}$, being the *closure* of the set $\{\mathcal{R}_L\colon L \text{ finite}\}$, is (see [28, p. 743])

$$\mathcal{C}_{\mathrm{DMN}}^{\mathrm{OUT}} = \{\mathcal{R}_L\colon L \text{ finite}\} \cup \lim_{L\to\infty} \mathcal{R}_L. \tag{19}$$

From Theorem 1, all points in $\{\mathcal{R}_L\colon L \text{ finite}\}$ are in $\mathcal{C}_{\mathrm{DMN}}$, and so are all points in $\lim_{L\to\infty} \mathcal{R}_L$ by definition. This establishes that $\mathcal{C}_{\mathrm{DMN}} = \mathcal{C}_{\mathrm{DMN}}^{\mathrm{OUT}}$.

It remains to show that $\{\mathcal{R}_L\colon L \text{ finite}\} \subseteq \lim_{L\to\infty} \mathcal{R}_L$, i.e., that for any point $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$ in $\mathcal{R}_L$ one can generate a sequence in $\{\mathcal{R}_L\colon L \text{ finite}\}$ whose limit is $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$. A trick used in [1, Sec. 15] is to write $\ell = mL + j$ for some integers $m \geq 0$ and $0 \leq j < L$, and then to randomize the code by repeating m times the distributions for $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$, and by sending no information in the last j uses of the channel. This means that $mL/(mL+j) \cdot (R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$ lies in $\mathcal{R}_\ell$. Moreover, the limit of these points as $m \to \infty$ is $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$. □

IV. Discussion and Simplifications

A. Time Sharing

Theorem 1 uses a *convex hull* operation to allow for time sharing. A better approach is to introduce a time-sharing random variable V, replace $I_L(\boldsymbol{A}_{(\mathcal{S})}; Y_t|\boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\mathcal{S}^C)})$ with $I_L(\boldsymbol{A}_{(\mathcal{S})}; Y_t|\boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\mathcal{S}^C)}V)$ in (16), and make the $\boldsymbol{A}_s^L$ conditionally independent given V (see, e.g., [15, p. 397]). The choice of technique is not critical here because one may include the time sharing in the code trees. However, for a fixed L, the latter approach can give larger regions [2, p. 289], [29]. A time-sharing random variable is also useful for improving error exponents, as discussed at the end of Section VII-B. At the same time, we show in Section V-F that a time-sharing random variable does not capture all time-sharing possibilities.

B. Max-Flow, Min-Cut Outer Bound

The outer bound of Theorem 2 is usually not computable. A simpler, but generally loose, outer bound was derived in [15, Sec. 14.10] and also applies to DMNs.

Proposition 1: $\mathcal{C}_{\mathrm{DMN}}$ is contained within the set of rate points $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$ satisfying

$$0 \leq \sum_{t\in\mathcal{S}} \sum_{s\in\mathcal{E}_t} R_s \leq I\left(X_{(\mathcal{S})}; Y_{(\mathcal{S}^C)}|X_{(\mathcal{S}^C)}\right) \tag{20}$$

for some joint distribution $P(x^T)$, where $\mathcal{S} \subseteq \{1, 2, \ldots, T\}$ and $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\{1, 2, \ldots, T\}$.

This bound can be proved by dividing the network terminals into two sets $\mathcal{S}$ and $\mathcal{S}^C$, letting the terminals in each set cooperate with each other to get a two-way channel between the sets, and applying Shannon's two-way channel outer bound [1, Sec. 9].

C. Codes With Code Trellises

The number of branches of a code tree grows exponentially with L. A natural way to curb this growth is to use a special type of code tree called a code *trellis*. By a code trellis we mean a code tree that collapses into a bounded number of states for all depths. For example, the code tree on the left in Fig. 3 could be extended by using the code trellis of Fig. 7 whose states correspond to the past received symbols $y_{1(n-1)}$. A potential advantage of code trellises is that they will be easier to decode than general code trees.

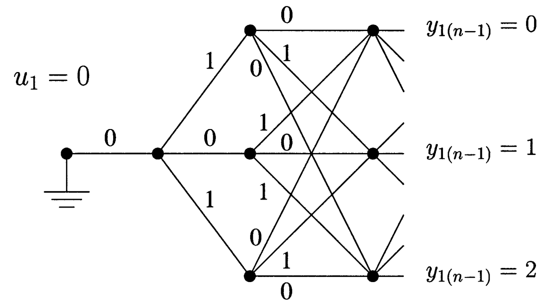

Fig. 7. A code trellis for the code tree on the left in Fig. 3.

D. No Feedback

We use the approach of [24] and say that terminal t is *used without feedback* if, for all m and n

$$P\left(x_{tmn}\left|x_{tm}^{n-1}y_t^{n-1}\right.\right) = P\left(x_{tmn}\left|x_{tm}^{n-1}\right.\right). \tag{21}$$

Without feedback, the code trees collapse to codewords, i.e., one can replace $\boldsymbol{A}_{tm}^L$ by X_{tm}^L.

E. Rate Regions in Terms of Directed Information

We are interested in expressing (16) with as few random variables as possible. We have

$$\begin{aligned} I\left(\boldsymbol{A}_{(\mathcal{S})}^N; Y_t^N\left|\boldsymbol{A}_{(\mathcal{E}_t)}^N\boldsymbol{A}_{(\mathcal{S}^C)}^N\right.\right) &= \sum_{n=1}^N H\left(Y_{tn}\left|Y_t^{n-1}\boldsymbol{A}_{(\mathcal{E}_t)}^n\boldsymbol{A}_{(\mathcal{S}^C)}^n\right.\right) \\ &\quad - H\left(Y_{tn}\left|Y_t^{n-1}\boldsymbol{A}_{(\mathcal{E}_t)}^n\boldsymbol{A}_{(\mathcal{D}_t)}^n\right.\right) \\ &= \sum_{n=1}^N H\left(Y_{tn}\left|Y_t^{n-1}X_t^n\boldsymbol{A}_{(\mathcal{S}^C)}^n\right.\right) \\ &\quad - H\left(Y_{tn}\left|Y_t^{n-1}X_t^n\boldsymbol{A}_{(\mathcal{D}_t)}^n\right.\right) \\ &= I\left(\boldsymbol{A}_{(\mathcal{S})}^N \to Y_t^N\left\|X_t^N\boldsymbol{A}_{(\mathcal{S}^C)}^N\right.\right). \end{aligned} \tag{22}$$

The first equality follows by the causal nature of the encoding. The second equality follows because X_t^n is a function of $\boldsymbol{A}_{(\mathcal{E}_t)}^n$ and Y_t^{n-1}, and because $[X_t^n, Y_t^{n-1}, \boldsymbol{A}_{(\mathcal{S}^C)}^n]$ and $[X_t^n, Y_t^{n-1}, \boldsymbol{A}_{(\mathcal{D}_t)}^n]$ d-separate $\boldsymbol{A}_{(\mathcal{E}_t)}^n$ from Y_{tn}. The third equality follows by definition.

Note that some code trees $\boldsymbol{A}_{(\mathcal{E}_t)}^L$ can be replaced by the channel input sequences X_t^L when using causally conditioned directed information. For example, for the two-way channel, $\mathcal{R}_L$ becomes the convex hull of the rate regions

$$\begin{aligned} 0 &\leq R_1 \leq I_L(\boldsymbol{A}_1 \to Y_2\|X_2) \\ 0 &\leq R_2 \leq I_L(\boldsymbol{A}_2 \to Y_1\|X_1) \end{aligned} \tag{23}$$

where $\boldsymbol{A}_1^L$ and $\boldsymbol{A}_2^L$ are independent. Moreover, for some channels, one can further replace the remaining code trees $\boldsymbol{A}_{tm}^L$ by the channel input sequences X_{tm}^L. In such cases, the random coding becomes simpler, as shown next.

F. Common Outputs

A simplification in the capacity expression is possible when a transmitting terminal t and a receiving terminal r have the same channel output Y_r. This means that terminal r knows X_{tm}^n given $\boldsymbol{A}_{tm}^n$ and Y_r^{n-1}, and one can replace the code trees $\boldsymbol{A}_{tm}^L$ by the codewords X_{tm}^L in the bounds corresponding to the errors terminal r can make when estimating the messages of terminal t.

Another simplification is possible if terminal t sees the outputs of *all* terminals to which it is sending messages. Terminal t can then restrict attention to random coding distributions that are based only on its channel input and output symbols. For example, for the common output two-way channel, $\mathcal{R}_L$ becomes the convex hull of the rate regions

$$\begin{aligned} &0 \le R_1 \le I_L(X_1 \to Y \| X_2) \\ &0 \le R_2 \le I_L(X_2 \to Y \| X_1) \end{aligned} \tag{24}$$

where $P(x_{1\ell}, x_{2\ell}|x_1^{\ell-1}, x_2^{\ell-1}, y^{\ell-1})$ factors as

$$P\left(x_{1\ell}\left|x_1^{\ell-1}, y^{\ell-1}\right.\right) \cdot P\left(x_{2\ell}\left|x_2^{\ell-1}, y^{\ell-1}\right.\right) \tag{25}$$

for $\ell = 1, 2, \ldots, L$. The factorization (25) follows from d-separation and implies that one can label the ℓth branches of the code trees *independently* for each vertex at depth ℓ by using $P(x_{t\ell}|x_t^{\ell-1}, y^{\ell-1})$ rather than having to label them *jointly* via $P(\boldsymbol{a}_{t\ell}|\boldsymbol{a}_t^{\ell-1})$.

Finally, we give the following outer bound on the capacity of a common-output DMN. This bound uses the *dependence balance* ideas of [30] and will be shown to give capacity for a class of DMNs in Section V-E. For the following theorem we write $I(X_1; X_2; \ldots; X_T|V)$ for $-H(X^T|V) + \sum_{t=1}^T H(X_t|V)$.

Theorem 4: $\mathcal{C}_{\rm DMN}$ of a common-output DMN is contained within the set of rate-tuples $(R_{(\mathcal{E}_1)}, \ldots, R_{(\mathcal{E}_T)})$ satisfying

$$0 \le \sum_{s \in \mathcal{S}} R_s \le I\left(X_{(\mathcal{S})}; Y \left| X_t X_{(\mathcal{S}^C)} X_{(\mathcal{D}_t^C)} V\right.\right) \tag{26}$$

for all $t = 1, \ldots, T$ and all $\mathcal{S} \subseteq \mathcal{D}_t$, where Y is the common output, $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{D}_t$, $\mathcal{D}_t^C$ is the complement of $\mathcal{D}_t$ in $\bigcup_{t'} \mathcal{E}_{t'}$, and the X_t satisfy

$$I(X_1; X_2; \ldots; X_T|V) \le I(X_1; X_2; \ldots; X_T|YV) \tag{27}$$

for some auxiliary random variable V that takes on at most $1 + \sum_t (2^{|\mathcal{D}_t|} - 1)$ values.

This theorem is proved in Appendix C. One can further extend many of the other results in [30] to common output DMNs. For example, the parallel channel extensions of [30, Secs. 5 and 6] will improve on Theorem 4.

V. Examples

We give several examples to show how the theory applies to specific DMNs. We begin by reviewing known results.

A. The DMC

The DMC is a two-terminal DMN for which terminal 1 transmits the message U_1 to terminal 2, possibly with noisy feedback Y_1 [26]. It is, of course, known that $\mathcal{R}_1$ is the capacity region [31], [32]. Proposition 1 serves as an outer bound.

B. The Two-Way Channel

The two-way channel was introduced in [1] and $\mathcal{R}_L$ is given by (23). Another capacity inner bound was developed in [33] but this region can be improved by using directed information rates [34, Ch. 4].

C. The MAC

The MAC is a T-terminal DMN for which the messages U_t, $t = 1, \ldots, T-1$, are transmitted to terminal T [1, Sec. 17]. If there is no feedback then $\mathcal{C}_{\rm DMN}$ is known to be $\mathcal{R}_1$ [9], [10], [12], [35] (see also the limiting characterization of [4]). Consider next the case $T = 3$ with full feedback, i.e., $Y = Y_1 = Y_2 = Y_3$. $\mathcal{R}_L$ is then the convex hull of the regions

$$\begin{aligned} &0 \le R_1 \le I_L(X_1 \to Y \| X_2) \\ &0 \le R_2 \le I_L(X_2 \to Y \| X_1) \\ &0 \le R_1 + R_2 \le I_L(X_1 X_2 \to Y). \end{aligned} \tag{28}$$

The discussion in Section IV-F shows that we may restrict our attention to those random coding distributions for which $P(x_{1\ell}, x_{2\ell}|x_1^{\ell-1}, x_2^{\ell-1}, y^{\ell-1})$ factors as in (25). A particular MAC of this type is considered in more detail in Section VI-A.

D. The BC

The BC [17] is a T-terminal DMN for which terminal 1 transmits up to $2^{T-1} - 1$ messages U_{1m}, $m = 1, 2, \ldots, 2^{T-1} - 1$ to the other $T - 1$ terminals. We will write that U_{1m} is transmitted to terminal t if the binary representation of m contains a 1 in the $(t-1)$th position. For instance, since $m = 6$ is 110 in binary digits, the message U_{16} is transmitted to terminals 3 and 4 (see [27] for a different choice of notation).

Consider the case $T = 3$ for which terminal 1 transmits U_{11} to terminal 2, U_{12} to terminal 3, and U_{13} to both terminals 2 and 3. $\mathcal{R}_L$ is then the convex hull of the (R_{11}, R_{12}, R_{13}) satisfying

$$\begin{aligned} &\text{terminal 2:} \\ &0 \le R_{11} \le I_L(\boldsymbol{A}_{11} \to Y_2 \| \boldsymbol{A}_{13}) \\ &0 \le R_{13} \le I_L(\boldsymbol{A}_{13} \to Y_2 \| \boldsymbol{A}_{11}) \\ &0 \le R_{11} + R_{13} \le I_L(\boldsymbol{A}_{11}\boldsymbol{A}_{13} \to Y_2) \\ &\text{terminal 3:} \\ &0 \le R_{12} \le I_L(\boldsymbol{A}_{12} \to Y_3 \| \boldsymbol{A}_{13}) \\ &0 \le R_{13} \le I_L(\boldsymbol{A}_{13} \to Y_3 \| \boldsymbol{A}_{12}) \\ &0 \le R_{12} + R_{13} \le I_L(\boldsymbol{A}_{12}\boldsymbol{A}_{13} \to Y_3) \end{aligned} \tag{29}$$

where $\boldsymbol{A}_{11}^L$, $\boldsymbol{A}_{12}^L$, and $\boldsymbol{A}_{13}^L$ are independent. Without feedback, we can replace the $\boldsymbol{A}_{tm}^L$ with X_{tm}^L. The resulting characterization of $\mathcal{C}_{\rm DMN}$ is the same as that in [11].

One is often interested in the case $R_{13} = 0$ for which our coding technique makes $\boldsymbol{A}_{13}^L$ a constant. The region (29) thus simplifies to

$$\begin{aligned} 0 &\le R_{11} \le I_L(\boldsymbol{A}_{11} \to Y_2) \\ 0 &\le R_{12} \le I_L(\boldsymbol{A}_{12} \to Y_3). \end{aligned} \tag{30}$$

However, as shown in [36], the region (29) includes a potentially larger region if $L = 1$. It is straightforward to generalize the derivation of [36, Appendix A] to take into account code trees. The result is that the following region is a capacity inner bound:

$$\begin{aligned} 0 &\le R_{11} \le I_L(\boldsymbol{A}_{11}\boldsymbol{A}_{13} \to Y_2) \\ 0 &\le R_{12} \le I_L(\boldsymbol{A}_{12}\boldsymbol{A}_{13} \to Y_3) \\ 0 &\le R_{11} + R_{12} \le \min[I_L(\boldsymbol{A}_{13} \to Y_2),\, I_L(\boldsymbol{A}_{13} \to Y_3)] \\ &\quad + I_L(\boldsymbol{A}_{11} \to Y_2 \| \boldsymbol{A}_{13}) + I_L(\boldsymbol{A}_{12} \to Y_3 \| \boldsymbol{A}_{13}) \end{aligned} \tag{31}$$

where $\boldsymbol{A}_{11}^L$, $\boldsymbol{A}_{12}^L$, and $\boldsymbol{A}_{13}^L$ are independent. Observe that setting $L = 1$ in (31) gives the capacity of the degraded BC [17].

E. DMNs With a Special Common Output

Consider the class of common-output DMNs having the special property that for every $y \in \mathcal{Y}$ there are $T - 1$ terminals whose channel inputs X_t are known (cf. the example in [30, p. 47]). The right-hand side of (27) is thus zero so that in Theorem 4 the X_t are independent given V.

Suppose additionally that each terminal either has no message destined for it or decodes all messages, i.e., $\mathcal{D}_t = \emptyset$ or $\mathcal{D}_t \cup \mathcal{E}_t = \bigcup_{t'} \mathcal{E}_{t'}$. The reason for this restriction is so that either X_{tn} is known to all terminals ($\mathcal{D}_t = \emptyset$) or Theorem 1 gives MAC-like inner bounds ($\mathcal{D}_t \cup \mathcal{E}_t = \bigcup_{t'} \mathcal{E}_{t'}$).

For DMNs satisfying these two constraints all points inside the region of Theorem 4 are in $\mathcal{R}_1$ if we use a time-sharing random variable (see Section IV-A). Consider, e.g., the following common-output three-way channel:

$$Y = \begin{cases} (X_1,\, X_2,\, X_3), & \text{if } X_1 \neq 0,\ X_2 \neq 0 \\ (0,\, 1,\, X_3), & \text{if } X_1 = 0,\ X_2 \neq 0 \\ (1,\, 0,\, X_3), & \text{if } X_1 \neq 0,\ X_2 = 0 \\ (0,\, 0,\, 0), & \text{if } X_1 = X_2 = 0 \end{cases} \tag{32}$$

where $\mathcal{X}_1 = \mathcal{X}_2 = \{0,\, 1,\, 2\}$ and $\mathcal{X}_3 = \{0,\, 1\}$. It is easy to check that every $y \in \mathcal{Y}$ determines at least two of X_1, X_2, or X_3. Suppose that each terminal sends a common message to both other terminals, i.e., $\mathcal{D}_t \cup \mathcal{E}_t = \bigcup_{t'} \mathcal{E}_{t'}$ for all t. The capacity region is thus $\mathcal{R}_1$ with the addition of a time-sharing random variable V. However, it turns out that V is not needed here. We choose $P_{X_1}(1) = P_{X_1}(2)$, $P_{X_2}(1) = P_{X_2}(2)$, and $P_{X_3}(0) = P_{X_3}(1) = 1/2$. Let $P_{X_1}(0) = p$ and $P_{X_2}(0) = q$. We find that $\mathcal{C}_{\text{DMN}}$ is the set of $(R_1,\, R_2,\, R_3)$ satisfying

$$\begin{aligned} 0 &\le R_1 \le H(Y|X_2X_3) = h(p) + (1-p)(1-q) \\ 0 &\le R_2 \le H(Y|X_1X_3) = h(q) + (1-p)(1-q) \\ 0 &\le R_3 \le H(Y|X_1X_2) = 1 - pq \end{aligned} \tag{33}$$

where $0 \le p \le 1$, $0 \le q \le 1$, and

$$h(x) = -x \log_2(x) - (1-x)\log_2(1-x)$$

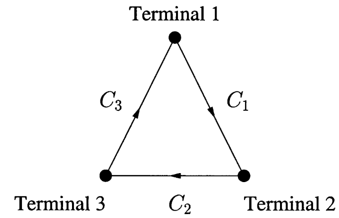

Fig. 8. A network of three DMCs.

is the binary entropy function. The best sum rate is $R_1 + R_2 + R_3 = 3.6890$ bits per use and is achieved by $p = q = 0.2263$.

F. Push-to-Talk DMC Networks

A *network of D DMCs* is a T-terminal DMN channel with D point-to-point links. For example, Fig. 8 shows a network of three DMCs. The DMCs are assumed to be independent in the sense that (9) factors as

$$P\left(y^T \,|x^T\right) = \prod_{d=1}^{D} P\left(\tilde{y}_d \,|\tilde{x}_d\right) \tag{34}$$

where the collection $\{\tilde{X}_1,\, \ldots,\, \tilde{X}_D\}$ (and $\{\tilde{Y}_1,\, \ldots,\, \tilde{Y}_D\}$) can be partitioned into T subsets such that each X_t (and Y_t) is a vector whose entries are the elements of one of the subsets. As in Fig. 8, a graph of such a network has T vertices representing the terminals and D directed edges representing the DMCs $P(\tilde{y}_d|\tilde{x}_d)$, where an edge goes from the vertex t to the vertex r if $\tilde{X}_d$ and $\tilde{Y}_d$ are entries of the respective X_t and Y_r.

We modify the problem slightly by making the DMCs *dependent* as follows: to each DMC we add an input symbol x_d^* such that the other DMCs can transmit information only if this symbol is used. Formally, (9) factors into a product of distributions $P(y_d|x^D)$ such that $P(y_d|x^D) = 1/|\mathcal{Y}_d|$ unless $X_c = x_c^*$ for all $c \neq d$ in which case $P(y_d|x^D) = P(y_d|x_d)$. We call such a DMN channel a *push-to-talk DMC network* because of its relation to Shannon's push-to-talk two-way channel [1, Sec. 1].

The best coding for these networks obviously involves time sharing. However, it turns out that neither the convex-hull operation nor the addition of a time-sharing random variable V makes $\mathcal{R}_1$ the capacity. The reason is that, although the DMN is memoryless, there are *delays* because the terminals can be separated from each other by more than one DMC. At the same time, one might expect that there *should* be a single-letter capacity expression for push-to-talk networks. In fact, the region of Proposition 1 is the capacity.

Consider, for example, the network of DMCs with

$$P\left(y^T \,|x^T\right) = P\left(y_1 \,|x_T\right) \cdot \prod_{t=2}^{T} P\left(y_t \,|x_{t-1}\right) \tag{35}$$

and suppose that the capacity of the DMC with input X_t is C_t. Such a network is shown in Fig. 8 for $T = 3$. We now add the symbols x_t^* described previously to get a push-to-talk DMC network. These networks are similar to a token ring [37, p. 320] because only one link can send at a time; however, there is no random accessing going on here. We will consider the case where each terminal is trying to send a common message to all other terminals.

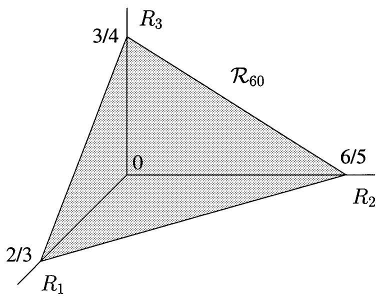

Fig. 9. The capacity region $\mathcal{R}_{60}$ of the push-to-talk counterpart of the DMN in Fig. 8 with $C_1 = 1$, $C_2 = 2$, and $C_3 = 3$.

Suppose first that $C_t = C$ for all t. The regions $\mathcal{R}_L$ can be computed to be the convex hull of the $T+1$ extreme points $(0, \ldots, 0)$, $(R, 0, \ldots, 0)$, $(0, R, \ldots, 0)$, …, $(0, 0, \ldots, R)$ where

$$R = C \cdot \left\lfloor \frac{L}{T-1} \right\rfloor \frac{1}{L} \tag{36}$$

and $\lfloor x \rfloor$ is the smallest integer less than or equal to x. We thus have $\mathcal{R}_{T-1} = \mathcal{C}_{\text{DMN}}$ but

$$\mathcal{R}_L = (0, \ldots, 0) \subset \mathcal{C}_{\text{DMN}}, \qquad \text{for } L < T-1.$$

For example, the capacity of the push-to-talk version of the network in Fig. 8 with $C_1 = C_2 = C_3$ is $\mathcal{R}_2$. The outer bound of Proposition 1 is also $\mathcal{R}_2$.

Consider next different capacities C_t with $C_t > 0$. Let $\mathcal{S}_t$ be the set $\{1, \ldots, T\}$ excluding $t-1$ if $t \neq 1$ and excluding T if $t = 1$. We define

$$N_t = \prod_{s \in \mathcal{S}_t} C_s \quad \text{and} \quad D_t = \sum_{r \in \mathcal{S}_t} \prod_{\substack{s \in \mathcal{S}_t \\ s \neq r}} C_s.$$

For example, the network of Fig. 8 has $N_1 = C_1 C_2$ and $D_1 = C_1 + C_2$. One can show that the capacity is the convex hull of the $T+1$ extreme points $(0, \ldots, 0)$, $(R_1, 0, \ldots, 0)$, $(0, R_2, \ldots, 0)$, …, $(0, 0, \ldots, R_T)$ where

$$R_t = N_t / D_t = \left[\sum_{s \in \mathcal{S}_t} C_s^{-1} \right]^{-1}. \tag{37}$$

In general, one needs $L \to \infty$ to get this result. However, if the C_t are integers then L need be at most the least common multiple of the D_t. For example, suppose Fig. 8 has $C_1 = 1$, $C_2 = 2$, and $C_3 = 3$. The maximal individual rates are then $R_1 = 2/3$, $R_2 = 6/5$, and $R_3 = 3/4$ which are all achieved if $L = 60$. The capacity $\mathcal{R}_{60}$ of this DMN is depicted in Fig. 9.

VI. Superposition Coding and Binning

The tools we used to derive the capacity results were code trees (to exploit feedback), merging functions (for broadcasting), and an appropriate maximum-likelihood (ML) decoding rule (56) (to deal with interference). A natural next step is to add *superposition coding* [17] and *binning* [18], [38]. Superposition coding with two codes can be interpreted as follows: the codewords of the first (coarse) code serve as "cloud centers" while the codewords of the second (fine) code are the "cloud" that refines the first code. Binning, on the other hand, is a technique invented for source coding and involves assigning source sequences to bins. The number of bins is smaller than the number of sequences, and one achieves compression by sending the indexes of the bins rather than those of the sequences.

The motivation for using superposition coding and binning is that these sophisticated techniques often give larger regions for the same L than Theorem 1. Furthermore, the generalization of Theorem 1 is often straightforward: one takes an existing region, replaces the codewords by code trees, adds a subscript L to the information symbol I, and the resulting region is a capacity inner bound. We demonstrate the above procedure for two DMNs: a MAC with full feedback and a BC with feedback from one terminal. Improvements over existing rate regions are found for both networks, and for the second network we even find new capacity points.

A. The Two-Transmitter MAC With Full Feedback

Cover and Leung [39] derived a capacity inner bound $\mathcal{R}^{\text{CL}}$ for the two-transmitter MAC with full feedback. A natural generalization of their region is the following lemma whose proof is basically the same as the proof in [39] with codewords replaced by code trees. An alternative proof using ML decoding is given in [34, p. 93].

Lemma 1: The set R_L^{CL} of rate pairs (R_1, R_2) satisfying

$$\begin{aligned} 0 &\le R_1 \le \frac{1}{L} I\left(\boldsymbol{A}_1^L; Y^L \middle| \boldsymbol{A}_2^L V\right) \\ 0 &\le R_2 \le \frac{1}{L} I\left(\boldsymbol{A}_2^L; Y^L \middle| \boldsymbol{A}_1^L V\right) \\ 0 &\le R_1 + R_2 \le \frac{1}{L} I\left(\boldsymbol{A}_1^L \boldsymbol{A}_2^L; Y^L\right) \end{aligned} \tag{38}$$

is contained within the capacity region of the MAC with full feedback, where V is a discrete random variable and

$$P\left(\boldsymbol{a}_1^L, \boldsymbol{a}_2^L, y^L \middle| v\right) = P\left(\boldsymbol{a}_1^L \middle| v\right) \cdot P\left(\boldsymbol{a}_2^L \middle| v\right) \cdot P\left(y^L \middle| \boldsymbol{a}_1^L, \boldsymbol{a}_2^L\right). \tag{39}$$

We emphasize that one *cannot* write

$$I(\boldsymbol{A}_1^L; Y^L | \boldsymbol{A}_2^L V) = I(X_1^L; Y^L | X_2^L V)$$

as might be expected. Observe that $\mathcal{R}^{\text{CL}} = \mathcal{R}_1^{\text{CL}}$ and that $\lim_{L \to \infty} \mathcal{R}_L^{\text{CL}}$ is the capacity region because $\mathcal{R}_L^{\text{CL}} \supseteq \mathcal{R}_L$. For $L = 1$, one can limit the cardinality of V to

$$\min(|\mathcal{X}_1||\mathcal{X}_2| + 1, \, |\mathcal{Y}| + 2)$$

[40], [41]. However, for large L, there does not seem to be a point in finding cardinality limits because the rate regions are computable only for very symmetric distributions anyway. As explained in Section IV-F, one can simplify the informations in this lemma because the three terminals have a common output. A further simplification results by replacing V with V^L and by choosing the coding distributions as shown later in (41).

Corollary 1: The set of rate pairs (R_1, R_2) satisfying

$$\begin{aligned} 0 &\le R_1 \le I_L(X_1 \to Y \| X_2 V) \\ 0 &\le R_2 \le I_L(X_2 \to Y \| X_1 V) \end{aligned}$$

$$0 \le R_1 + R_2 \le I_L(X_1X_2 \to Y) \tag{40}$$

is contained within $\mathcal{R}_L^{\mathrm{CL}}$, where V^L is a sequence of discrete random variables such that

$$P(x_{1\ell},\, x_{2\ell},\, y_\ell | v^L,\, x_1^{\ell-1},\, x_2^{\ell-1},\, y^{\ell-1})$$

factors as

$$P\left(x_{1\ell}\left|v^\ell,\, x_1^{\ell-1},\, y^{\ell-1}\right.\right) \cdot P\left(x_{2\ell}\left|v^\ell,\, x_2^{\ell-1},\, y^{\ell-1}\right.\right) \cdot P\left(y_\ell\left|x_{1\ell}x_{2\ell}\right.\right) \tag{41}$$

for $\ell = 1, 2, \ldots, L$.

The region of this corollary is strictly larger than $\mathcal{R}^{\mathrm{CL}}$ for several channels [34, Ch. 4]. For example, consider the MAC with $Y = X_1 + X_2 + Z$ where X_1, X_2, and Z are binary (0 and 1) random variables and $P_Z(0) = P_Z(1) = 1/2$. Theorem 4 gives an equal-rate upper bound of $R_1 = R_2 = 0.45915$ bit per use. In Appendix D, we show that the point with $R_1 = R_2 = 0.43621$ bit per use lies on the boundary of $\mathcal{R}^{\mathrm{CL}}$, while $R_1 = R_2 = 0.43879$ bit per use is achievable with Corollary 1. Although the improvement is small, this is the first example of a discrete MAC with full feedback for which $\mathcal{R}^{\mathrm{CL}}$ could be shown to be strictly inside $\mathcal{C}_{\mathrm{DMN}}$. Improvements over $\mathcal{R}^{\mathrm{CL}}$ had previously been shown only for the power-constrained Gaussian MAC with full feedback [42].

B. The Two-Receiver BC With Partial Feedback

Consider the BC with two receivers and $R_{13} = 0$ (see Section V-D). For simplicity, we drop certain indexes and write that the transmitter sends X and the receivers see Y_1 and Y_2. The best known single-letter rate region for this DMN is that of Marton [43], and this region can be derived by using binning [44]. A natural generalization of her region $\mathcal{R}^{\mathrm{Mar}}$ is given in the following lemma. Note that this lemma allows the transmitter to see noisy feedback $\tilde{Y}$.

Lemma 2: The set $\mathcal{R}_L^{\mathrm{Mar}}$ of rate pairs $(R_1,\, R_2)$ satisfying

$$\begin{aligned}
&0 \le R_1 \le I_L(\boldsymbol{A}_1\boldsymbol{A}_3;\, Y_1)\\
&0 \le R_2 \le I_L(\boldsymbol{A}_2\boldsymbol{A}_3;\, Y_2)\\
&0 \le R_1 + R_2 \le \min[I_L(\boldsymbol{A}_3;\, Y_1),\, I_L(\boldsymbol{A}_3;\, Y_2)]\\
&\qquad + I_L(\boldsymbol{A}_1;\, Y_1|\boldsymbol{A}_3) + I_L(\boldsymbol{A}_2;\, Y_2|\boldsymbol{A}_3) - I_L(\boldsymbol{A}_1;\, \boldsymbol{A}_2|\boldsymbol{A}_3)
\end{aligned} \tag{42}$$

is contained within the capacity region of the BC with feedback, where $P(x^L,\, y_1^L,\, y_2^L,\, \tilde{y}^L | \boldsymbol{a}_1^L,\, \boldsymbol{a}_2^L,\, \boldsymbol{a}_3^L)$ factors as

$$\prod_{\ell=1}^{L} P\left(x_\ell\left|\boldsymbol{a}_1^\ell,\, \boldsymbol{a}_2^\ell,\, \boldsymbol{a}_3^\ell,\, \tilde{y}^{\ell-1}\right.\right) \cdot P\left(y_{1\ell},\, y_{2\ell},\, \tilde{y}_\ell | x_\ell\right). \tag{43}$$

Note that $\mathcal{R}_1^{\mathrm{Mar}} = \mathcal{R}^{\mathrm{Mar}}$ and that $\lim_{L\to\infty} \mathcal{R}_L^{\mathrm{Mar}}$ is the capacity region. Also, if the $\boldsymbol{A}_1^L$, $\boldsymbol{A}_2^L$, and $\boldsymbol{A}_3^L$ are independent then (42) collapses to (31).

Lemma 2 can be proved by observing that nothing prevents the auxiliary random variables U, V, and W in [43] from being the respective $\boldsymbol{A}_1^L$, $\boldsymbol{A}_2^L$, and $\boldsymbol{A}_3^L$. Once the length N code trees (consisting of $\tilde{N}$ concatenated code trees of length L) have been chosen, one channel input sequence x^N is generated for every possible $(\boldsymbol{a}_1^N,\, \boldsymbol{a}_2^N,\, \boldsymbol{a}_3^N,\, \tilde{y}^{N-1})$ by using the distribution

$$\prod_{\ell=1}^{L} P\left(x_\ell\left|\boldsymbol{a}_1^\ell,\, \boldsymbol{a}_2^\ell,\, \boldsymbol{a}_3^\ell,\, \tilde{y}^{\ell-1}\right.\right) \tag{44}$$

for each of the $\tilde{N}$ subblocks of length L. An alternative code construction could proceed along the lines of [45].

Consider, e.g., the binary-symmetric BC (BSBC). This channel has a binary (0 and 1) input X and binary outputs $Y_t = X \oplus Z_t$, $t = 1, 2$, where Z_1 and Z_2 are independent binary random variables taking on the value 1 with probability ϵ_1 and ϵ_2, respectively. Suppose that U_1 and U_2 are destined for receivers 1 and 2, respectively. $\mathcal{C}_{\mathrm{DMN}}$ is then given by (31) with $L = 1$, or by $\mathcal{R}^{\mathrm{Mar}}$ [15, p. 457] (see also [46]–[48]). If $\epsilon_1 \le \epsilon_2$, the boundary of $\mathcal{R}^{\mathrm{Mar}}$ can be approached by setting $X = X_1$, making X_2 a constant, making X_3 a coin-flipping random variable, and setting $\Pr(X = X_3) = q$. The resulting rate region is

$$\begin{aligned}
&0 \le R_2 \le I(X_3;\, Y_2) = 1 - h(q * \epsilon_2)\\
&0 \le R_1 + R_2 \le I(X;\, Y_1|X_3) + I(X_3;\, Y_2)\\
&\qquad = [h(q * \epsilon_1) - h(\epsilon_1)] + [1 - h(q * \epsilon_2)]
\end{aligned} \tag{45}$$

where $p_1 * p_2 = p_1(1 - p_2) + p_2(1 - p_1)$. The capacity region boundary is plotted for $\epsilon_1 = 0.15$ and $\epsilon_2 = 0.2$ in Fig. 10 as the curve labeled $L = 1$. Observe that $\mathcal{C}_{\mathrm{DMN}}$ is slightly larger than the time-sharing region (see also [15, p. 426]).

Suppose next that U_2 is destined for *both* receivers. It turns out that $\mathcal{C}_{\mathrm{DMN}}$ remains unchanged because the BSBC is degraded (see also [49]). Furthermore, an outer bound on the sum rate is clearly the capacity of the X to Y_1 channel, i.e., $R_1 + R_2 \le 1 - h(\epsilon_1)$.

Suppose now that U_2 is destined for both receivers *and* that there is full feedback from receiving terminal 1, i.e., $\tilde{Y} = Y_1$. We consider code trees of depth $L = 2$ and choose X^2 from the branch labels of $\boldsymbol{A}_1^2$. We further make $\boldsymbol{A}_2^2$ a constant and $\boldsymbol{A}_3^2$ a codeword. The bounds (42) become

$$\begin{aligned}
&0 \le R_2 \le I_2(X_3;\, Y_2)\\
&0 \le R_1 + R_2 \le I_2(\boldsymbol{A}_1;\, Y_1|X_3)\\
&\qquad + \min[I_2(X_3;\, Y_1),\, I_2(X_3;\, Y_2)].
\end{aligned} \tag{46}$$

We choose X_3^2 to be a binary codeword that takes on all four possible values with probability $1/4$, and set $\Pr(X_1 = X_{31}) = q$ and

$$\begin{aligned}
&\Pr(X_2 = X_{32}|X_1 = 0,\, Y_{11} = 0) = q_0\\
&\Pr(X_2 = X_{32}|X_1 = 0,\, Y_{11} = 1) = q_1\\
&\Pr(X_2 = X_{32}|X_1 = 1,\, Y_{11} = 0) = 1 - q_1\\
&\Pr(X_2 = X_{32}|X_1 = 1,\, Y_{11} = 1) = 1 - q_0.
\end{aligned} \tag{47}$$

Optimizing over q, q_0, and q_1 we find that the region labeled $L = 2$ in Fig. 10 is a capacity inner bound. Observe that feedback enlarges the capacity region. Moreover, for some distributions we have $I_2(X_3;\, Y_1) < I_2(X_3;\, Y_2)$ and $I_2(X_3\boldsymbol{A}_1;\, Y_1) = 1 - h(\epsilon_1)$ so that the $L = 2$ curve lies on the sum rate bound for an interval of R_1. *This means that two-letter coding gives capacity points that single-letter coding does not.* For example, by

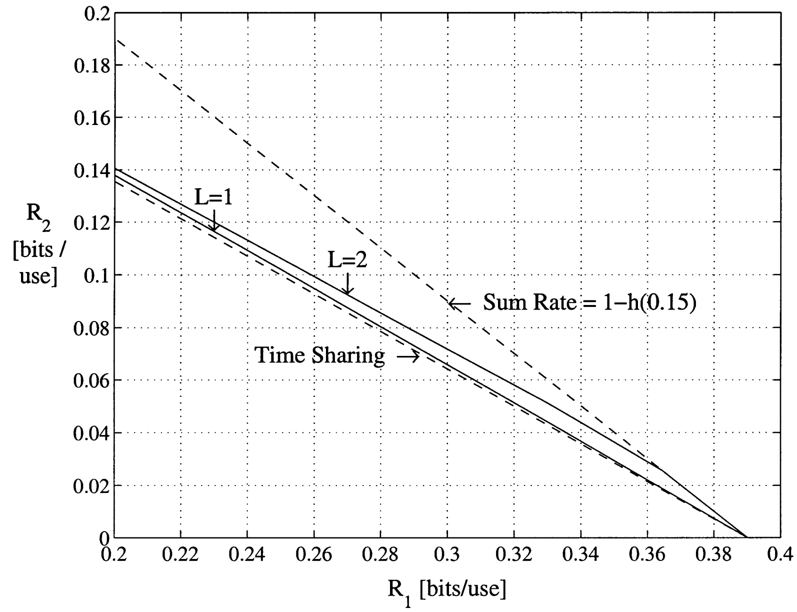

Fig. 10. Rate regions for a binary-symmetric BC with crossover probabilities $\epsilon_1 = 0.15$ and $\epsilon_2 = 0.2$. Note the x-axis range.

using $q = 0.647$, $q_0 = 0.696$, and $q_1 = 1$ the point $(R_1, R_2) = (0.3668, 0.0234)$ with sum rate $1 - h(0.15) = 0.3902$ is in $\mathcal{R}_2^{\mathrm{Mar}}$. In contrast, the point $(R_1, R_2) = (0.3668, 0.0172)$ lies on the boundary of $\mathcal{R}_1$. Enlargements over the no-feedback capacity region of other BCs had already been discovered in [50], [51].

VII. Proofs and Random Coding Exponents

A. A Capacity Outer Bound

Consider terminal t and define the average bit error probability $P_{t,\mathcal{S}}$ as (cf. (14))

$$P_{t,\mathcal{S}} = \frac{1}{B_{\mathcal{S}}} \sum_{s\in\mathcal{S}} \sum_{b=1}^{B_s} \Pr\left(\hat{U}_{t,sb} \neq U_{sb}\right) \tag{48}$$

where $B_{\mathcal{S}} = \sum_{s\in\mathcal{S}} B_s$ and $\hat{U}_{t,sb}$ is terminal t's estimate of U_{sb}. We can bound $P_{t,\mathcal{S}}$ by using a generalization of Fano's inequality to bit sequences [16, p. 79]

$$h(P_{t,\mathcal{S}}) \geq \frac{1}{B_{\mathcal{S}}} H\left(U_{(\mathcal{S})}^{B_{(\mathcal{S})}} \middle| \hat{U}_{t,(\mathcal{S})}^{B_{(\mathcal{S})}}\right). \tag{49}$$

We can further bound

$$h(P_{t,\mathcal{S}}) \geq \frac{1}{B_{\mathcal{S}}} H\left(U_{(\mathcal{S})}^{B_{(\mathcal{S})}} \middle| \hat{U}_{t,(\mathcal{S})}^{B_{(\mathcal{S})}} U_{(\mathcal{S}^C)}^{B_{(\mathcal{S}^C)}}\right) \tag{50}$$

where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{D}_t$. The reason for using (50) is that we want a bound that includes conditioning on the event that $U_{(\mathcal{S}^C)}^{B_{(\mathcal{S}^C)}}$ is decoded correctly.

From here on, we no longer consider individual message bits so we write U_s in place of $U_s^{B_s}$. We have

$$H\left(U_{(\mathcal{S})} \middle| \hat{U}_{t,(\mathcal{S})} U_{(\mathcal{S}^C)}\right) = H\left(U_{(\mathcal{S})}\right) - I\left(U_{(\mathcal{S})}; \hat{U}_{t,(\mathcal{S})} U_{(\mathcal{S}^C)}\right)$$

and note that $\hat{U}_{t,(\mathcal{S})} = d_{t,\mathcal{S}}(U_{(\mathcal{E}_t)}, Y_t^N)$ for some function $d_{t,\mathcal{S}}$. We thus have

$$\begin{aligned} h(P_{t,\mathcal{S}}) &\geq \frac{1}{B_{\mathcal{S}}} \left[H\left(U_{(\mathcal{S})}\right) - I\left(U_{(\mathcal{S})}; U_{(\mathcal{E}_t)} Y_t^N U_{(\mathcal{S}^C)}\right)\right] \\ &= \frac{1}{B_{\mathcal{S}}} \left[H\left(U_{(\mathcal{S})}\right) - I\left(\boldsymbol{A}_{(\mathcal{S})}^N; \boldsymbol{A}_{(\mathcal{E}_t)}^N Y_t^N \boldsymbol{A}_{(\mathcal{S}^C)}^N\right)\right] \\ &= \frac{1}{B_{\mathcal{S}}} \left[H\left(U_{(\mathcal{S})}\right) - I\left(\boldsymbol{A}_{(\mathcal{S})}^N; Y_t^N \middle| \boldsymbol{A}_{(\mathcal{E}_t)}^N \boldsymbol{A}_{(\mathcal{S}^C)}^N\right)\right] \end{aligned} \tag{51}$$

where the first step follows by the data processing theorem [16, p. 80], and the second because the code trees are functions of the messages and because of d-separation in the FDG. We can rewrite (51) as

$$h(P_{t,\mathcal{S}}) \geq \frac{H\left(U_{(\mathcal{S})}\right)}{B_{\mathcal{S}}} \left[1 - \frac{I_N\left(\boldsymbol{A}_{(\mathcal{S})}; Y_t \middle| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\boldsymbol{\mathcal{S}}^C)}\right)}{R_{\mathcal{S}}}\right] \tag{52}$$

where $R_{\mathcal{S}} = \sum_{s\in\mathcal{S}} R_s$. We use (52) to prove the following proposition.

Proposition 2: Consider a DMN that is used N times, and let $\mathcal{E}_{\mathrm{all}} = \bigcup_t \mathcal{E}_t$ and $k = H(U_{(\mathcal{S})})/B_{\mathcal{S}} > 0$. If there is some $\epsilon > 0$ such that

$$R_{\mathcal{S}} \geq I_L\left(\boldsymbol{A}_{(\mathcal{S})}; Y_t \middle| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\boldsymbol{\mathcal{S}}^C)}\right) + \epsilon \tag{53}$$

for all $P_{\boldsymbol{A}_{(\mathcal{E}_{\mathrm{all}})}^L}$ and all L then the average bit error probability $P_{t,\mathcal{S}}$ satisfies

$$P_{t,\mathcal{S}} \geq h^{-1}\left(\frac{k\cdot\epsilon}{\epsilon + \log(|\mathcal{Y}_t|)}\right) \tag{54}$$

where $h^{-1}(\cdot)$ is the "inverse" binary entropy function taking on values between 0 and $1/2$.

Proof: The condition (53) for *all* L implies that

$$R_{\mathcal{S}} = I_N(\boldsymbol{A}_{(\mathcal{S})};\, Y_t|\boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}) + \delta$$

for some $\delta \geq \epsilon$. We now manipulate (52) as

$$\begin{aligned} h(P_{t,\mathcal{S}})/k &\geq 1 - \frac{1}{R_{\mathcal{S}}}\, I_N\left(\boldsymbol{A}_{(\mathcal{S})};\, Y_t \left| \boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}\right.\right) \\ &= 1 - \frac{I_N\left(\boldsymbol{A}_{(\mathcal{S})};\, Y_t \left| \boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}\right.\right)}{I_N\left(\boldsymbol{A}_{(\mathcal{S})};\, Y_t \left| \boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}\right.\right) + \delta} \\ &= \frac{\delta}{\delta + I_N\left(\boldsymbol{A}_{(\mathcal{S})};\, Y_t \left| \boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}\right.\right)}. \end{aligned} \tag{55}$$

Since $f(x) = x/(x+c)$ is increasing with x for $x > 0$ and $c \geq 0$, we have

$$\delta/(\delta + c) \geq \epsilon/(\epsilon + c)$$

for $c = I_N(\boldsymbol{A}_{(\mathcal{S})};\, Y_t|\boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\mathcal{S}^C)})$. Furthermore, we have

$$I_N(\boldsymbol{A}_{(\mathcal{S})};\, Y_t|\boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}) \leq H_N(Y_t)$$

and from a basic bound on entropy we have $H_N(Y_t) \leq \log(|\mathcal{Y}_t|)$ (see [15, p. 27]). Combining these results yields (54). □

Proposition 2 allows the number of channel uses N to be any positive integer, which lets us prove Theorem 2. For this theorem we need to add a *closure* operation because the definition of the capacity region $\mathcal{C}_{\mathrm{DMN}}$ includes those rate-tuples which one gets arbitrarily close to. For example, the information rates $I_N(\boldsymbol{A}_{(\mathcal{S})};\, Y_t|\boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)})$ may approach, but never reach, a limit as the block length N increases.

Proof of Theorem 2: Any rate point lying outside $\mathcal{C}_{\mathrm{DMN}}^{\mathrm{OUT}}$ must satisfy (53) for some t, some $\mathcal{S} \subseteq \mathcal{D}_t$ with $\mathcal{S} \neq \emptyset$, some $\epsilon > 0$, and all L. For such a rate point, Proposition 2 guarantees that no codes can achieve an error probability $P_{t,\mathcal{S}}$ below that specified by (54), which is positive for $k > 0$ and $\epsilon > 0$. Thus, this rate point is not in $\mathcal{C}_{\mathrm{DMN}}$. □

Note that $\mathcal{C}_{\mathrm{DMN}}^{\mathrm{OUT}}$ is convex because allowing L to be any positive integer implicitly allows time sharing [1, Sec. 8].

B. Capacity Inner Bounds

The transmission of message U_s at rate $R_s = H(U_s)/N$ requires $\lceil 2^{NR_s} \rceil$ code trees. The code trees $\boldsymbol{a}_s^N$ are generated by concatenating $\tilde{N}$ randomly chosen code trees of length L, so that $N = L \cdot \tilde{N}$. We denote by $Q(\boldsymbol{a}_s^L)$ the probability of selecting $\boldsymbol{a}_s^L$ when *random coding* to distinguish it from the probability $P(\boldsymbol{a}_s^N)$ that the code tree $\boldsymbol{a}_s^N$ is chosen for transmission. The merging functions can be any functions with appropriate domains and ranges, and they do not play an important role in what follows.

We simplify notation by writing $\boldsymbol{a}_{u_{(\mathcal{S})}}^N$ in place of $\boldsymbol{a}_{(\mathcal{S})}^N(u_{(\mathcal{S})})$. We define $\mathcal{E}_{\mathrm{all}}$ as the set of indexes of all transmitted messages (see Proposition 2). We further denote by $\tilde{\mathcal{E}}_t$ the set of encoding indexes *not* in $\mathcal{E}_t \cup \mathcal{D}_t$, i.e., those indexes corresponding to the messages that are not encoded or decoded at terminal t. For example, in Fig. 2, we have $\tilde{\mathcal{E}}_1 = \{22\}$, $\tilde{\mathcal{E}}_2 = \{1\}$, $\tilde{\mathcal{E}}_3 = \emptyset$, and $\tilde{\mathcal{E}}_4 = \{21, 22\}$.

We will decode with ML decoders. There are several possible ML rules to choose from and we adopt the following one [48, Sec. 4]: terminal t chooses the messages $\hat{u}_{(\mathcal{D}_t)}$ if the $\hat{u}_{(\mathcal{D}_t)}$ are any of the messages that maximize

$$\begin{aligned} &P_Q\left(y_t^N \left| \boldsymbol{a}_{(\mathcal{E}_t)}^N,\, \boldsymbol{a}_{\hat{u}_{(\mathcal{D}_t)}}^N\right.\right) \\ &\quad := \sum_{\boldsymbol{a}_{(\tilde{\mathcal{E}}_t)}^N} Q\left(\boldsymbol{a}_{(\tilde{\mathcal{E}}_t)}^N\right) \cdot P\left(y_t^N \left| \boldsymbol{a}_{(\mathcal{E}_t)}^N,\, \boldsymbol{a}_{\hat{u}_{(\mathcal{D}_t)}}^N,\, \boldsymbol{a}_{(\tilde{\mathcal{E}}_t)}^N\right.\right) \end{aligned} \tag{56}$$

where the sum is over those $\boldsymbol{a}_{(\tilde{\mathcal{E}}_t)}^N$ constructed by concatenating $\tilde{N}$ code trees of length L, and where we have abused the notation by not including the appropriate subscripts for the distributions. Note that this rule is *not* a maximum *a posteriori* (MAP) decoding rule because the *random coding* distribution is used for averaging and not the distribution $P(u_{(\tilde{\mathcal{E}}_t)})$. This means that terminal t's decoding regions are independent of the $\boldsymbol{a}_{u_{(\tilde{\mathcal{E}}_t)}}^N$, a fact that we make use of.

Our approach to bounding the decoding error probability follows closely that of [16, Ch. 5]. We define the event

$$\mathrm{Error}_{t,\mathcal{S}} = \left\{\hat{U}_{t,(\mathcal{S})} \neq u_{(\mathcal{S})},\, \hat{U}_{t,(\mathcal{S}^C)} = u_{(\mathcal{S}^C)}\right\} \tag{57}$$

and the block error probability

$$P_{B,t,\mathcal{S}}\left(u_{(\mathcal{E}_{\mathrm{all}})}\right) = \Pr\left(\mathrm{Error}_{t,\mathcal{S}} \left| U_{(\mathcal{E}_{\mathrm{all}})} = u_{(\mathcal{E}_{\mathrm{all}})}\right.\right) \tag{58}$$

where $\mathcal{S} \subseteq \mathcal{D}_t$ and $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{D}_t$. The average of this error probability over the ensemble of all codes generated with $Q(\boldsymbol{a}_{(\mathcal{E}_{\mathrm{all}})}^N)$ is

$$\begin{aligned} \overline{P}_{B,t,\mathcal{S}}\left(u_{(\mathcal{E}_{\mathrm{all}})}\right) &= \sum_{\boldsymbol{a}_{u_{(\mathcal{E}_{\mathrm{all}})}}^N,\, y_t^N} Q\left(\boldsymbol{a}_{u_{(\mathcal{E}_{\mathrm{all}})}}^N\right) P\left(y_t^N \left| \boldsymbol{a}_{u_{(\mathcal{E}_{\mathrm{all}})}}^N\right.\right) \\ &\qquad \cdot \Pr\left(\mathrm{Error}_{t,\mathcal{S}} \left| u_{(\mathcal{E}_{\mathrm{all}})},\, \boldsymbol{a}_{u_{(\mathcal{E}_{\mathrm{all}})}}^N,\, y_t^N\right.\right) \end{aligned} \tag{59}$$

where the notation $\Pr(A|b)$ for the event A and the random variable B is used as a shorthand for $\Pr(A|B = b)$.

Continuing as in [16, Sec. 5.6] we have, for $0 \leq \rho \leq 1$

$$\overline{P}_{B,t,\mathcal{S}}\left(u_{(\mathcal{E}_{\mathrm{all}})}\right) \leq 2^{-N[E_{o,t,\mathcal{S}}^L(\rho,\, Q_{\boldsymbol{A}_{(\mathcal{E}_{\mathrm{all}})}^L}) - \rho R_{\mathcal{S}}]} \tag{60}$$

where $E_{o,t,\mathcal{S}}^L(\rho,\, Q_{\boldsymbol{A}_{(\mathcal{E}_{\mathrm{all}})}^L})$ is defined as

$$\begin{aligned} &-\frac{1}{L}\, \log_2 \sum_{\boldsymbol{a}_{(\mathcal{E}_t)}^L,\, \boldsymbol{a}_{(\mathcal{S}^C)}^L,\, y_t^L} Q\left(\boldsymbol{a}_{(\mathcal{E}_t)}^L,\, \boldsymbol{a}_{(\mathcal{S}^C)}^L\right) \\ &\quad \cdot \left[\sum_{\boldsymbol{a}_{(\mathcal{S})}^L} Q\left(\boldsymbol{a}_{(\mathcal{S})}^L\right) \cdot P_Q\left(y_t^L \left| \boldsymbol{a}_{(\mathcal{E}_t)}^L,\, \boldsymbol{a}_{(\mathcal{S})}^L,\, \boldsymbol{a}_{(\mathcal{S}^C)}^L\right.\right)^{\frac{1}{1+\rho}}\right]^{1+\rho} \end{aligned} \tag{61}$$

and $P_Q(y_t^L|\boldsymbol{a}_{(\mathcal{E}_t)}^L,\, \boldsymbol{a}_{(\mathcal{S})}^L,\, \boldsymbol{a}_{(\mathcal{S}^C)}^L)$ is defined as in (56). We further define the *random-coding exponents* by

$$\begin{aligned} &E_{r,t,\mathcal{S}}^L\left(R_{\mathcal{S}},\, Q_{\boldsymbol{A}_{(\mathcal{E}_{\mathrm{all}})}^L}\right) \\ &\qquad = \max_{0 \leq \rho \leq 1}\left[E_{o,t,\mathcal{S}}^L\left(\rho,\, Q_{\boldsymbol{A}_{(\mathcal{E}_{\mathrm{all}})}^L}\right) - \rho R_{\mathcal{S}}\right]. \end{aligned} \tag{62}$$

At this point, we prefer not to maximize over $Q_{\boldsymbol{A}_{(\mathcal{E}_{\mathrm{all}})}^L}$ because, given $R_{(\mathcal{E}_{\mathrm{all}})}$, this must be done for all terminals and exponents simultaneously. As in [16, p. 139], we would denote the distribution-optimized exponents by $E_{r,t,\mathcal{S}}^L(R_{\mathcal{S}})$.

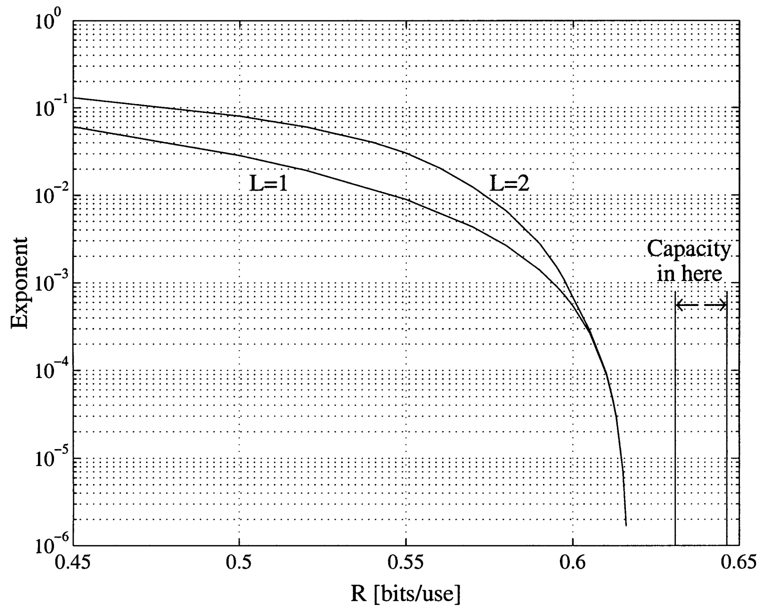

Fig. 11. Random coding exponents for the binary multiplying channel.

Next, we have

$$\left.\frac{\partial E^L_{o,\,t,\,\mathcal{S}}\left(\rho,\, Q_{\boldsymbol{A}^L_{(\mathcal{E}_{\rm all})}}\right)}{\partial\rho}\right|_{\rho\downarrow 0} = I_L\left(\boldsymbol{A}_{(\mathcal{S})};\, Y_t \middle| \boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}\right) \tag{63}$$

where $\rho \downarrow 0$ means ρ approaches zero from above. Thus, if R_S is less than (63) then the exponent $E^L_{r,\,t,\,\mathcal{S}}(R_\mathcal{S},\, Q_{\boldsymbol{A}^L_{(\mathcal{E}_{\rm all})}})$ will increase with ρ from $E^L_{o,\,t,\,\mathcal{S}}(0,\, Q_{\boldsymbol{A}^L_{(\mathcal{E}_{\rm all})}}) = 0$, i.e., $E^L_{r,\,t,\,\mathcal{S}}(R_\mathcal{S},\, Q_{\boldsymbol{A}^L_{(\mathcal{E}_{\rm all})}})$ is positive for these $R_\mathcal{S}$. Combining these results we have the following generalization of part of [16, Theorem 5.6.4, p. 143] (see also [52] for random-coding exponents for the MAC).

Theorem 5: The exponent $E^L_{r,\,t,\,\mathcal{S}}(R_\mathcal{S},\, Q_{\boldsymbol{A}^L_{(\mathcal{E}_{\rm all})}})$ is a positive function of $R_\mathcal{S}$ if

$$0 \le R_\mathcal{S} < I_L\left(\boldsymbol{A}_{(\mathcal{S})};\, Y_t \middle| \boldsymbol{A}_{(\mathcal{E}_t)}\boldsymbol{A}_{(\boldsymbol{S}^C)}\right) \tag{64}$$

where $\mathcal{S} \subseteq \mathcal{D}_t$ and $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{D}_t$.

This theorem gives a sequence of inner bounds to the reliability function [16, p. 160] of a DMN. One can additionally prove other generalizations of results known for the DMC (see [16, Secs. 5.6 and 5.7]). Theorem 5 implies Theorem 1 by fixing L and letting $\tilde{N}$ become large. Theorem 5 also implies the achievability of certain steady-state directed information rates (see Section II-C) by setting $\tilde{N} = 1$ and letting L become large.

One can, in fact, improve the error exponents in several ways [53, p. 129]. First, one can include a time-sharing random variable as discussed in Section IV-A (see [53, eq. (2.25)]). Second, one can use the method of types with an appropriate decoder to get universal error bounds [54]. Third, for low rates, one can use expurgated random coding ensembles [16, Sec. 5.7]. Fourth, one can derive random coding bounds for tree and convolutional codes [55]. It would also be useful to have *lower* bounds on the error probability which are similar to the bounds for the DMC [16, Sec. 5.8], [54, eq. (7)].

As an example, consider the binary multiplying two-way channel [1, Sec. 13]. This channel has binary (0 and 1) inputs and outputs where

$$Y_t = X_1 \cdot X_2, \qquad t = 1,\, 2. \tag{65}$$

It has long been known that $\mathcal{R}_1$ is strictly interior to the capacity region of this channel [1, Secs. 9 and 13] (see also [50], [56]–[58]). Consider the symmetric rates $R = R_1 = R_2$ and choose $Q_{X_1}(0) = Q_{X_2}(0) = q$. The exponents $E^1_{r,\,t,\,t}(R,\, Q_{X_1X_2})$ for $L = 1$ are then (see (61) and (62))

$$-\log_2\left[q + (1-q)\cdot\left(q^{1+\rho} + (1-q)^{1+\rho}\right)\right] \tag{66}$$

where $0 \le \rho \le 1$. For $L = 2$, we generate the code trees with $Q_{X_{t1}}(0) = q$ and $Q_{X_{t2}|X_{t1}Y_1}(0|x_{t1},\, y_1) = q_{x_{t1}y_1}$. Equivalently

$$Q\left(\boldsymbol{a}_t^2\right) = \begin{cases} q\cdot q_{00}\cdot q_{01}, & \boldsymbol{a}_t^2 = [0,\, 00] \\ q\cdot q_{00}\cdot(1-q_{01}), & \boldsymbol{a}_t^2 = [0,\, 01] \\ \vdots & \vdots \\ (1-q)(1-q_{10})(1-q_{11}), & \boldsymbol{a}_t^2 = [1,\, 11]. \end{cases} \tag{67}$$

See (11) for the interpretation of the code tree notation.

We consider the exponents as a function of the common rate R. Optimizing over ρ, q, q_{00}, q_{01}, and q_{11} we obtain the curves shown in Fig. 11.[1] These curves are, however, *not* the best ones for $L = 1$ or $L = 2$. For example, one can achieve zero error probability for $R \le 0.5$ by time sharing. One would notice this if one used a time-sharing random variable V and the expurgated ensemble of codes described in [16, Sec. 5.7]. In any case, the curves in Fig. 11 give a flavor of what happens for large R as L increases.

Finally, consider the rate R at which the error exponents go to zero. The best R for $L = 1$ is 0.61695. The best R for $L = 2$

[1]These curves are traditionally plotted with a linearly scaled y-axis rather than a logarithmic one. However, a logarithmic scale more clearly shows what happens near capacity.

is again 0.61695 for the distributions in (67), so that no rate increase is achieved. Thus, code trees can improve the *reliability* of communication even if they do not increase the *rate*. For $L = 3$, we obtain a rate increase with $R = 0.61964$. The best equal-rate inner bound we are aware of is $R = 0.63072$ [59]. The best equal-rate *outer* bound to date is $R \leq 0.64628$ [30], [60].

VIII. Concluding Remarks

The main ingredient to getting the capacity results for DMNs, namely, random coding with strategies/code trees, can already be found in [1, Sec. 15]. The strength of the code tree approach is its generality—two obvious shortcomings are the complicated codes it describes and the difficulty in computing rate regions.

There are several extensions of the theory presented here. First, extensions to certain continuous-alphabet networks and networks with memory can be found in [61]. Second, an open issue is how to build codes having practical encoders and decoders. Code trellises might prove useful in this respect. Finally, we point out that an interesting achievable rate region for DMNs was put forward in [62]. This region combines and generalizes several superposition coding and binning regions. Some improvements on [62] can be found in [63].

Appendix A
Equivalence of d-Separation Rules

We begin with the d-separation rule of [19, p. 117] which considers an undirected path (a sequence of edges) P between certain vertices. The path is assumed to be cycle free so that if the vertex v lies along P, then v has exactly two edges of P touching it.

Definition 5: Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be disjoint subsets of the vertices of an FDG. Then $\mathcal{Z}$ *d-separates* $\mathcal{X}$ from $\mathcal{Y}$ if along every cycle-free path P between a vertex in $\mathcal{X}$ and a vertex in $\mathcal{Y}$ there is a vertex v such that either

1. v has two incoming edges along P *and* neither v nor its descendants are in $\mathcal{Z}$, or
2. v has at most one incoming edge along P *and* v is in $\mathcal{Z}$.

The requirement that P be cycle free is needed to avoid paths such as the following one in Fig. 1:

$$U^B \to X_3 \to Y_3 \leftarrow X_3 \leftarrow Y_2 \to X_3 \to Y_3.$$

Lemma 3: The d-separation rules of Definitions 1 and 5 are equivalent.

Proof: Suppose that $\mathcal{Z}$ does not d-separate $\mathcal{X}$ from $\mathcal{Y}$ according to Definition 1. Then after Step 2 there must be a path between $\mathcal{X}$ and $\mathcal{Y}$. Suppose that this path has no intermediate vertices. Then clearly $\mathcal{Z}$ does not d-separate $\mathcal{X}$ from $\mathcal{Y}$ according to Definition 5. Next, consider the two cases for which there is at least one vertex v along the path between $\mathcal{X}$ and $\mathcal{Y}$.

1. If neither v nor its descendants are in $\mathcal{Z}$, then the two edges touching v must have been retained in Step 1 by moving backward from $\mathcal{X}$ or $\mathcal{Y}$. But then one of these edges must go out of v.

2. If v is in $\mathcal{Z}$, then it cannot have an outgoing edge because these edges were cut in Step 2.

Thus, d-separation according to Definition 5 implies d-separation according to Definition 1. Conversely, suppose that $\mathcal{Z}$ d-separates $\mathcal{X}$ from $\mathcal{Y}$ according to Definition 1. Then along any path P between $\mathcal{X}$ and $\mathcal{Y}$ one edge must have been cut in Step 1 or Step 2.

1. Suppose that P is cut in Step 1. Then there is a vertex v along this path which is not in $\mathcal{Z}$ and has no descendants in $\mathcal{X}$, $\mathcal{Y}$, or $\mathcal{Z}$. If v has no outgoing edges, then we have d-separation according to Definition 5. If v has an outgoing edge, then the vertex w to which this edge goes is a descendant of v, and hence w is not in $\mathcal{Z}$ and cannot have descendants in $\mathcal{X}$, $\mathcal{Y}$, or $\mathcal{Z}$. We can now perform the same steps for vertex w as for vertex v until we arrive at a vertex u that has an edge going to a vertex x in $\mathcal{X}$ or y in $\mathcal{Y}$. However, this contradicts the result that u cannot have descendants in $\mathcal{X}$ or $\mathcal{Y}$. Hence, u must have two incoming edges and we have d-separation according to Definition 5.

2. Suppose that P was not cut in Step 1 but was cut in Step 2. Then the vertex v out of which the cut edge came must be in $\mathcal{Z}$. Furthermore, this edge is outgoing so that v has at most one incoming edge.

Thus, d-separation according to Definition 1 implies d-separation according to Definition 5. □

Appendix B
Properties of Causally Conditioned Entropy

We begin with two bounds whose proofs are omitted because they are rather easy (see [34, Ch. 3]).

Property 1 (Bounds on Causally Conditioned Entropy):

$$H\left(X^N \middle| Y^N\right) \leq H\left(X^N \middle\| Y^N\right) \leq H\left(X^N\right) \tag{68}$$

with equality on the left if and only if

$$H(X_n|X^{n-1}Y^n) = H(X_n|X^{n-1}Y^N)$$

for all $n = 1, 2, \ldots, N$, and with equality on the right if and only if $I(X_n; Y^n|X^{n-1}) = 0$ for all $n = 1, 2, \ldots, N$.

Property 2 (Bounds on Directed Information [24]):

$$0 \leq I\left(X^N \to Y^N\right) \leq I\left(X^N; Y^N\right) \tag{69}$$

with equality on the left if and only if $I(X^n; Y_n|Y^{n-1}) = 0$ for all $n = 1, 2, \ldots, N$, and with equality on the right if and only if $H(Y_n|Y^{n-1}X^n) = H(Y_n|Y^{n-1}X^N)$ for all $n = 1, 2, \ldots, N$.

One might be tempted to guess that equality holds on the right in both cases only if X^N and Y^N are independent. However, this is not true, as the following example demonstrates. Let X_1 and Y_1 be independent with $\Pr(X_1 = 0) = \Pr(X_1 = 1) = 1/2$ and $\Pr(Y_1 = 0) = \Pr(Y_1 = 1) = 1/2$ and let $X_2 = X_1$ and $Y_2 = X_1 \oplus Y_1$, where $\oplus$ denotes addition modulo 2. Then we have $I(X^2 \to Y^2) = 0$ but $I(X^2; Y^2) = 1$ bit.

Property 3 (Chain Rules):

$$\begin{aligned} I\left(X^N Y^N \to Z^N\right) &= I\left(X^N \to Z^N\right) \\ &\quad + I\left(Y^N \to Z^N \middle\| X^N\right) \end{aligned} \tag{70}$$

$$\begin{aligned} I\left(X^N \to Y^N Z^N\right) &= I\left(X^N \to Y^N \middle\| DZ^N\right) \\ &\quad + I\left(X^N \to Z^N \middle\| Y^N\right) \end{aligned} \tag{71}$$

where $DZ^N := 0, Z_1, Z_2, \dots, Z_{N-1}$ is the *delay* of Z^N by one time instant (with discard of the last digit).

Recall that $I(X^N \to Y^N) \neq I(Y^N \to X^N)$ is possible. However, it would be intuitively pleasing if

$$I(X^N \to Y^N) + I(Y^N \to X^N) = I(X^N; Y^N)$$

as suggested by a result of Marko [23, eq. (14)]. However, this relation does not hold in general.

Property 4 (Oppositely Directed Informations):

$$\begin{aligned} I\left(X^N \to Y^N\right) &+ I\left(Y^N \to X^N\right) \\ &= I\left(X^N; Y^N\right) + I\left(X^N \to Y^N \middle\| DX^N\right). \end{aligned} \tag{72}$$

We next consider stationarity properties. A *discrete stationary source* (DSS) is a device that emits a sequence $U_1, U_2, U_3, \dots$ of discrete random variables such that, for every $n \geq 1$ and $L \geq 1$, the random vectors $[U_1 \cdots U_L]$ and $[U_{n+1} \cdots U_{n+L}]$ have the same probability distribution. This means that for every window length L along the DSS output sequences one sees the same statistical behavior regardless of where the window is placed along the output sequences [16, p. 56]. We will consider the case where $U_n = (X_n, Y_n, Z_n)$, i.e., the sequences X, Y, and Z are the outputs of a DSS.

Property 5 (Entropy Rates of a DSS): If the sequences $X_1, X_2, X_3, \dots$ and $Y_1, Y_2, Y_3, \dots$ are the output sequences of a DSS, then

1. $H(X_L|X^{L-1}Y^L) \leq H_L(X\|Y)$ for all $L \geq 1$,
2. $H(X_L|X^{L-1}Y^L)$ is nonincreasing with L,
3. $H_L(X\|Y)$ is nonincreasing with L,
4. $\lim_{L\to\infty} H(X_L|X^{L-1}Y^L) = \lim_{L\to\infty} H_L(X\|Y)$, i.e., both of these limits exist and have the same value $H_\infty(X\|Y)$.

These four properties are known if the DSS has a single output or, equivalently, if the Y sequence is independent of the X sequence. The proof of the more general Property 5 follows the same steps as the proof in [16, p. 57] and is omitted. Property 5 can be applied to prove the following.

Property 6 (Directed Information Rates of a DSS): If X and Y are output sequences of a DSS, then $\lim_{L\to\infty} I_L(X \to Y)$ exists and is given by

$$I_\infty(X \to Y) = H_\infty(Y) - H_\infty(Y\|X). \tag{73}$$

If Z is also an output sequence of the DSS, then $\lim_{L\to\infty} I_L(X \to Y\|Z)$ exists and is given by

$$I_\infty(X \to Y\|Z) = H_\infty(Y\|Z) - H_\infty(Y\|XZ). \tag{74}$$

Proof: Consider (73). By Property 5, $H_\infty(Y)$ and $H_\infty(Y\|X)$ exist. But the limit of a real sequence whose elements are the term-by-term differences of the elements of two real convergent sequences exists. Furthermore, this limit is the difference of the two limits of the original sequences [28, p. 223]. Thus, we have

$$\begin{aligned} \lim_{L\to\infty} I_L(X \to Y) &= \lim_{L\to\infty} [H_L(Y) - H_L(Y\|X)] \\ &= \lim_{L\to\infty} H_L(Y) - \lim_{L\to\infty} H_L(Y\|X). \end{aligned}$$

This proves (73); (74) is proved in the same manner. □

APPENDIX C
AN OUTER BOUND FOR COMMON-OUTPUT DMNs

We prove Theorem 4. First, we bound the right-hand side of (16) by using the independence of the $\boldsymbol{A}_s$

$$\begin{aligned} I_N &\left(\boldsymbol{A}_{(\mathcal{S})}; Y \middle| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\mathcal{S}^C)}\right) \\ &\leq H_N\left(\boldsymbol{A}_{(\mathcal{S})} \middle| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\mathcal{S}^C)} \boldsymbol{A}_{(\mathcal{D}_t^C)}\right) \\ &\quad - H_N\left(\boldsymbol{A}_{(\mathcal{S})} \middle| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\mathcal{S}^C)} \boldsymbol{A}_{(\mathcal{D}_t^C)} Y\right) \\ &= I_N\left(\boldsymbol{A}_{(\mathcal{S})}; Y \middle| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\mathcal{S}^C)} \boldsymbol{A}_{(\mathcal{D}_t^C)}\right). \end{aligned} \tag{75}$$

Expanding (75) and simplifying we obtain

$$\begin{aligned} I_N &\left(\boldsymbol{A}_{(\mathcal{S})}; Y \middle| \boldsymbol{A}_{(\mathcal{E}_t)} \boldsymbol{A}_{(\mathcal{S}^C)} \boldsymbol{A}_{(\mathcal{D}_t^C)}\right) \\ &= \frac{1}{N} \sum_{n=1}^{N} I\left(X_{(\mathcal{S})n}; Y_n \middle| X_t^n X_{(\mathcal{S}^C)}^n X_{(\mathcal{D}_t^C)}^n Y^{n-1}\right) \\ &\leq \frac{1}{N} \sum_{n=1}^{N} I\left(X_{(\mathcal{S})n}; Y_n \middle| X_{tn} X_{(\mathcal{S}^C)n} X_{(\mathcal{D}_t^C)n} Y^{n-1}\right). \end{aligned} \tag{76}$$

Next, consider the chain of (in)equalities in [30, eq. (14)]. We can follow virtually the same steps to find that

$$\begin{aligned} 0 &\leq I\left(U_{(\mathcal{E}_1)}; \dots; U_{(\mathcal{E}_T)} \middle| Y^N\right) \\ &\leq \sum_{n=1}^{N} I\left(X_{1n}; \dots; X_{Tn} \middle| Y_n Y^{n-1}\right) \\ &\quad - I\left(X_{1n}; \dots; X_{Tn} \middle| Y^{n-1}\right). \end{aligned} \tag{77}$$

Now let $V = [W, Y^{W-1}]$ and $X_s = X_{sW}$ where $P_W(n) = 1/N$ for $n = 1, \dots, N$. We then have (26) and (27) from (76) and (77), respectively. The cardinality bound on V follows from [2, p. 310].

APPENDIX D
APPROACHABLE RATES FOR A NOISY BINARY ADDER CHANNEL WITH FULL FEEDBACK

A. Equal-Rate Point on the Cover-Leung Region Boundary

Consider the $Y = X_1 + X_2 + Z$ channel of Section VI-A and the rate region $\mathcal{R}_1^{\mathrm{CL}}$. We use a binary V with $P_V(0) = P_V(1) = 1/2$, and $\Pr(X_1 \neq V) = \Pr(X_2 \neq V) = q$. The rate $R = R_1 = R_2$ is then bounded by

$$\begin{aligned} R &\leq h(q)/2 \\ 2R &\leq h\left(\left[q^2 + (1-q)^2\right]/2\right). \end{aligned} \tag{78}$$

The bounds meet if $q = [q^2 + (1-q)^2]/2$, or $q = 1 - 1/\sqrt{2}$. The best rate is thus $R = h(1/\sqrt{2})/2 \approx 0.436\,21$ bits per use.

For general $P(v, x_1, x_2, y)$ define $q_{tv} = \Pr(X_t = 0|V = v)$. Straightforward manipulations of the informations yield

$$\begin{aligned} I(X_1; Y|X_2V) &= \sum_v P(v)h(q_{1v})/2 \\ I(X_2; Y|X_1V) &= \sum_v P(v)h(q_{2v})/2 \\ I(X_1X_2; Y) &= h^{(4)}\left(P_Y(0), P_Y(1), P_Y(2), P_Y(3)\right) - 1 \end{aligned} \tag{79}$$

where

$$h^{(4)}(p_1,\, p_2,\, p_3,\, p_4) = \sum_{i=1}^{4} -p_i \log(p_i)$$

and

$$\begin{aligned}
P_Y(0) &= \sum_v P(v) q_{1v} q_{2v}/2 \\
P_Y(1) &= \sum_v P(v)[q_{1v}q_{2v} + q_{1v}(1-q_{2v}) + (1-q_{1v})q_{2v}]/2 \\
P_Y(2) &= \sum_v P(v)[(1-q_{1v})(1-q_{2v}) + q_{1v}(1-q_{2v}) \\
&\qquad + (1-q_{1v})q_{2v}]/2 \\
P_Y(3) &= \sum_v P(v)(1-q_{1v})(1-q_{2v})/2.
\end{aligned}$$

We would like to show that a binary V is best for equal-rate points. We first use the convexity of the entropies and Jensen's inequality [15, pp. 25–30] to write

$$\begin{aligned}
I(X_1X_2;\, Y) &\le h^{(4)}\left(\frac{P_Y(0)+P_Y(3)}{2},\, \frac{P_Y(1)+P_Y(2)}{2},\right. \\
&\qquad \left.\frac{P_Y(1)+P_Y(2)}{2},\, \frac{P_Y(0)+P_Y(3)}{2}\right) - 1 \\
&= h(P_Y(0)+P_Y(3)) \\
&= h\left(\sum_v P(v)(1-t_v)/2\right) \qquad (80)
\end{aligned}$$

where $t_v := q_{1v}(1-q_{2v}) + (1-q_{1v})q_{2v}$.

The informations (79) are virtually identical with the informations in [41, eqs. (3) and (5)]. We may thus use the same function $\phi(\cdot)$ defined there, namely,

$$\phi(t) = \begin{cases} \left(1-\sqrt{1-2t}\right)/2, & \text{for } 0 \le t \le 1/2 \\ \left(1-\sqrt{2t-1}\right)/2, & \text{for } 1/2 < t \le 1. \end{cases} \qquad (81)$$

In [41], it is shown that the composite function $h(\phi(\cdot))$ is symmetrical around $t = 1/2$ and convex-$\cap$ in t for $0 \le t \le 1$. Following the same steps as in [41, eq. (8)] we arrive at

$$R \le h\left(\phi\left(\bar{t}\right)\right)/2 \qquad (82)$$

where $\bar{t} = \sum_v P(v) t_v$. Combining (82) and (80), we find that R satisfies

$$R \le \min_{0 \le t \le 1/2} \{h(\phi(t))/2,\, h((1-t)/2)/2\} \qquad (83)$$

or, by setting $q = \phi(t)$ so that $t = 2q(1-q)$, we have

$$R \le \min_{0 \le q \le 1/2} \left\{h(q)/2,\, h\left(\left[q^2 + (1-q)^2\right]/2\right)/2\right\} \qquad (84)$$

which is the same as the bounds (78). Thus, the rate point

$$(h(1/\sqrt{2})/2,\, h(1/\sqrt{2})/2) \approx (0.43621,\, 0.43621)$$

bits per use lies on the boundary of $\mathcal{R}^{\mathrm{CL}}$.

B. *Equal-Rate Point With Directed Informations*

Consider the rate region of Corollary 1. We use a memory 1 random coding technique with $P_{V_n}(0) = P_{V_n}(1) = 1/2$ for all n, and

$$\Pr\left(X_{tn} \ne V_n \,\middle|\, X_{t(n-1)} = x,\, Y_{n-1} = y\right) = q_{t,\,xy}. \qquad (85)$$

The random coding can thus be described by a Markov chain having six states. We call the states $\sigma_{tn} = (x_{t(n-1)},\, y_{n-1})$ and the random variable corresponding to these states Σ_{tn}. The state diagram of the entire system can be shown to have eight states $\Sigma_n = (X_{1(n-1)},\, X_{2(n-1)},\, Y_{n-1})$.

The best $q_{t,\,xy}$ we found were

$$\begin{array}{lll}
q_{1,\,00} = 0.2584, & q_{1,\,01} = 0.7148, & q_{1,\,02} = 1 \\
q_{1,\,11} = 0, & q_{1,\,12} = 1 - 0.7148, & q_{1,\,13} = 0.2584
\end{array}$$

$$\begin{array}{lll}
q_{2,\,00} = 0.2584, & q_{2,\,01} = 1 - 0.7148, & q_{2,\,02} = 0 \\
q_{2,\,11} = 1, & q_{2,\,12} = 0.7148, & q_{2,\,13} = 0.2584.
\end{array}$$

The resulting steady-state distribution of the system is

$$\begin{array}{lll}
p_{000} = 0.1486, & p_{001} = 0.1486, & p_{011} = 0.1014 \\
p_{012} = 0.1014, & p_{101} = 0.1014, & p_{102} = 0.1014 \\
p_{112} = 0.1486, & p_{113} = 0.1486 &
\end{array}$$

where $p_{ijk} = \Pr(\Sigma_n = (i,\, j,\, k))$. We bound the steady-state entropies as follows (see Appendix B and [15, p. 71]):

$$\begin{aligned}
&H\left(Y_\ell \,\middle|\, \Sigma_{\ell-2} V^\ell X_2^\ell Y^{\ell-1}\right) = 1.43879 \\
&\quad \le H_\infty(Y \| X_2 V) \\
&\quad \le H\left(Y_\ell \,\middle|\, \Sigma_{2(\ell-2)} V^{\ell-2\cdots\ell} X_2^{\ell-2\cdots\ell} Y^{\ell-2\cdots\ell-1}\right) = 1.44559
\end{aligned} \qquad (86)$$

where $V^{k\cdots\ell}$ denotes $V_k,\, V_{k+1},\, \ldots,\, V_\ell$, and similarly for $X_2^{k\cdots\ell}$ and $Y^{k\cdots\ell}$. Furthermore, we have

$$\begin{aligned}
H\left(Y_\ell \,\middle|\, Y^{\ell-1}\Sigma_{\ell-2}\right) &= 1.87758 \le H_\infty(Y) \\
&\le H\left(Y_\ell \,|\, Y_{\ell-2}Y_{\ell-1}\right) = 1.87764.
\end{aligned} \qquad (87)$$

All of the quantities are in bits per use. Because $H(Y|X_1X_2) = 1$ and because both users have the same directed information rates, we have

$$\begin{aligned}
I_\infty(X_1 \to Y \| X_2 V) &\ge 1.43879 - 1 = 0.43879 \\
I_\infty(X_2 \to Y \| X_1 V) &\ge 1.43879 - 1 = 0.43879 \\
I_\infty(X_1X_2 \to Y) &\ge 1.87758 - 1 = 0.43879 \cdot 2.
\end{aligned}$$

Thus, $R_1 = R_2 = 0.43879$ is approachable. This is beyond the rate point $R_1 = R_2 = 0.43621$ that lies on the boundary of $\mathcal{R}^{\mathrm{CL}}$.

ACKNOWLEDGMENT

The author is grateful to Prof. Emeritus J. L. Massey for providing the impetus for this work. The author also wishes to thank Prof. em. H. Marko for sending him papers on causality, Dr. J. Sayir and Dr. M. Gastpar for helpful comments on drafts, and the associate editor and the three reviewers for their useful remarks.

REFERENCES

[1] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Mathematical Statistics and Probability*, J. Neyman, Ed. Berkeley, CA: Univ. Calif. Press, 1961, vol. 1, pp. 611–644. Reprinted in *Claude Elwood Shannon: Collected Papers*, N. J. A. Sloane and A. D. Wyner, Eds. New York: IEEE Press, 1993, pp. 351–384..

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

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

[4] ——, "The discrete memoryless channel with two senders and one receiver," in *Proc. 2nd Int. Symp. Information Theory (1971)*, Tsahkadsor, Armenian S.S.R., 1973, pp. 103–135.

[5] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 648–657, Nov. 1980.

[6] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inform. Theory*, vol. 35, pp. 605–619, May 1989.

[7] ——, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 733–751, July 1989.

[8] R. S. Cheng and S. Verdú, "On limiting characterizations of memoryless multiuser capacity regions," *IEEE Trans. Inform. Theory*, vol. 39, pp. 609–612, Mar. 1993.

[9] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory (1971)*, Tsahkadsor, Armenian S.S.R., 1973, pp. 23–52.

[10] ——, "The capacity region of a channel with two senders and two receivers," *Ann. Probab.*, vol. 2, pp. 805–814, Oct. 1974.

[11] E. C. van der Meulen, "Random coding theorems for the general memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 180–190, Mar. 1975.

[12] ——, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 1–37, Jan. 1977.

[13] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inform. Theory*, vol. 41, pp. 35–43, Jan. 1995.

[14] A. B. Carleial, "On the capacity of multiple-terminal communication networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975.

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

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

[17] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[18] ——, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 226–228, Mar. 1975.

[19] J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artificial Intell.*, vol. 29, pp. 241–288, 1986. Reprinted in *Great Papers in Computer Science*, P. Laplante, Ed. Amsterdam, The Netherlands: Elsevier, 1996, pp. 586–627.

[20] J. T. A. Koster, "Markov properties of nonrecursive causal models," *Ann. Statist.*, vol. 24, no. 5, pp. 2148–2177, Oct. 1996.

[21] J. Pearl and R. Dechter, "Identifying independencies in causal graphs with feedback," in *Uncertainty in Artificial Intelligence, Proc. 12th Conf.*, E. Horvitz and F. Jensen, Eds. San Francisco, CA: Morgan Kaufmann, 1996, pp. 420–426.

[22] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[23] H. Marko, "The bidirectional communication theory—A generalization of information theory," *IEEE Trans. Commun.*, vol. COM-21, pp. 1345–1351, Dec. 1973.

[24] J. L. Massey, "Causality, feedback and directed information," in *Proc. 1990 IEEE Int. Symp. Information Theory and Its Applications*, Hawaii, Nov. 27–30, 1990, pp. 303–305.

[25] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[26] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948. Reprinted in *Claude Elwood Shannon: Collected Papers*, N. J. A. Sloane and A. D. Wyner, Eds. New York: IEEE Press, 1993, pp. 5–83.

[27] T. M. Cover, "An achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 399–404, July 1975.

[28] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*, 3rd ed. Berlin, Germany: Springer-Verlag, 1997.

[29] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 49–60, Jan. 1981.

[30] A. P. Hekstra and F. M. J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 44–53, Jan. 1989.

[31] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 221–238, Sept. 1956. Reprinted in *Claude Elwood Shannon: Collected Papers*, N. J. A. Sloane and A. D. Wyner, Eds. New York: IEEE Press, 1993, pp. 221–238.

[32] R. L. Dobrushin, "Information transmission in a channel with feedback," *Theory Probab. Appl.*, vol. 34, pp. 367–383, Dec. 1958.

[33] T. S. Han, "A general coding scheme for the two-way channel," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 35–44, Jan. 1984.

[34] G. Kramer, *Directed Information for Channels With Feedback*, ser. ETH Series in Information Processing. Konstanz, Switzerland: Hartung-Gorre Verlag, 1998, vol. 11.

[35] H. Liao, "Multiple access channels," Ph.D. dissertation, Univ. Hawaii, Honolulu, HI, 1972.

[36] B. E. Hajek and M. B. Pursley, "Evaluation of an achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 36–46, Jan. 1979.

[37] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[38] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[39] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 292–298, May 1981.

[40] F. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 93–95, Jan. 1982.

[41] ——, "On multiple access channels with feedback," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 842–845, Nov. 1984.

[42] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 623–629, July 1984.

[43] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 306–311, May 1979.

[44] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 120–122, Jan. 1981.

[45] ——, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 120–122, Jan. 1981.

[46] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.

[47] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 772–777, Nov. 1973.

[48] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Pered. Inform.*, vol. 10, no. 3, pp. 3–14, 1974.

[49] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 60–64, Jan. 1977.

[50] G. Dueck, "Partial feedback for two-way and broadcast channels," *Inform. Contr.*, vol. 46, no. 1, pp. 1–15, July 1980.

[51] L. H. Ozarow and S. K. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 667–671, July 1984.

[52] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sept. 1973.

[53] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124–142, Mar. 1985.

[54] Y. Liu and B. L. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, pp. 376–386, Mar. 1996.

[55] R. L. Peterson and D. J. Costello, Jr., "Error probability and free distance bounds for two-user tree codes on multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 658–670, Nov. 1980.

[56] G. Dueck, "The capacity region of the two-way channel can exceed the inner bound," *Inform. Contr.*, vol. 40, pp. 258–266, Mar. 1979.

[57] J. P. M. Schalkwijk, "The binary multiplying channel—A coding scheme that operates beyond Shannon's inner bound region," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 107–110, Jan. 1982.

[58] ——, "On an extension of an achievable rate region for the binary multiplying channel," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 445–448, May 1983.

[59] H. B. Meeuwissen, "Coding strategies for two-way communication," Inst. Continuing Education, Eindhoven Univ. Technol., Eindhoven, The Netherlands, Res. Rep., 1995.

[60] Z. Zhang, T. Berger, and J. P. M. Schalkwijk, "New outer bounds to capacity regions of two-way channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 383–386, May 1986.

[61] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, Sept. 2000.

[62] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," in *Proc. 2001 IEEE Int. Symp. Information Theory*, Washington, D.C., June 24–29, 2001, p. 159.

[63] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc. 2002 IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June 30–July 5, 2002, p. 136.

[64] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 155–165, Mar. 1980.

[65] T. S. Han and K. Kobayashi, "A unified achievable rate regionfor a general class of multiterminalsource coding systems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 277–288, May 1980.