Information Networks With In-Block Memory

Gerhard Kramer

Abstract—A class of channels is introduced for which there is memory inside blocks of a specified length and no memory across the blocks. The multi-user model is called an information network with in-block memory (NiBM). It is shown that block-fading channels, channels with state known causally at the encoder, and relay networks with delays are NiBMs. A cut-set bound is developed for NiBMs that unifies, strengthens, and generalizes existing cut bounds for discrete memoryless networks. The bound gives new finite-letter capacity expressions for several classes of networks including point-to-point channels, and certain multiaccess, broadcast, and relay channels. Cardinality bounds on the random coding alphabets are developed that improve on existing bounds for channels with action-dependent state available causally at the encoder and for relays without delay. Finally, quantize-forward network coding is shown to achieve additive, Gaussian noise channels, symmetric power constraints, and a multicast session.

Index Terms—capacity, feedback, relay channels, networks

I. INTRODUCTION

Communication channels often have memory, e.g., due to bandwidth limitations and dispersion. The memory is often modeled as being finite and of a sliding-window type, e.g., a convolution. However, in a network environment with bursty traffic and interference one often schedules users to dedicated time-frequency slots and with time-frequency offsets between successive slots. A pragmatic approach is then to model the channel as having memory inside a block and as being memoryless across blocks. We say that such channels have in-block memory or iBM.

This paper studies networks with iBM (NiBMs) where two central themes are memory and feedback. Several classes of channels fall into the NiBM framework, including block-fading channels [1], channels with state known causally at the encoder [2], and relay networks with delays [3]. In fact, the original motivation for this work was to show that the theory for relay networks with delays can be derived from theory for discrete memoryless networks (DMNs). We only later realized that NiBMs include block fading channels and channels with state known causally at the encoders.

This document is organized as follows. Section II presents the NiBM model. Section III defines the capacity region of a NiBM and introduces notation. Section IV states our main technical result: a cut-set bound on reliable communication rates. Sections V and VI apply the bound to point-to-point and multiuser channels, and they show that NiBMs let us unify, strengthen, and generalize existing theory for several classes of networks. For example, we derive new capacity theorems and new cardinality bounds on random variables. Section VII extends the approaches to relay networks. Several proofs are developed in the Appendices.

II. MODEL

The general DMN model was studied in [4] and a bounding tool for a class of DMNs called relay networks was developed in [5] (see also [6]). We use terminology and notation from [7]. Recall that a DMN with K nodes has each node k, k = 1, 2, . . . , K, dealing with four types of random variables.

- Messages W_km, m = 1, 2, . . . , M_k, that have entropy H(W_km) = B_km bits where M_k is the number of messages at node k. The rate of message W_km is thus R_km = B_km/n bits per channel use. The {W_km} are mutually statistically independent for all m and k.
- Channel inputs X_k,i, i = 1, 2, . . . , n, with alphabet X_k. We interpret i as a time index but it could alternatively represent frequency or space, for example.
- Channel outputs Y_k,i, i = 1, 2, . . . , n, with alphabet Y_k.
- Message estimates W_{k\ellm}, \ellm \in D(k), where D(k) is a decoding index set whose elements are selected pairs \ellm, \ell \neq k, of message indices from other nodes.

Let K = \{1, 2, . . . , K\} be the set of nodes; let E(k) = \{k1, k2, . . . , kM_k\} encode the encoding index set of node k; let Y_k^i = Y_{k1}Y_{k2} . . . Y_{ki};\; let r(x, y) be the remainder when x is divided by y. For a set S \subseteq K we write E(S) = \bigcup_{k \in S} E(k) and X_{S,i} = \{X_{ki} : k \in S\}. For a set S of integer pairs km we write W_S = \{W_{km} : km \in S\}. The relationships between the random variables are as follows.

- Without feedback, node k chooses X_{ki} as a function of W_{E(k)} only. The X_{ki}(W_{E(k)}) are called codewords.
- With feedback, node k chooses functions a_{ki}, i = 1, 2, . . . , n, such that

\[ X_{ki} = a_{ki}(W_{E(k)}, Y_k^{i-1}). \]  

(1)

We call a_k(W_{E(k)}::) a code function or an adaptive codeword since it replaces the notion of a codeword. For a finite alphabet Y_k one may interpret a_k(W_{E(k)}::) as a code tree (see [8] Sec. 15), [4] Sec. 5), and [9] Ch. 9). We write a_k(W_{E(k)}::) as A_k(W_{E(k)}::) when we wish to emphasize that A_k is a random variable. The alphabet of
The capacity region $C$ of a NiBM is the closure of the set of rate-tuples $(R_{kn}: 1 \leq k \leq K, 1 \leq m \leq M_k)$ such that for any positive $\epsilon$ there is an $n$ and code functions and decoders for which the error probability

$$P_e = \Pr \left[ \bigcup_k \bigcup_{\ell m \in D(k)} \{\hat{W}^{(k)}_{\ell m} \neq W_{\ell m}\} \right]$$

is at most $\epsilon$.

B. Causal Conditioning and Directed Information

We use notation from [7] for causal conditioning and directed information (see also [11, 12, 13]). The probability of $x^L$ causally conditioned on $y^L$ and conditioned on $a$ is defined as

$$P(x^L|y^L) = \prod_{i=1}^L P(x_i|x^{i-1}, y^i)$$

$$P(x^L|y^L|a) = \prod_{i=1}^L P(x_i|x^{i-1}, y^i, a).$$

As done here, we will drop subscripts on probability distributions if the argument is the lowercase version of the random variable. Causally-conditioned entropy is defined as

$$H(X^L|Y^L) = \sum_{i=1}^L H(X_i|X^{i-1}Y^i)$$

$$H(X^L|Y^L|A) = \sum_{i=1}^L H(X_i|X^{i-1}Y^i, A)$$

The commas in [12] and [13] emphasize that the pair $X^L, Z^L$ should here be considered as a length-$L$ sequence of pairs $(X_1, Z_1), (X_2, Z_2), \ldots, (X_L, Z_L)$. As another example of such notation, we write the directed information flowing from $X_1^L, X_2^L$ to $Y^L$ when causally conditioned on $Z_1^L, Z_2^L$ as

$$I(X_1^L, X_2^L \to Y^L|Z_1^L, Z_2^L) = H(Y^L|Z_1^L, Z_2^L) - H(Y^L|X_1^L, X_2^L, Z_1^L, Z_2^L).$$

C. Further Notation

The functional dependence [11] implies that $P(x_{k,i}|a_k^L, y_k^{i-1})$ takes on the value 1 only for that letter $x_{k,i}$ satisfying [11], and is 0 otherwise. To emphasize such dependence, we write $1(x_{k,i}|a_k^L, y_k^{i-1})$ in place of $P(x_{k,i}|a_k^L, y_k^{i-1})$, and similarly $1(x_{k,i}^L|a_k^L, 0y_k^{i-1})$ in place of $P(x_{k,i}^L|a_k^L, 0y_k^{i-1})$. The expression $0y_k^{i-1}$ denotes the concatenation of 0 and $y_k^{i-1}$.

It will be convenient to split symbol strings into blocks of length $L$. We use the notation

$$a_k^L = a_k, l(i-1)+1, a_k, l(i-1)+2, \ldots, a_k, l(i-1)+L$$

$$x_k^L, i = x, k, l(i-1)+1, x, k, l(i-1)+2, \ldots, x, k, l(i-1)+L$$

$$y_k^L, i = y, k, l(i-1)+1, y, k, l(i-1)+2, \ldots, y, k, l(i-1)+L.$$
**D. Channel Distribution**

We have defined the channel using the function (4). It will be convenient to alternatively define the channel by a probability distribution. Consider $P(a_k^n, x_k^n, y_k^n)$ that factors as

$$
\prod_{k=1}^{K} P(a_k^n) P(x_k^n|x_k^0, 0y_k^{n-1}) P(y_k^n|x_k^n).
$$

(15)

The $P(y_k^n|x_k^n)$ further factors into $m = \lceil n/L \rceil$ blocks as

$$
\prod_{i=1}^{m-1} P(y_{k,i}^L|x_{k,i}^L) P(y_{k,i+1}^{L'}|x_{k,i}^{L'}, x_{k,i+1}^L).
$$

(16)

where the last block has length $L' = n - (m-1)L$. We focus on $n = mL$ so that $L' = L$ and all blocks have length $L$.

**Remark 3:** The expressions (15)-(16) let us define the channel by using the block-invariant distribution $P(y_k^n|x_k^n)$ rather than by using $Z$ and the functions in (4). We further have

$$
P(y_k^n|a_k^n) = P(y_k^n|x_k^n).
$$

(17)

Thus, we may view the channel as being defined by the functional relations (4), by $P(y_k^n|x_k^n)$, or by $P(y_k^n|a_k^n)$.

**E. Linear Channels**

We consider several examples where the channel alphabets are the field $\mathbb{F}$. We write the channel inputs and outputs as vectors $X_k = [X_{k,1} \ldots X_{k,L}]^T$ and $Y_k = [Y_{k,1} \ldots Y_{k,L}]^T$, respectively. For instance, a scalar, linear, and additive-noise channel has

$$
Y_k = \sum_{j \neq k} G_{kj} X_j + Z_k
$$

(18)

where the $G_{kj}$ are $L \times L$ lower-triangular matrices and $Z_k = [Z_{k,1} \ldots Z_{k,L}]^T$, $k \in \mathcal{K}$. The noise $Z_k^c$ is independent of $A_k^c$. We write the covariance matrix of a random vector $X$ as $Q_X$ and its determinant as $|Q_X|$.

**IV. Cut-Set Bound**

We develop a cut-set bound for NiBMs that generalizes the classic cut-set bound for DMNs. Consider a set $\mathcal{S}$ of nodes and let $\mathcal{S}^c$ be the complement of $\mathcal{S}$ in $\mathcal{K} = \{1, 2, \ldots, K\}$. We say that $(\mathcal{S}, \mathcal{S}^c)$ is a cut separating a message $W_{km}$ and its estimate $\hat{W}_{km}$ if $k \in \mathcal{S}$ and $\ell \in \mathcal{S}^c$. Let $\mathcal{M}(\mathcal{S})$ be the set of indexes (which are integer pairs $km$) of those messages separated from one of their estimates by the cut $(\mathcal{S}, \mathcal{S}^c)$, and let $R_{\mathcal{M}(\mathcal{S})}$ be the sum of the rates of these messages.

There is a subtlety in that the NiBM can have high mutual information at the start of each block and low mutual information elsewhere. For example, consider a point-to-point channel (13) where $\mathbb{F}$ is the Galois field of size two, $K = 2$, $L = 2$, the channel matrix is

$$
G_{21} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}
$$

and $Z_2 = [0 0]^T$. We find that using the channel once gives larger mutual information per letter than using the channel twice or more. But this fact is not very interesting because we wish to transmit information reliably and can (usually) accomplish this only by using the channel often. To avoid such formal details, we will require that $n = mL$ for a positive integer $m$. Alternatively, we could require that $n$ be much larger than $L$. We have the following result that we prove in Appendix A.
Theorem 1: The capacity region $\mathcal{C}$ of a NiBM with block length $L$ that is used a multiple of $L$ times satisfies

$$\mathcal{C} \subseteq \bigcup_{P_{A_k^L} \in \mathcal{P}_{A_k^L}} \mathcal{R}(P_{A_k^L}, S)$$

where $\mathcal{R}(P_{A_k^L}, S)$ is the set of non-negative rate-tuples satisfying

$$R_{M(S)} \leq I(A_{S_k}; Y_{S_k}^L | A_{S_k}^L)/L.$$  \hspace{1cm} (20)

The joint probability distribution $P(a_{k}^L, x_{k}^L, y_{k}^L)$ factors as

$$P(a_{k}^L) \prod_{k=1}^{K} 1(x_{k}^L | a_{k}^L, 0y_{k}^{L-1}) P(y_{k}^L | x_{k}^L).$$  \hspace{1cm} (21)

Remark 4: The code functions in Theorem 1 are statistically dependent. This is different than in Sec. 4 where the code functions are independent (see Fig. 1 and (15)). Similarly, Shannon’s outer bound for the two-way channel [8, Eq. (36)] and the classic cut-set bound for DMNs [7, 10, Ch. 10, 14, p. 477] have statistically dependent inputs (see Sec. V-B).

Remark 5: The $1x_{k}^L | a_{k}^L, 0y_{k}^{L-1}$, $k = 1, 2, \ldots, K$, are fixed functions and $P(y_{k}^L | x_{k}^L)$ is fixed by the channel.

Remark 6: Remark 3 states that we may view the channel as being $P(y_{k}^L | a_{k}^L)$. This insight is useful for deriving achievable rates and for computing the cut-set bound (see [4, Sec. 5]). For instance, $I(A_{S_k}; Y_{S_k}^L | A_{S_k}^L)$ is concave in $P_{A_k^L}$. This result follows by the concavity of $I(A; B | C = c)$ in $P_{A|C=c}$ when $P_{B|A=c}$ is held fixed, and because $P(y_{k}^L | a_{k}^L)$ is fixed.

Remark 7: The $A_{k}^L$ are not “auxiliary” random variables, i.e., they are explicit components of the communication problem just like the channel inputs $X_{k}^L$. Moreover, the cardinalities $|A_{k}^L|$ are bounded by the channel alphabets (see (4)).

Remark 8: Average per-letter cost constraints can be dealt with in the usual way (see Remark 34 below). More precisely, if we have $J$ cost functions $s_j(\cdot)$ and constraints

$$\frac{1}{n} \sum_{i=1}^{n} E[s_j(X_{K,i}, Y_{K,i})] \leq S_j, \quad j = 1, 2, \ldots, J$$

then one may add the requirement that the union in (19) is over distributions (21) that satisfy

$$\frac{1}{L} \sum_{i=1}^{L} E[s_j(X_{K,i}, Y_{K,i})] \leq S_j, \quad j = 1, 2, \ldots, J.$$  \hspace{1cm} (23)

One may treat average per-block cost constraints similarly.

Remark 9: The bound (20) may be weakened as follows:

$$I(A_{S_k}; Y_{S_k}^L | A_{S_k}^L) \leq \sum_{i=1}^{L} H(Y_{S_k,i} | Y_{S_k,i}^{-1} A_{S_k}^L) - H(Y_{S_k,i} | Y_{S_k,i}^{-1} A_{S_k}^L)$$

$$= I(A_{S_k}^L \rightarrow Y_{S_k}^L | A_{S_k}^L)$$  \hspace{1cm} (24)

where $(a)$ follows by the chain rule for entropy and because

$$(A_{K,i+1}, \ldots, A_{K,i}, Y_{S_k,i}) \rightarrow A_{K,i}^L \rightarrow Y_{S_k}^L - Y_{S_k,i}$$  \hspace{1cm} (25)

forms a Markov chain. The bound (24) is further weakened by replacing code functions with channel inputs and outputs:

$$I(A_{S_k}^L \rightarrow Y_{S_k}^L | A_{S_k}^L) \leq \sum_{i=1}^{L} H(Y_{S_k,i} | Y_{S_k,i}^{-1} X_{K,i}^L) - H(Y_{S_k,i} | Y_{S_k,i}^{-1} X_{K,i}^L)$$

$$= I(X_{S_k}^L \rightarrow Y_{S_k}^L | X_{S_k}^L)$$  \hspace{1cm} (26)

However, the identity (27) may not be valid when considering dependent code functions such as in Theorem 1.

Remark 12: The cut-set bound with the normalized (26) in place of the right-hand side of (20) was derived in [15, Thm. 1] for causal relay networks and in [16, Thm. 1] for generalized networks. The authors of [15, 16] restrict attention to multiple unicast sessions as in [6, Sec. 15.10]. Theorem 1 improves these bounds and extends them to multiple multicast sessions. We discuss these bounds in more detail in Sec. VII-D.

Example 2: Consider additive noise channels with

$$Y_{k,i} = f_{k,i}(X_{k}^L) + Z_{k,i}$$

for $i = 1, 2, \ldots, L$, $k = 1, 2, \ldots, K$, where $Y_{k,i}$, $Z_{k,i}$ and $f_{k,i}(X_{k}^L)$ take on values in the field $F$. The noise variables $Z_{k,i}$ are independent of $A_{k}^L$. For finite fields, the bound (26) is

$$I(A_{S_k}^L; Y_{S_k}^L | A_{S_k}^L) \leq I(X_{S_k}^L, Y_{S_k}^L | X_{S_k}^L) = H(Y_{S_k}^L | X_{S_k}^L) - H(Z_{S_k}^L | 0Z_{S_k}^{L-1}).$$

Since $H(Z_{S_k}^L | 0Z_{S_k}^{L-1})$ is fixed by the channel, the cut-set bound with the normalized (29) in place of the right-hand side of (20) is a maximum (conditional) entropy problem.

Example 3: A special case of (28) is a deterministic NiBM for which the noise is a constant and

$$I(A_{S_k}^L; Y_{S_k}^L | A_{S_k}^L) \leq H(Y_{S_k}^L | X_{S_k}^L).$$  \hspace{1cm} (30)
B. DMNs

For $L = 1$ the NiBM is a DMN and Theorem 1 is the classic cut-set bound. Alternatively, we may view the DMN as a NiBM with block length $L$ and with

$$P(y^L_k|x^L_k) = \prod_{i=1}^{L} P_{X_i|X_{K,i}}(y_{K,i}|x_{K,i}).$$  \hspace{1cm} (31)

The weakened bound (26) becomes

$$I(X^L_k, 0Y^L_S \rightarrow Y^L_S \parallel X^L_S)$$

$$= \sum_{i=1}^{L} H(Y_{S^c,i}|X^L_S, Y^{i-1}_S) - H(Y_{S^c,i}|X^L_S)$$

$$\leq \sum_{i=1}^{L} I(X_{S^c,i}; Y_{S^c,i}|X^L_S).$$  \hspace{1cm} (32)

If we choose the code functions as codewords and

$$P(x^L_k) = \prod_{i=1}^{L} P_{X_i}(x_{K,i})$$  \hspace{1cm} (33)

then we achieve equality in (32). We recover the classic cut-set bound by choosing $P(x_{K,i}) = P_{X_i}(x_{K,i})$ for all $i$.

**Remark 13:** Consider a DMN that is time varying in blocks of length $L$, i.e., we have a NiBM of length $L$ and

$$P(y^L_k|x^L_k) = \prod_{i=1}^{L} P_{X_i|X_{K,i}}(y_{K,i}|x_{K,i})$$  \hspace{1cm} (34)

The cut-set bound of Theorem 1 may now be computed with independent inputs as in (33).

V. POINT-TO-POINT CHANNELS

Consider a point-to-point channel with input $X^L$ taking on values in $X^L$, receive output $Y^L$, feedback $\tilde{Y}^L$ taking on values in $\tilde{Y}^L$, and feedback $\tilde{Y}^L$ taking on values in $\tilde{Y}^L$. A FDG for $L = 2$ and $n = 4$ channel uses is shown in Fig. 2.

**Theorem 2:** The capacity of a point-to-point channel with iBM and block length $L$ is

$$C = \max_{P_{A^L}} I(A^L; Y^L) / L$$  \hspace{1cm} (35)

where $P(a^L, y^L, \tilde{y}^L)$ factors as

$$P(a^L)I(x^L\parallel a^L, 0y^{L-1})P(y^L, \tilde{y}^L|x^L).$$  \hspace{1cm} (36)

**Proof:** Achievability follows by random coding with a maximizing $P_{A^L}$. For example, one may use the steps outlined in [2], Sec. VII.B). The converse follows by Theorem 1. \hfill \blacksquare

**Remark 14:** The distribution (36) gives

$$I(A^L; Y^L) = I(A^L \rightarrow Y^L).$$  \hspace{1cm} (37)

**Remark 15:** The feedback $\tilde{Y}^L$ can be noisy.

**Remark 16:** In-block feedback can increase $C$ but across-block feedback does not increase $C$. This statement refines Shannon’s classic theorem on feedback capacity [17, Thm. 6]. For instance, in Fig. 2 we can remove the dashed lines across blocks without changing $C$.

**Remark 17:** If $\tilde{Y}^L$ is a constant then there is no feedback and we have

$$I(A^L; Y^L) = I(X^L; Y^L) = I(X^L \rightarrow Y^L).$$  \hspace{1cm} (38)

The corresponding capacity result is not new, however, since the model is a special case of a point-to-point channel with vector alphabets.

**Remark 18:** $I(A^L; Y^L)$ is concave in $P_{A^L}$ and the Arimoto-Blahut algorithm [18], [19] can perform the maximization (35).

The cardinality $|A^L|$ is bounded by the channel alphabets (see (2) and Remark 7) and we have

$$|A^L| = \prod_{i=1}^{L} |X_i|^{5^{i-1}}.$$  \hspace{1cm} (39)

The identity (39) means that $|A^L|$ is double exponential in $L$ if the alphabet sizes are similar for all $i$. However, we prove the following theorem by using classic results [20] p. 96], [21] p. 310] on bounding set sizes.

**Theorem 3:** The maximum in Theorem 2 is achieved by a $P_{A^L}$ for which $|\text{supp}(P_{A^L})|$ is at most

$$\min \left( |Y^L|, |X_i| + \sum_{i=2}^{L} |X^{i-1}| \cdot |\tilde{Y}^{i-1}| \cdot (|X_i| - 1) \right).$$  \hspace{1cm} (40)

**Proof:** See Appendix B. \hfill \blacksquare

**Remark 19:** Theorem 3 states that $|\text{supp}(P_{A^L})|$ can be exponential, and not double exponential, in $L$. Of course, one must still determine $\text{supp}(P_{A^L})$ which can be a high-complexity search problem for even small $L$.

**Example 4:** Consider a binary-alphabet channel with $L = 2$ and

$$Y_1 = X_1, \quad \tilde{Y}_1 = Z, \quad Y_2 = X_2 \oplus Z$$  \hspace{1cm} (41)

where the bit $Z$ has $P(Z) = e$. This is an additive noise channel of the form (28) whose capacity without feedback is achieved by uniformly-distributed $X^2$ so that

$$I(X^2; Y^2) = 2 - H_2(e).$$  \hspace{1cm} (42)
To compute the feedback capacity, consider the simple bound
\[ I(A^2;Y^2) = H(Y^2) - H(Y^2|A^2) \leq 2 \] (43)
and observe that we achieve equality in (43) with \( X_2 = X_2^t \oplus Z \) where \( X_2^t \) is independent of \( X_1 \), and where \( X_1 \) and \( X_2 \) are uniformly distributed bits. Feedback thus enlarges the capacity.

We translate this strategy into a code function (here a code tree) distribution. We label \( A^2 \) as \( b_b b_1 \) by which we mean that \( X_1 = b, X_2 = b_1 \) if \( Y_1 = 0 \), and \( X_2 = b_1 \) if \( Y_1 = 1 \). We choose
\[
\begin{align*}
P_{A^2}(0,0) &= P_{A^2}(0,1) = P_{A^2}(1,0) = P_{A^2}(1,1) = 0 \\
P_{A^2}(0,0) &= P_{A^2}(0,1) = P_{A^2}(1,0) = P_{A^2}(1,1) = 1/4
\end{align*}
\]
and achieve capacity with four code trees, as predicted by Theorem [3].

**Example 5:** We demonstrate the deficiencies of the weakened bound based on (26). Suppose the channel is
\[
Y_1 = X_1 + Z_1 + Z_2, \quad Y_2 = Z_2 \tag{44}
\]
where \( Z_1 \) and \( Z_2 \) are independent with \( P_{Z_1}(1) = \epsilon_1 \) and \( P_{Z_2}(1) = \epsilon_2 \). We achieve the capacity
\[
C = ((1 - H_z(\epsilon_1))/2 \tag{45}
\]
by having the receiver compute \( Y_1 \oplus Y_2 = X_1 \oplus Z_1 \). In fact, we can achieve capacity by not using the feedback.

For the weakened bound (29), observe that (44) has the form (28). Defining \( \epsilon_1 + \epsilon_2 = \epsilon_1 (1 - \epsilon_2) + (1 - \epsilon_1) \epsilon_2 \) and \( \epsilon = \{1\} \) we compute (see (29))
\[
\begin{align*}
H(Z_{S}^{L}, ||Z_{S}^{L-1}| = H(Z_{1} \oplus Z_{2}) + H(Z_{2}|Z_{1} \oplus Z_{2}, Z_{1})
= H_{2}(\epsilon_{1} + \epsilon_{2}).
\end{align*}
\tag{46}
\]
The weakened bound based on (29) is therefore
\[
2C \leq \max_{P_{X}^{2}} H(Y^2) - H_{2}(\epsilon_{1} + \epsilon_{2})
= 1 + H_{2}(\epsilon_{2}) - H_{2}(\epsilon_{1} + \epsilon_{2}) \tag{47}
\]
with equality if \( X_1 \) is uniform. This bound is loose in general, e.g., if \( \epsilon_1 = 1/2 \) then \( C = 0 \) but (47) gives \( C \leq H_{2}(\epsilon_2)/2 \).

**A. Noise-Free Feedback**

The feedback is noise-free if \( \bar{Y}^L \) is a causal function of \( X^L \) and \( Y^L \), i.e., if \( \bar{Y}_i = \hat{f}(X_i, Y_i) \) for \( i = 1, 2, \ldots, L \). The receiver can therefore track, or observe, the choice of \( X^L \) for each tree \( A^L \). The expression (35) simplifies to
\[
C = \max_{P_{X}^{L}} I(X^L \rightarrow Y^L)/L. \tag{48}
\]

**Example 6:** Consider the additive noise channel (18) with \( Y = GX + Z \) and noise-free feedback. We have
\[
I(X^L \rightarrow Y^L) = H(Y^L) - H(Z^L) \tag{49}
\]
so that computing (48) reduces to maximizing \( H(Y^L) \).

**Remark 20:** As in (49), one is sometimes interested in maximizing the output entropy \( H(Y^L) \). We observe that for noisy or noise-free feedback we have
\[
\max_{P_{A^L}} H(Y^L) = \max_{P_{X^L} \mid_{o Y^L}} H(Y^L). \tag{50}
\]

**B. Block Fading Channels**

Channels with block fading [11] or block interference [22] have a state \( S \) that is memoryless across blocks of length \( L \) and whose realization \( S = s \) specifies the memoryless channel in each block. In other words, when \( S = s \) we have
\[
P(y^L, \tilde{y}^L|x^L, s) = \prod_{i=1}^{L} P_{Y|X}(y_i, \tilde{y}_i|x_i, s). \tag{51}
\]
We may view such channels as NiBMs for which \( Z = SN^L \) in (3), i.e., \( Z \) includes the state \( S \) and a noise string \( N^L \) where the \( N_i, i = 1, 2, \ldots, L \), are statistically independent and identically distributed. Equation (4) thus becomes
\[
Y_i = f_{i(i+1)}(X_i, S_{i/L}|N_i) \tag{52}
\]
\[
Y_i = \hat{f}_{i(i+1)}(X_i, S_{i/L}|N_i) \tag{53}
\]
for \( i = 1, 2, \ldots, n \).

**C. Channels with State Known Causally at the Encoder**

Shannon’s channel with state known causally at the encoder [14] is a point-to-point channel with input and output sequences \( X^n \) and \( Y^n \), respectively, and where a state sequence \( S^n \) is revealed causally to the encoder in the sense that \( X_i \) can be a function of \( W \) and \( S_i, i = 1, 2, \ldots, n \). The \( S_i, i = 1, 2, \ldots, n \), are statistically independent realizations of a state random variable \( S \). The channel outputs are
\[
Y_i = f(X_i, S_i, Z_i) \tag{54}
\]
for some function \( f(\cdot) \) where the \( Z_i, i = 1, 2, \ldots, n \), are statistically independent realizations of a noise random variable \( Z \). The FDG is shown in Fig. [3].

The channel is usually considered memoryless. However, an alternative and insightful interpretation is that this channel has iBM and block length \( L = 2 \). To see this, observe that Fig. [3] is a subgraph of Fig. [2] up to relabeling the nodes. In other words, in Fig. [3] we choose \( X_1 = Y_1 = Y_2 = \{0\} \) and \( \tilde{Y}_1 = S \). Observe that the “feedback” \( S \) can be noisy in the sense of Sec. V-A. For the FDG in Fig. [3] we have renamed \( A_2, \tilde{Y}_1, X_2, Y_2 \) as \( A_1, S_1, X_1, Y_1 \), respectively, so that the subscripts enumerate the block. The same random variables without the block indices are the respective \( A, S, X, Y \). The channel functions \( A \) for this type of problem are sometimes called Shannon strategies [14] p. 176].

The capacity is given by Theorem [2] which here is
\[
2C = \max_{P_{A}} I(A;Y). \tag{55}
\]

The alphabet size of \( A \) is \(|X|^{|S|} \) but (40) tells us that
\[
|\text{supp}(P_{A})| \leq \min(|X|, 1 + |S|) \cdot (|X| - 1) \tag{56}
\]
suffices. The \(|Y| \) bound is due to Shannon [2] and the second bound was reported in [23] Thm. 1 (see also [14] p. 177).
X = b_1 if S = 1. The capacity turns out to be 2C = 1 bit and is achieved with
\[ P_A(00) = P_A(11) = 0 \]
\[ P_A(01) = P_A(10) = 1/2. \]
We thus require at most three code trees, as predicted by (56).
Moreover, the weakened bound based on (55) gives
\[ 2C \leq \max_{P_{X|b}} I(X;S;Y) = \log_2(3) \text{ bits.} \]
A better upper bound follows by giving S to the receiver to obtain
\[ 2C \leq \max_{P_{X|b}} I(X;Y|S) = 1 \text{ bit.} \]

**Remark 21:** The above construction extends in an obvious way to show that any DMN with state(s) known causally at the encoder(s) is effectively a NiBM with block length L = 2. The cut-set bound (19) thus applies to these problems.

### D. Channels with Action-Dependent State

Weissman’s channel with action-dependent state modifies Shannon’s model and lets the transmitter influence the state \( \{B, S, X, Y\} \) at time \( i \). In other words, at time \( i \) the transmitter can choose a letter \( B_i \) as a function of \( W \) and \( S_{i-1} \) and the next state is
\[ S_i = g(B_i, Z_i) \]
for some function \( g(\cdot) \). The FDG is shown in Fig. 4. Observe that \( Z_i \) could be a random vector so that the noise influencing \( S_i \) and \( Y_i \) is statistically independent.

This channel is again usually considered memoryless. However, we interpret the channel as having iBM and block length \( L = 2 \), since Fig. 4 is a subgraph of Fig. 2 up to relabeling the nodes. More precisely, in Fig. 2 we choose \( Y_1 = Y_2 = \{0\} \) and \( \tilde{Y}_1 = \tilde{Y}_2 = S \). For the FDG in Fig. 4 we have renamed \( X_1, Y_1, X_2, Y_2 \) as \( B_1, S_1, X_1, Y_1 \), respectively, so that the subscripts enumerate the block. The same random variables without the block indices are the respective \( B, S, X, Y \). Theorem 2 gives the capacity
\[ 2C = \max_{P_{X}} I(A;Y) = \max_{P_{BA}} I(BA;Y) \]
and Theorem 3 gives
\[ |\text{supp}(P_{BA})| \leq \min(|Y|, |S| + |B| (|X| - 1)). \]

**Remark 22:** The expression (61) is the same as in [24] Thm. 2] because \( U \) plays the role of \( BA_2 \).

**Remark 23:** The constraint (62) is slightly stronger than that in [24] Thm. 2] because \( Z \) may influence both \( S \) and \( Y \). However, the associations described in [24] p. 5405] show that the original model includes more problems than apparent at first glance (see also comments in [24] Sec. VII).

**Remark 25:** The model in Fig. 4] may seem different than in [24] because \( S \) may influence future actions as well as the present and future \( X \). However, across-block feedback does not increase capacity (see Remark 16 so we may remove the \( S \)-to-\( B \) functional dependence without affecting capacity (see also comments in [24] Sec. VII concerning feedback).

**Remark 26:** We may add functional dependence from \( B \) to \( Y \) without changing the capacity expression. Similar comments are made in [24] p. 5398 and Sec. VII.

**Example 8:** Consider a channel with a rewrite option [24] Sec. V.A] which means that the \( B \)-to-\( S \) and \( X \)-to-\( Y \) channels are effectively the same. At time \( i = 1 \) the encoder “writes” on the \( B \)-to-\( S \) channel. At time \( i = 2 \), if the encoder is happy with the outcome \( S \) then it sends a no-rewrite symbol \( N \) which means that \( Y = S \). But if the encoder is unhappy with \( S \) then it “rewrites” a symbol on the \( X \)-to-\( Y \) channel.

We have \( X = B \cup \{N\}, S = Y \), and the bound (62) is \( |\text{supp}(P_{BA})| \leq |Y| \). For example, suppose the \( B \)-to-\( S \) channel is a binary symmetric channel (BSC) with crossover probability \( \delta \), \( 0 \leq \delta \leq 1/2 \) (see [24]). We label \( BA_2 \) as \( b, b_0, b_1 \) by which we mean that \( B = b, X = b_0 \) if \( S = 0 \), and \( X = b_1 \) if \( S = 1 \). We have \( |Y| = 2 \) and achieve \( C = I(BA;Y) = 1 - H_2(\delta^2) \) by choosing
\[ P_{BA_2}(0, N_0) = P_{BA_2}(1, 1 N) = 1/2. \]
We require only two code trees, as predicted by (62).

**Remark 27:** Multiple rewrites are modeled by increasing \( L \).
VI. MULTIUSER CHANNELS

A. Multiaccess Channels

Consider a two-user (three-terminal) MAC with iBM and with inputs \(X_1^L, X_2^L\), and outputs \(Y_1^L, Y_2^L\). Node 3 is the receiver and the variables \(X_3, Y_3\) should be considered constants. The FDG for \(L = 2\) and \(n = 4\) is the same as Fig. 1 except that the variables \(X_3, i = 1, 2, 3, 4\), are missing in Fig. 1. The cut-set bound of Theorem 3 is

\[
\{ (R_1, R_2) : 
\begin{align*}
0 \leq R_1, 0 \leq R_2 \\
R_1 \leq I(A_1^L; Y_2^L | X_1^L)/L \\
R_2 \leq I(A_2^L; Y_2^L | X_1^L)/L \\
R_1 + R_2 \leq I(A_1^L A_2^L; Y_3^L)/L
\end{align*}
\}
\]

(63)

If there is no feedback, then \(Y_1^L\) and \(Y_2^L\) can be considered constants. The resulting cut-set bound can be strengthened in the usual way to become

\[
\{ (R_1, R_2) : 
\begin{align*}
0 \leq R_1, 0 \leq R_2 \\
R_1 \leq I(X_1^L Y_2^L X_1^L T)/L \\
R_2 \leq I(X_2^L Y_2^L X_1^L T)/L \\
R_1 + R_2 \leq I(X_1^L X_2^L Y_3^L | T)/L
\end{align*}
\}
\]

(64)

where the union is over distributions such that \(X_1^L - T - X_2^L\) forms a Markov chain (\(T\) is the usual time-sharing random variable). This modified cut-set bound is the capacity region without feedback. The result is not new, however, since the model is a special case of a classic MAC with vector alphabets.

Remark 28: MACs with state known causally at the encoders were treated in [25, Sec. IV]. As pointed out in Remark 21 such channels are NiBM with block length \(L = 2\). For example, the outer bound of Theorem 3 in [25, Sec. IV] is the same as the cut-set bound of Theorem 3.

B. Multiaccess Channels with Feedback

Several capacity results for DMNs generalize to problems with iBM. For example, consider Willems’ result [26] that the Cover-Leung region [27] is a channel for full feedback \((Y_1 = Y_2 = Y_3 = Y)\) and where one channel input, say \(X_1\), is a function of \(Y\) and \(X_2\). A natural generalization to MACs with iBM is to consider full feedback \((Y_{1,i} = Y_{2,i} = Y_{3,i} = Y_i)\) and require \(X_{1,i} = f_i(X_{2,i}, Y_i)\) for \(i = 1, 2, \ldots, L\). A MAC of this type is the binary adder channel (BAC) with \(\{0, 1\}\) input alphabets and the integer-addition output

\[
Y = G_1 X_1 + G_2 X_2
\]

(65)

where \(G_1\) and \(G_2\) are lower-triangular matrices with \(\{0, 1\}\) entries, and where \(G_1\) has ones on the diagonal.

Theorem 4: The capacity region of a MAC with iBM and full feedback and where \(X_{1,i} = f_i(X_{2,i}, Y_i)\) for all \(i\) is

\[
\{ (R_1, R_2) : 
\begin{align*}
0 \leq R_1, 0 \leq R_2 \\
R_1 \leq I(A_1^L; Y_2^L | A_1^L V)/L \\
R_2 \leq I(A_2^L; Y_2^L | A_1^L V)/L \\
R_1 + R_2 \leq I(A_1^L A_2^L; Y_2^L)/L
\end{align*}
\}
\]

(66)

where the union is over distributions that factor as

\[
P(v) \prod_{k=1}^{2} P(a_k^L | v) I(x_k^L | a_k^L, 0y^{L-1}) P(y^L | x_1^L, x_2^L).
\]

(67)

A cardinality bound on \(V\) is \(|V| \leq |Y^L| + 2\).

Proof: The proof mimics that in [26] and is given in Appendix 4.

Proposition 1: An alternative way of writing (66) is

\[
\{ (R_1, R_2) : 
\begin{align*}
0 \leq R_1, 0 \leq R_2 \\
R_1 \leq I(X_1^L Y_2^L | X_2^L V)/L \\
R_2 \leq I(X_2^L Y_2^L | X_1^L V)/L \\
R_1 + R_2 \leq I(X_1^L X_2^L Y_2^L)/L
\end{align*}
\}
\]

(68)

where the union is over distributions that factor as

\[
P(v) \prod_{k=1}^{2} P(x_k^L | 0y^{L-1} | v) P(y^L | x_1^L, x_2^L).
\]

(69)

Note that one conditions on \(V\) for all times.

Proof: Consider the distribution (67). The chains

\[
A_1^L - V X_2^L Y_i - 1 - Y_i
\]

(70)

\[
A_2^L - V X_1^L Y_i - 1 - Y_i
\]

(71)

\[
A_1^L A_2^L - V X_2^L Y_i - 1 - Y_i
\]

(72)

are Markov such that

\[
I(A_1^L; Y^L | A_2^L V)
\]

\[
= \sum_{i=1}^{L} H(Y_i | A_2^L Y_i - 1 X_2^L V) - H(Y_i | A_2^L A_1^L Y_i - 1 X_1^L X_2^L V)
\]

(73)

and similarly

\[
I(A_1^L; Y^L | A_1^L A_2^L V) = I(X_2^L Y^L | X_1^L V).
\]

(74)

\[
I(A_1^L A_2^L; Y^L) = I(X_1^L X_2^L Y^L).
\]

(75)

The distribution (69) follows from (67).

C. Broadcast Channels

Consider a two-user (three terminal) BC with iBM. We label the transmitter inputs and outputs as \(X^L\) and \(Y^L\), respectively, and the receiver outputs as \(Y_1^L\) and \(Y_2^L\). Suppose there are only dedicated messages and no common message. The cut-set bound of Theorem 1 is

\[
\{ (R_1, R_2) : 
\begin{align*}
0 \leq R_1, 0 \leq R_2 \\
R_1 \leq I(A_1^L; Y_2^L)/L \\
R_2 \leq I(A_2^L; Y_2^L)/L \\
R_1 + R_2 \leq I(A_1^L A_2^L; Y_2^L)/L
\end{align*}
\}
\]

(76)

An achievable region follows by extending Marton’s region as in [27, Lemma 2]: the non-negative rate pair \((R_1, R_2)\) is achievable if it satisfies

\[
L R_1 \leq I(T U_1; Y_1^L)
\]

(77)

\[
L R_2 \leq I(T U_2; Y_2^L)
\]

\[
L(R_1 + R_2) \leq \min \{ I(T; Y_1^L), I(T; Y_2^L) \}
\]

(78)

\[
+ I(U_1; Y_1^L | T) + I(U_2; Y_2^L | T) - I(U_1; U_2 | T)
\]

for some auxiliary random variables \(TU_1 U_2\) for which the joint distribution of the random variables factors as

\[
P(t, u_1, u_2)1(x^L | 0y^{L-1} | t, u_1, u_2) P(y_1^L, y_2^L | x^L).
\]

Marton’s region is known to be the same as (76) for \(L = 1\) and deterministic broadcast channels. For \(L > 1\), suppose that
\(Y_{1,i}\) and \(Y_{2,i}\) are functions of \(X^i\) for all \(i\). We may choose \(T = 0\), \(U_1 = X_i^L\), and \(U_2 = Y_i^L\) without violating the Markov condition (78) and achieve

\[
\mathcal{C} = \left\{ \begin{array}{c}
0 \leq R_1 \leq H(Y_i^L)/L \\
0 \leq R_2 \leq H(Y_i^L)/L \\
R_1 + R_2 \leq H(Y_i^L Y_i^L)/L
\end{array} \right\}.
\] (79)

The cut-set region (76) is the same as (79), and therefore (79) is \(\mathcal{C}\). In fact, feedback does not increase capacity because the transmitter knows, and controls, the channel outputs.

\textbf{Remark 29:} The capacity region of a physically degraded BC with two receivers and state known causally at the encoder was derived in [25, Sec. II]. Such channels are NiBMs with block length \(L = 2\), see Remark [24]. The cut-set bound of Theorem [1] is loose but the capacity region is achieved by using the coding method described above. In particular, we choose \(U_2\) in (77) to be a constant and recover the achievability part of Theorem 1 of [25, Sec. II].

\section{Interference Channels}

The cut-set bound is often not so interesting for BCs or interference channels (ICs) with \(L = 1\) because better capacity bounds exist. The same will be true for \(L > 1\). On the other hand, studying extensions of existing bounds and achievable regions is interesting, e.g., extensions of the Han-Kobayashi region [28] to \(L > 1\). It may also be interesting to study interference alignment [29], [30] and interference focusing [31] for NiBMs.

\section{Relay Networks}

Causal relay networks [13] and generalized networks [14] effectively extend relay networks with delays [3] in the sense that for every relay network with delays there is a causal relay network having the same capacity region. Furthermore, causal relay networks and generalized networks are special NiBMs. This section focuses on relay networks with iBM and applies Theorem [1] to this class of problems.

\subsection{Relay Channels}

Consider a three-node relay channel (RC) with iBM and source inputs \(X_i^L\), relay inputs \(X_i^L\) and outputs \(Y_i^L\), and destination outputs \(Y_3^L\). The RC is a special case of the MAC in Sec. VI-A where node 2 (the relay) has no message and node 1 (the source) has no feedback. An FDG for \(L = 2\) and \(n = 4\) is shown in Fig. 5. The cut-set bound of Theorem [1] is

\[
\mathcal{C} = \max \min \left\{ I(X_i^L; Y_i^L Y_3^L | A_i^L), I(X_i^L A_i^L; Y_3^L | Y_i^L) \right\}.
\] (80)

where the maximization is over \(P_{X_i^L A_i^L}\).

We list several classic coding strategies [32], [33]. The achievable rates follow by standard random coding arguments (see [2] Sec. VII).

- **Decode-forward (DF)** achieves rates \(R\) satisfying

\[
R = \max \min \left\{ I(X_i^L; Y_i^L | A_i^L), I(X_i^L A_i^L; Y_3^L) \right\}
\] (81)

where the maximization is over \(P_{X_i^L A_i^L}\) and where the joint distribution factors as

\[
P(x_i^L, a_i^L) 1(x_i^L \parallel a_i^L, 0 y_i^{L-1}) P(y_i^L, y_i^L \parallel x_i^L, x_2^L).
\] (82)

- **Partial decode-forward (PDF)** achieves \(R\) satisfying

\[
R = \max \min \left\{ I(U; Y_2^L | A_i^L), I(X_i^L Y_3^L U), I(X_i^L A_i^L, Y_3^L) \right\}
\] (83)

where the maximization is over \(P_{U X_i^L A_i^L}\) and where the joint distribution factors as

\[
P(u, x_i^L, a_i^L) 1(x_i^L \parallel a_i^L, 0 y_i^{L-1}) P(y_i^L, y_i^L \parallel x_i^L, x_2^L).
\] (84)

\textbf{The rate (83) generalizes [3, Prop. 5].}

- **Compress-forward (CF)** achieves \(R\) satisfying

\[
\begin{aligned}
R &= \max \min \left( I(X_i^L; Y_2^L Y_3^L | A_i^L T), \\
&\quad I(X_i^L A_i^L; Y_3^L | T) - I(Y_i^L; Y_2^L | X_i^L A_i^L Y_3^L T) \right)
\end{aligned}
\] (85)

where the maximization is over joint distributions that factor as

\[
P(t) P(x_i^L | t) P(a_i^L | t) 1(x_i^L \parallel a_i^L, 0 y_i^{L-1}) \\
\cdot P(y_i^L, y_i^L \parallel x_i^L, x_2^L).
\] (86)

\textbf{Example 9:} Remark [3] states that we can view the channel as being \(P(y_i^L, y_i^L | x_i^L, a_i^L)\). The RC is physically degraded if \(X_i^L - A_i^L Y_i^L - Y_2^L\) forms a Markov chain so that \(I(X_i^L; Y_3^L | A_i^L Y_2^L) = 0\). The DF rate (81) thus matches (80). This capacity result generalizes [3, Prop. 6].

\textbf{Example 10:} The RC is reversely physically degraded if \(X_i^L - A_i^L Y_i^L - Y_2^L\) forms a Markov chain so that \(I(X_i^L; Y_3^L | A_i^L Y_2^L) = 0\). The cut-set bound (80) reduces to

\[
\mathcal{C} = \max \min \mathcal{C}(X_i^L; Y_3^L | A_i^L Y_2^L = a_i^L).
\] (89)

The rate on the right-hand side of (89) is achieved by random coding with \(A_i^L = \Phi_i^L\).

\textbf{Remark 30:} Physically degraded RCs with state known causally at the encoder are treated in [25, Sec. III]. Such channels are NiBMs with block length \(L = 2\), see Remark [21] and Theorem [1] gives the converse for [25, Thm. 2]. However, these channels are not treated in this section because the source node receives the channel state as “feedback”.

\textbf{Example 11:} Suppose the RC is semi-deterministic in the sense that \(Y_2,i = f_i(X_1,X_3)\) for \(i = 1, 2, \ldots, L\). We may choose \(U = Y_2^L\) and (83) becomes the cut-set bound (80). This capacity result generalizes [3, Prop. 7].

\textbf{Example 12:} Suppose the RC is semi-deterministic in the (more general) sense that \(Y_2,i = f_i(X_1,X_3)\) for \(i = 1, 2, \ldots, L\). Consider (85) for which we have

\[
I(Y_2^L; Y_3^L | X_1^L A_2^L Y_3^L T) = 0.
\] (90)

We choose \(T\) as a constant and \(Y_2,i = Y_2^L\) so that (85) is the right-hand side of (80) but with independent \(X_i^L\) and \(A_i^L\).

\textbf{Example 13:} A special case of Example [2] is where \(Y_2,i = f_i(X_1, X_3)\) and there is a separate channel with iBM and capacity \(R_0\) from the relay to the destination (see [34]). The best \(X_i^L\) and \(A_i^L\) are independent so the choice \(Y_2,i = Y_2^L\) lets CF achieve the cut-set bound (80).
\[ L = 2 \]

we can model the channel as a RC with iBM and block length \( L \).

channel is therefore

and the FDG for two channel uses is shown in Fig. 6.

and output \( Y_2 \), and destination output \( Y_3 \). The channel is

\[ P(y_2|x_1) \cdot P(y_3|x_1, x_2, y_2) \]  \hspace{1cm} (91)

and the FDG for two channel uses is shown in Fig. 6.

This channel is usually considered memoryless. However, we can model the channel as a RC with iBM and block length \( L = 2 \) and where \( \mathcal{X}_{2,1} = \mathcal{Y}_{3,1} = \mathcal{Y}_{2,2} = \mathcal{X}_{1,2} = \emptyset \). The channel is therefore

\[ P(y_2^2, y_3^2|x_1^2, x_2^2) = P(y_2, y_2, x_1, x_1) \cdot P(y_3, y_3, x_2, x_2, y_2) \]  \hspace{1cm} (92)

as long as \( x_{2,1} = y_{3,1} = y_{2,2} = x_{1,2} = 0 \). Note that every node has at most one channel input and output in each block. We can thus remove the time indices and (92) becomes (91). Observe that Fig. 6 is a subgraph of Fig. 5 up to relabeling the nodes.

We apply the cut-set bound (80) and remove the time indices to obtain

\[ 2C \leq \max \{ I(X_1; Y_2 Y_3|A_2), I(X_1 A_2; Y_3) \} \]  \hspace{1cm} (93)

where the maximization in (93) is over \( P_{X_1 A_2} \) and \( |A_2| = |\mathcal{X}_2| |\mathcal{Y}_2| \). In fact, (93) combined with this cardinality constraint is attributed to Willems’ in [3, p. 3419]. We show in Appendix C that one can choose

\[ |\text{supp}(P_{A_2})| \leq \min(|\mathcal{Y}_3| + 1, |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1) \]  \hspace{1cm} (94)

Remark 31: The cut-set bound in [3 Thm. 2] is the same as (93) except that the maximization is different. The bound of [3 Thm. 2] requires

\[ X_2 = f(A_2, Y_2) \]  \hspace{1cm} (95)

for some function \( f(\cdot) \) and one optimizes over all \( f(\cdot) \) and \( P_{X_1 A_2} \) such that \( |A_2| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1 \).

We claim that the formulation (93) combined with (94)

\[ \sup \{ |\mathcal{X}_2| \cdot |\mathcal{Y}_2| \} \leq |\mathcal{X}_2| \cdot |\mathcal{Y}_2| \]  \hspace{1cm} (96)

out of \(|\mathcal{X}_2| \cdot |\mathcal{Y}_2|\) code functions. We must therefore perform at most

\[ \frac{|\mathcal{X}_2| \cdot |\mathcal{Y}_2|}{N_A} \]  \hspace{1cm} (97)

optimizations in \(|\mathcal{X}_1| \cdot N_A - 1 \) dimensions. In contrast, (93) and (95) require optimizing \( P_{X_1 A_2} \) for \(|\mathcal{X}_2| \cdot |\mathcal{Y}_2|\) functions \( f(\cdot) : \mathcal{A}_2 \times \mathcal{Y}_2 \to \mathcal{X}_2 \) where \(|\mathcal{A}_2| \) is at most

\[ N_V = |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1 \]  \hspace{1cm} (98)

We thus have at most \(|\mathcal{X}_2|^{N_V} \cdot |\mathcal{Y}_2|\) optimizations in \(|\mathcal{X}_1| \cdot N_V - 1 \) dimensions. But we have \( N_A \leq N_V \) and

\[ \left( \frac{|\mathcal{X}_2| \cdot |\mathcal{Y}_2|}{N_A} \right) \leq \frac{|\mathcal{X}_2| \cdot |\mathcal{Y}_2|}{N_V} \leq \frac{|\mathcal{X}_2|^{N_A} \cdot |\mathcal{Y}_2|}{N_V} \]  \hspace{1cm} (99)

so the optimization of (93) and (95) is generally simpler than the optimization of (93) and (95). This discussion shows that
one may as well consider code functions directly rather than introducing auxiliary random variables and auxiliary functions.

Example 14: Suppose that $|X_1| = |X_2| = 2$ and $|Y_2| = 4$. Then [22] states that at most 5 code functions (here code trees) out of 16 need have positive probability. Our search is thus over $\binom{16}{5} = 4368$ combinations of code trees. In comparison, [3] Thm. 2] requires a search over $2^{20} \approx 10^6$ mappings $f(\cdot)$.

C. Relay Networks with Delays

Relay networks with delays [3] have the simplifying feature that every node has at most one channel input and output in each block. Furthermore, there is exactly one network message that originates at a designated source node $k = 1$ and that is destined for a designated node $K = K$. Nodes 1 and $K$ have no channel outputs and inputs, respectively, i.e., we effectively have $Y_{1,i} = X_{K,i} = 0$ for all $i$.

A cut bound for such networks was developed in [3] Thm. 4] that is almost the same as Theorem [1]. The difference between the bounds is similar to the difference described in Remark [3] above, i.e., [3] Thm. 4 uses auxiliary variables for the code functions (in this case Shannon strategies) and specifies cardinality bounds on these variables. Theorem [1] instead uses the code functions directly, and these functions have finite cardinality if the channel input and output alphabets are finite (see Remark [7]). One may develop improved cardinality bounds as in [3] Thm. 4] that are useful if the channel input or output alphabets are continuous.

D. Causal Relay Networks and Generalized Networks

Causal relay networks [15] and generalized networks [16] are NiBMs that extend relay networks with delays by considering more than one unicast session. We describe these networks by using an example with $K = 5$ nodes whose FDG for one block is shown in Fig. 7. We show in Example 15 below that if the causal relays have no messages then Theorem [1] can be strictly better than [15, Thm. 2] due to inequality (20). We show in Example [15] below that if the causal relays have no messages then Theorem [1] can be strictly better than [15, Thm. 2] due to inequality (20). Furthermore, the auxiliary random variables $U_k$ in [15] Thm. 2] are not specified to be code functions. The optimization is thus more complex than by using Theorem [1] in general (see Remark [31]).

Example 15: Consider Fig. 7 with $X_k = Y_k = \{0\}$ for $k = 2, 4$, i.e., nodes 2 and 4 are removed from the problem. Consider $Y_3 = [X_1, Z]$ where $X_1 = \{0, 1\}$ and $P_Z(0) = P_Z(1) = 1/2$, and $Y_5 = Z$. Suppose there is only one message with rate $R_{15}$ at node 1 destined for node 5 (so the causal

Fig. 7. FDG for a causal relay network with $K = 5$ nodes and $n = 3$ channel uses. The network is a NiBM with block length $L = 3$.
relays at nodes 3 and 5 have no messages). We effectively have a RC with no delay and the capacity is zero because \( X_{1,A3} \) has no influence on \( Y_5 \). For instance, the cut-set bound \((20)\) with \( S = \{1,3\} \) gives \( 3R_{15} \leq I(X_1|A_3; Y_5|A_5) = 0 \).

Next, consider the cut-set bound of \((15)\) Thm. 2. There are two cuts to consider without nodes 2 and 4. The cut \( S = \{1,3\} \) gives (see \((101)\) after step (c))

\[
3R_{15} \leq I(X_1X_2Y_3;Y_5|A_5) = 1
\]

and the cut \( S = \{1\} \) gives

\[
3R_{15} \leq I(X_1;Y_3|A_3A_5) = H(X_1|A_3A_5).
\]

But we have \( H(X_1|A_3A_5) = 1 \) by choosing \( X_1 \) independent of \( A_3A_5 \) and \( P_{X_1}(0) = P_{X_1}(1) = 1/2 \). Thus, the cut-set bound of \((15)\) Thm. 2 is loose while Theorem 1 is tight.

**Example 16:** Consider the generalized network called a “BSC with correlated feedback” in \((16)\) Sec. VI. This network is a two-way channel with iBM and block length \( L = 2 \) and with binary inputs and outputs

\[
\begin{align*}
Y_{2,1} &= X_{1,1} + Z \\
Y_{1,2} &= X_{2,2} + Y_{2,1}
\end{align*}
\]

where \( P_Z(1) = 1 - P_Z(0) = \epsilon \). The rate pair \((R_1,R_2) = (1 - H_2(\epsilon),1)/2\) is achievable by choosing \( X_{1,1} \) as uniform over \( \{0,1\} \) and \( X_{2,2} = X_{2,2}' + Y_{2,1} \) where \( X_{2,2}' \) is independent of \( Y_{2,1} \) and uniform over \( \{0,1\} \). For the converse, the cut-set bound of Theorem 1 is

\[
\bigcup_{P_{X_{1,1}}A_{2,2}} \left\{ (R_1,R_2): 0 \leq R_1 \leq I(X_{1,1};Y_{2,1}|A_{2,2})/2 \\ 0 \leq R_2 \leq I(A_{2,2};Y_{1,2}|X_{1,1})/2 \right\}
\]

and we have \( I(X_{1,1};Y_{2,1}|A_{2,2}) \leq 1 - H_2(\epsilon) \) with equality if \( X_{1,1} \) is uniform and independent of \( A_{2,2} \). We further have \( I(A_{2,2};Y_{1,2}|X_{1,1}) \leq 1 \) since \( Y_{1,2} \) is binary. This shows that Theorem 1 is tight.

Finally, we translate the capacity-achieving strategy into a code tree distribution. We label the branch-pairs of our tree \( A_{2,2} \) as \( b_0b_1 \) by which we mean that \( X_{2,2} = b_0 \) if \( Y_{2,1} = 0 \) and \( X_{2,2} = b_1 \) if \( Y_{2,1} = 1 \). We choose \( A_{2,2} \) independent of \( X_{1,1} \) and

\[
\begin{align*}
P_{A_{2,2}}(00) &= P_{A_{2,2}}(11) = 0 \\
P_{A_{2,2}}(01) &= P_{A_{2,2}}(10) = 1/2
\end{align*}
\]

and compute \( I(A_{2,2};Y_{1,2}|X_{1,1}) = 1 \), as desired.

### E. Quantize-Forward Network Coding

The final channels we consider are relay networks with iBM. Suppose node 1 multicasts a message of rate \( R \) to sink nodes in the set \( T \). The quantize-map-forward (QMF) and noisy network coding (NNC) strategies in \((35)\), \((36)\), \((37)\) generalize to NiBMs and we call the resulting strategies quantize-forward (QF) network coding. QF network coding achieves \( R \) satisfying

\[
LR \leq \min_{k \in \mathcal{S}^c \cap T} \left[ I(A_k^L;\tilde{Y}_S^L|A_{S^c}^L,T) - I(Y_S^L;\tilde{Y}_S^L|A_k^L\tilde{Y}_S^L,T) \right]
\]

for all \( S \subset K \) with \( 1 \in S \) and \( S^c \cap T \neq \emptyset \). The \( A_k^L, k = 1,2,\ldots,K \), are independent and \( \tilde{Y}_S^L \) is a noisy function of \( A_k^L \) and \( Y_S^L \) for all \( k \).

**Remark 32:** A simple lower bound on the first mutual information expression in \((105)\) is

\[
I(A_S^L;\tilde{Y}_S^L|Y_k|A_k^L,T) \geq I(A_S^L;\tilde{Y}_S^L|A_k^L,T).
\]

We use the right-hand side of \((106)\) below because it better matches \((20)\) with \( Y_S^L \) replacing \( Y_k^L \).

**Example 17:** We extend results of \((35)\), \((36)\), \((37)\). If the network is deterministic then \( A_k^L \) determines \( X_k^LinY_k^L \). We thus have

\[
I(Y_S^L;\tilde{Y}_S^L|A_k^LY_k^L,T) = 0
\]

and can choose \( Y_k^L = Y_k^L \) to achieve the cut-set bound but evaluated with independent code functions only. As a result, we obtain the multicast capacity of networks of deterministic point-to-point channels with iBM, for instance. However, QF network coding does not give the capacity region for all deterministic networks because dependent code functions may increase rates.

### F. QF Network Coding for Gaussian Networks

Consider the channel \((18)\) with additive Gaussian noise (AGN), i.e., the \( Z_k \) are Gaussian noise vectors and where \( Z_K \) has a positive definite covariance matrix. For simplicity, we assume that the \( Z_1, Z_2, \ldots, Z_K \) are mutually independent.

Suppose again that node 1 multicasts a message of rate \( R \) to sink nodes in \( T \). Let \( S \) be a cut, i.e., \( 1 \in S \) and \( S^c \cap T \neq \emptyset \). We use the notation

\[
\begin{align*}
Y_{S^c} &= G_{S^c}S_SX_S + G_{S^c}S'_S X_{S'} + Z_{S^c}
\end{align*}
\]

for the \(|S^c|\) equations \((18)\) with \( k \in S^c \), where \( G_{k,j} \) is a \( |U|L \times |V|L \) matrix with block-entries \( G_{k,j} \), \( k \in U \), \( j \in V \). Recall that the \( G_{k,j} \) is \( L \times L \) lower-triangular matrices.

We begin with the upper bound \((29)\) which we write as

\[
\begin{align*}
&h(G_{S^c}S_SX_S + Z_{S^c}||X_{S^c}) - h(Z_{S^c}) \leq h(G_{S^c}S_SX_S + Z_{S^c}) - h(Z_{S^c}) \\
&\leq \frac{1}{2} \log \left| \frac{Q_{Z_{S^c}}}{{Q_{Z_{S^c}}} + G_{S^c}S'_S Q_{X_{S'}} G_{S^c}^{T_S} S_{Z_{S^c}}} \right| \quad (a)
\end{align*}
\]

where \((a)\) follows by a classic maximum entropy theorem. The (positive definite) noise covariance matrix has a Cholesky decomposition \( Q_{Z_{S^c}} = S_{Z_{S^c}} S_{Z_{S^c}}^T \) where \( S_{Z_{S^c}} \) is lower triangular and invertible. We can thus rewrite \((109)\) as

\[
I(X_k^L \rightarrow Y_k^L || X_k^L) \leq \frac{1}{2} \log \left| I_{X_k^L} + \tilde{G}_{S^c} S_{X_{S'}} G_{S^c}^{T_S} S_{Z_{S^c}} \right| \quad (110)
\]

where \( I_{X_k^L} \) is the \( |U|L \times |U|L \) identity matrix and \( \tilde{G}_{S^c} S_{X_{S'}} G_{S^c}^{T_S} S_{Z_{S^c}} \).

For achievability, we choose \( T \) to be a constant and the code functions (effectively) as codewords

\[
A_k^L(\cdot) = X_k^L, \quad k = 1,2,\ldots,K
\]
where \( X^L_k \) is Gaussian. We further choose
\[
Y^L_k = Z^L_k + \hat{Z}^L_k, \quad k = 1, 2, \ldots, K
\]
where \( \hat{Z}^L_k \) is independent of \( X^L_k Y^L_k \) and has the same statistics as \( Z^L_k \). Consider the right-hand side of (106) with codewords rather than code functions. We have
\[
\begin{align*}
I(X^L_S; \hat{Y}^L_S | X^L_K) & \overset{(a)}{=} h(G_{S^c S} X_S + Z_{S^c} + \hat{Z}_{S^c}) - h(Z_{S^c} + \hat{Z}_{S^c}) \\
& \overset{(b)}{=} \frac{1}{2} \log \left| 2Q_{S^c} + G_{S^c S} Q_{X_S} G_{S^c S}^T \right| \\
& = \frac{1}{2} \log \left| I_{S^c} + \frac{1}{2} G_{S^c S} Q_{X_S} G_{S^c S}^T \right| \\
& \overset{(c)}{\geq} \frac{1}{2} \log \left| I_{S^c} + G_{S^c S} Q_{X_S} G_{S^c S}^T \right| - \frac{|S^c|L}{2}
\end{align*}
\]
where (a) is because the \( X^L_k \) are independent, (b) is because the \( X^L_k \) are Gaussian, and (c) follows by using \( |A + B|/2 \geq (|A| + |B|)/2 \) for \( b \times b \) positive definite matrices \( A \) and \( B \). We also have
\[
\begin{align*}
I(Y^L_S; \hat{Y}^L_S | X^L_S) & = I(Z^L_S; \hat{Z}^L_S | X^L_S) \\
& = |S|L/2
\end{align*}
\]
where the last step is because \( \hat{Z}^L_S \) has the same statistics as \( Z^L_S \). Combining (113) and (114) we find that \( R \) satisfying
\[
LR \leq \frac{1}{2} \log \left| I_{S^c} + G_{S^c S} Q_{X_S} G_{S^c S}^T \right| - \frac{KL}{2}
\]
for all \( S \subset K \) with \( 1 \in S \) and \( S^c \cap T \neq \emptyset \) are achievable.

It remains to study the first expression on the right-hand side of (115), both without and with independent \( X^L_k \). Suppose that \( G_{S^c S} \) has the singular value decomposition \( UV^T \Sigma V^T \) so that this expression is
\[
\frac{1}{2} \log \left| I_{S^c} + \Sigma V Q_{X_S} V^T \Sigma V^T \right|.
\]
Suppose there are \( K \) power constraints \( \sum_{k=1}^n E[X^2_{k,i}] \leq P \), \( k = 1, 2, \ldots, K \), i.e., we have symmetric power constraints. Optimizing over \( Q_{X_S} \) we obtain \( \min((|S|, |S^c|)) \cdot L \) parallel channels on which we can put at most power \( |S|P \). We thus have the capacity upper bound
\[
LR \leq \sum_{j} \frac{1}{2} \log (1 + s_j^2 |S|P)
\]
where the sum is over the parallel channels and the \( s_j \) are the singular values.

For a lower bound we simplify (111) even further and choose \( Q_{X_S} = (P/L) \cdot I_{|k|} \). The expression (116) becomes
\[
\begin{align*}
\frac{1}{2} \sum_{s_j} \log (1 + s_j^2 (P/L)) \\
\geq \frac{1}{2} \sum_{s_j} \log (1 + s_j^2 (P/L)) - \frac{|S|L}{2} \log (|S|L).
\end{align*}
\]
We thus have the following theorem that implies that QF network coding approaches capacity at high signal-to-noise ratio. This extends results in [35], [36], [37] to NiBMs.

**Theorem 5:** QF network coding for scalar, linear, AGN channels, symmetric power constraints, and a multicast session achieves capacity to within
\[
K(1 + \log (KL))/2 \text{ bits.}
\]

One may derive better results than (119) by using the approach in (37), for example. Extensions to asymmetric power constraints and multiple multicast sessions are clearly possible.

**APPENDIX A**

**PROOF OF CUT-SET BOUND**

The bound follows from classic steps and the factorizations [15] and [16]. There is one new subtlety, however, namely how to define the random code functions that appear in (20). Fano’s inequality states that for \( P_e \to 0 \) we have
\[
nR_{M(S)} \leq I(W_{M(S)}; \{\tilde{W}^{(t)}_{M(S)} : t \in S^c\})
\]
where (a) follows because \( \tilde{W}_{M(S)} \) is a subset of \( W_{E(S)} \) and because \( \{\tilde{W}^{(t)}_{M(S)} : t \in S^c\} \) is a function of \( Y^{n}_{S^c} \) and \( W_{E(S')} \); (b) follows because the messages are independent and \( A^R_k \) is a function of the messages at node \( k \); and (c) follows because
\[
W_{E(S)} - A^R_k - Y^{L}_{S^c}
\]
forms a Markov chain for any \( S \) and \( S' \). Recall that \( n = mL \) for some integer \( m \). We may thus write
\[
\begin{align*}
I(A^{R}_{S^c} Y^{n}_{S^c} | A^{R}_{S^c}) & \overset{(a)}{=} \sum_{i=1}^m I(A^{R}_{S^c Y^{L}_{S^c}} | A^{R}_{S^c} Y^{(i-1)L}_{S^c}) \\
& \overset{(b)}{=} \sum_{i=1}^m I(A^{R}_{S^c} Y^{(i-1)L}_{S^c} | A^{R}_{S^c} Y^{(i-1)L}_{S^c}) \\
& \leq \sum_{i=1}^m I(A^{R}_{S^c} Y^{(i-1)L}_{S^c} | Y^{L}_{S^c} A^{R}_{S^c} Y^{(i-1)L}_{S^c})
\end{align*}
\]
where (a) follows by choosing \( Y^{L}_{k,i} \) to be the channel output of node \( k \) from time \((i - 1)L + 1 \) to time \( iL \), and where (b) follows by Markovity.

Now let \( A^{R}_{S^c,i} \) be the string of functions \( A_{k,j} \cdot Y^{(i-1)L}_{k,i} \), \( j = (i - 1)L + 1, (i - 1)L + 2, \ldots, iL \). We then have
\[
\begin{align*}
I( & A^{R}_{S^c} Y^{(i-1)L}_{S^c} | Y^{L}_{S^c} A^{R}_{S^c} Y^{(i-1)L}_{S^c}) \\
& \overset{(a)}{=} H(Y^{L}_{S^c,i} | A^{R}_{S^c} Y^{(i-1)L}_{S^c}) - H(Y^{L}_{S^c,i} | A^{R}_{S^c} Y^{(i-1)L}_{S^c}) \\
& \leq H(Y^{L}_{S^c,i} | A^{R}_{S^c} Y^{(i-1)L}_{S^c}) = H(Y^{L}_{S^c,i} | A^{R}_{S^c} Y^{(i-1)L}_{S^c}) \\
& = I(A^{R}_{S^c} Y^{L}_{S^c,i} | A^{R}_{S^c} Y^{L}_{S^c,i})
\end{align*}
\]
(123)
where \((a)\) follows because \(\hat{A}_{k,i}^L\) is a function of \(A_k^L Y_{k}^{(i-1)L}\) and \((b)\) follows because
\[
\hat{A}_{k,i}^L Y_{k}^{(i-1)L} - \hat{B}_{k,i}^L - Y_{k}^{(i-1)L}\]
forms a Markov chain (this step permits \(L\)-letterization).

The remaining steps follow because the \(\hat{A}_{k}^L\)-to-\(Y_{k}^L\) channel does not depend on the block index \(i\). More precisely, we have
\[
P(y_{k,i}^L | \hat{A}_{k,i}^L) = P_{Y_{k}^L | A_k^L}(y_{k,i}^L | \hat{A}_{k,i}^L)
\]
\[
= \left[ \prod_{i=1}^{K} 1(x_{k,i}^{L} | \hat{A}_{k,i}^L, 0 y_{k,i}^{L-1}) \right] P_{Y_{k}^L | X_{k}^L}(y_{k,i}^L | x_{k}^L)
\]
where \(P_{Y_{k}^L | X_{k}^L}\) refers to the first \(L\) channel uses. Inserting \([123]\) into \([122]\), we have
\[
I(A_{k,i}^{m L}; Y_{k,i}^{m L} | A_{k,i}^{m L}) \leq \sum_{i=1}^{m} I(A_{k,i}^{L}; Y_{k,i}^{L} | A_{k,i}^{L})
\]
\[
= m I(A_{k,T}^{L}; Y_{k,T}^{L} | A_{k,T}^{L})
\]
where \(T\) takes on the value \(i, i = 1, 2, \ldots, m\), with probability \(1/m\), and where \((a)\) follows because
\[
T - \hat{A}_{k,T}^L - Y_{k,T}^L
\]
forms a Markov chain. Inserting \([126]\) into \([20]\), we have
\[
L \cdot R_{M(S)} \leq I(A_{k,T}^{L}; Y_{k,T}^{L} | A_{k,T}^{L})
\]
where the joint distribution of the random variables factors as
\[
P(a_{k,T}^L) P_{Y_{k,T}^L | A_k^L}(y_{k,i}^L | \hat{A}_{k,i}^L, T)
\]
and where the second term in \([129]\) is computed using \([125]\) (this step permits the factorization \([21]\)).

Remark 33: If \(n \neq m L\) then we may as well consider \(n = (m-1)L + L'\) where \(0 < L' < L\). The sum in \([129]\) changes and has as its \(mth\) term
\[
I(A_{k,m L'; Y_{k,m L'} | A_{k,m L'})
\]
where the code functions have depth \(L'\). The term \((130)\) could be larger than the right-hand side of \([128]\). However, if \(m\) is large then the capacity is effectively limited by \([128]\).

Remark 34: Consider the \(jth\) cost constraint in \([22]\). We may rewrite \([22]\) as
\[
\frac{1}{L} \sum_{\ell=1}^{L} \sum_{i=1}^{m} 1 \sum_{i=1}^{m} E [s_j (X_{k,(m-1)L+\ell}, Y_{k,(m-1)L+\ell})] = \frac{1}{L} \sum_{\ell=1}^{L} E [s_j (X_{k,(T-1)L+\ell}, Y_{k,(T-1)L+\ell})] \leq S_j
\]
and the inequality in \([131]\) is the \(jth\) inequality in \([23]\).

APPENDIX B
CARDINALITY BOUNDS FOR POINT-TO-POINT CHANNELS
Consider a point-to-point channel with NiBM. We write
\[
P(y_{k}^{L}) = \sum_{a_{k}^{L}} P(a_{k}^{L}) P(y_{k}^{L} | a_{k}^{L})
\]
\[
H(Y_{k}^{L} | A_{k}^{L}) = \sum_{a_{k}^{L}} P(a_{k}^{L}) H(Y_{k}^{L} | A_{k}^{L} = a_{k}^{L})
\]
where \(P(y_{k}^{L} | a_{k}^{L})\) and \(H(Y_{k}^{L} | A_{k}^{L} = a_{k}^{L})\) are determined by the channel \(P(y_{k}^{L} | x_{k}^{L})\). Equations \([132]\) and \([133]\) imply that \(P(y_{k}^{L})\) and \(H(Y_{k}^{L} | A_{k}^{L})\) are convex combinations of \(P(a_{k}^{L})\). Furthermore, if we fix \(P(y_{k}^{L})\) for all \(y_{k}^{L}\) but one, and if we fix \(H(Y_{k}^{L} | A_{k}^{L})\), then we have fixed \(I(A_{k}^{L} ; Y_{k}^{L})\). We can therefore focus on \(|Y_{k}^{L}|\) constraints and \([21]\) Lemma 3.4 guarantees that we need only \(|Y_{k}^{L}|\) non-zero values of \(P(a_{k}^{L})\).

Similarly, observe that
\[
P(y_{k}^{L}) = \sum_{x_{k}^{L}} P(x_{k}^{L} | 0 y_{k}^{L-1}) P(\tilde{y}_{k}^{L}, y_{k}^{L} | x_{k}^{L})
\]
so that if we fix \(P(x_{k}^{L} | 0 y_{k}^{L-1})\) then we have fixed \(P(y_{k}^{L})\). Our approach will be to replace \(|Y_{k}^{L}|\) constraints of the form \([132]\) with (hopefully fewer) constraints to fix \(P(x_{k}^{L} | 0 y_{k}^{L-1})\).

We proceed by induction. We may fix \(P(x_{1})\) with \(|X_{1}|\) constraints of the form
\[
P(x_{1}) = \sum_{a_{1}} P(a_{1}) 1(x_{1} | a_{1}).
\]
This fixes \(P(x_{1}, \tilde{y}_{1})\) because the channel specifies \(P(\tilde{y}_{1} | x_{1})\).

Now suppose that \(P(x_{i-1}, \tilde{y}_{i-1})\) is fixed and write
\[
P(x_{i} | x_{i-1}, \tilde{y}_{i-1}) = \sum_{a_{i}} P(a_{i}) P(x_{i} | \tilde{y}_{i-1}, a_{i}) P(x_{i-1} | \tilde{y}_{i-1}, a_{i})
\]
where \(P(x_{i}, \tilde{y}_{i-1}, a_{i})\) is fixed because \(a_{i}\) is in the conditioning.

We thus define
\[
|X_{i}| \cdot |\tilde{y}_{i-1}| \cdot (|X_{i}| - 1)
\]
constraints of the form \([136]\) to fix \(P(x_{i} | x_{i-1}, \tilde{y}_{i-1})\) for all its arguments. This in turn fixes \(P(x_{i}, \tilde{y}_{i} | x_{i-1}, \tilde{y}_{i-1})\) because the channel specifies \(P(\tilde{y}_{i} | x_{i}, \tilde{y}_{i-1})\).

This we thus find that \(P(x_{i}, \tilde{y}_{i})\) is fixed which completes the induction step. Collecting all the constraints including \([133]\) we have
\[
|X_{1}| + \sum_{i=2}^{L} |X_{i-1}| \cdot |\tilde{y}_{i-1}| \cdot (|X_{i}| - 1)
\]
constraints in total. This number may be less than \(|Y_{k}^{L}|\), e.g., if one of the \(L\) channel outputs is continuous.

APPENDIX C
CARDINALITY BOUNDS FOR RELAYS WITHOUT DELAY
Consider an RC without delay and suppose that \(P(x_{1} | a_{2})\) is specified. This fixes \(P(x_{1}, x_{2}, y_{1}, y_{1} a_{2})\) because the channel fixes \(P(y_{2} | x_{1})\) and \(P(y_{2} | x_{1}, x_{2}, y_{2})\), and \(a_{2}\) specifies \(1(x_{2} | a_{2}, y_{2})\) due to \([1]\). We have thus fixed \(P(y_{2} | a_{2})\), \(H(Y_{3} | X_{1}, A_{2} = a_{2})\), and \(I(X_{1}, Y_{2} Y_{3} | A_{2} = a_{2})\). We further have
\[
P(y_{3}) = \sum_{a_{2}} P(a_{2}) P(y_{3} | a_{2})
\]
\[
H(Y_{3} | X_{1} A_{2} = a_{2}) = \sum_{a_{2}} P(a_{2}) H(Y_{3} | X_{1}, A_{2} = a_{2})
\]
\[
I(X_{1}; Y_{2} Y_{3} | A_{2} = a_{2}) = \sum_{a_{2}} P(a_{2}) I(X_{1}; Y_{2} Y_{3} | A_{2} = a_{2})
\]
Finally, if we fix \( P(y_3) \) for all \( y_3 \) but one, and if we fix \( H(Y_3|X_1A_2) \) and \( I(X_1;Y_3|A_2) \), then we have fixed \( I(X_1A_2|Y_3) \) and (obviously) \( I(X_1;Y_2|A_2) \). We thus have \( |Y_3| + 1 \) constraints in total to specify the bound \([33]\).

Next, note that

\[
P(y_3) = \sum_{x_1,x_2,y_2} P(x_1,x_2)P(y_2|x_1)P(y_3|x_1,x_2,y_2) \tag{142}
\]

so that if we fix \( P(x_1,x_2) \) then we have fixed \( P(y_3) \). We proceed by writing

\[
P(x_1,x_2) = \sum_{a_2} P(a_2)P(x_1,x_2|a_2) \tag{143}
\]

which gives us \(|X_1||X_2| - 1 \) constraints instead of the \(|Y_3| - 1 \) before. Together with \([140]\) and \([141]\) we arrive at \(|X_1||X_2| + 1 \) constraints in total.

**APPENDIX D**

**CONVERSE FOR A CLASS OF MACS WITH FEEDBACK**

Let \( V_i = X_i^{(i-1)L}Y^{(i-1)L} \) for \( i = 1, 2, \ldots, m \). Fano’s inequality, \( P_e \rightarrow 0 \), and the independence of messages give

\[
nR_1 \leq I(W_1;Y^n|W_2) = I(A^n_1;Y^n|A_2^n) = \sum_{i=1}^{m} H(Y_i|A_2^n) - H(Y_i|A_1^nA_2^nY^{(i-1)L}) \leq mI(A_1^n;Y^n|A_2^n) \tag{144}\]

where \((a)\) follows because \( A_2^nY^{(i-1)L} \) defines \( X_i^n \) and therefore also \( X_i^{(i-1)L} \). Step \((b)\) follows by using \( T \) as our time-sharing random variable, \( A_1^n \), as in Appendix A and similar steps as in \([123]\); step \((c)\) follows because

\[
T - V_T A_1^n T A_2^n T - Y_T^n \tag{145}
\]

forms a Markov chain. The chains

\[
T - A_1^n T A_2^n T - Y_T^n \tag{146}
\]

\[
A_1^n T T - V_T - A_2^n T \tag{147}
\]

are also Markov.

By symmetry, we have a similar bound as \([144]\) for \( nR_2 \). The corresponding sum-rate bound is

\[
n(R_1 + R_2) \leq I(W_1W_2;Y^n) = I(A^n_1;A_2^n;Y^n) \leq \sum_{i=1}^{m} H(Y_i|A_2^n) - H(Y_i|A_1^nA_2^nY^{(i-1)L}) = mI(A_1^n;A_2^n;Y^n|T) \leq mI(A_1^n;A_2^n;Y^n|T) \tag{148}\]

Collecting the bounds, we arrive at the region of Theorem 2.

The cardinality bound follows by using similar steps as in Appendices B and C see also \([38\ App. B]\).

**APPENDIX E**

**WEAKENED BOUND FOR CAUSAL RELAY NETWORKS**

The bound \([15\ Thm. 2]\) follows from Theorem 2 in a different way than \([24\ and 26]\). We have

\[
I(A^L_S;Y^L_S|A^L_S) = \sum_{i=1}^{L} H(Y_{S^L,i}|Y_{S^L,i-1}X_{S^L,i}A^L_S) \tag{149}
\]

because \( X_{S^L,i} \) is a function of \( Y_{S^L,i-1} \) and \( A^L_S \). We bound the first entropy in the sum in \([149]\) as

\[
H(Y_{S^L,i}|Y_{S^L,i-1}X_{S^L,i}A^L_S) \leq H(Y_{S^L,i}|Y_{S^L,i-1}X_{S^L,i}A^L_S\cap N_0) \tag{150}
\]

For the second entropy in \([149]\) we use two approaches. For \( 1 \leq i \leq L - 1 \) we bound

\[
H(Y_{S^L,i}|Y_{S^L,i-1}X_{S^L,i}A^L_K) \geq H(Y_{S^L,i}|Y_{K^L,i}X_{S^L,i}A^L_K) \tag{151}\]

where \((a)\) follows because (cf. \([25]\))

\[
A^L_K - Y_{K^L,i}X_{S^L,i}A^L_S - Y_{S^L,i} \tag{152}\]

forms a Markov chain for all \( i = 1, 2, \ldots, L \). Next, for time \( i = L \) we use

\[
H(Y_{S^L,L}|Y_{S^L,L-1}X_{S^L,L}A^L_L) \tag{153}\]

\[
H(Y_{S^L,L}|Y_{S^L,L-1}X_{S^L,L}A^L_L) \tag{154}\]

forms a Markov chain, and \((b)\) follows because \( X_{S^L,N_1} \) is a function of \( A^L_S\cap N_1 \) and because

\[
A^L_S\cap N_1 - Y_{S^L,L-1}X_{S^L,L}X_{S^L,N_1}A^L_S\cap N_0 - Y_{S^L,L} \tag{155}\]

forms a Markov chain.

Summarizing, we insert \([150]\), \([151]\), and \([153]\) into \([149]\) and obtain the following bound that appeared in \([15\ Thm. 2]\):

\[
I(A^L_S;Y^L_S|A^L_S) \leq I(X_{S^L,L-1},0;Y_{S^L,L-1}|X_{S^L,L-1}A^L_S\cap N_0) \tag{156}\]

**ACKNOWLEDGMENTS**

The results reported here were motivated by the paper \([10\] by S. Fong and R. Yeung. I am grateful to these authors for sending me of an early version of their work. I am also grateful to S.-Y. Chung, H. Permuter, Y.-H. Kim, and the two reviewers for their detailed and constructive criticisms of the document.
