Relay Networks With Delays

Abbas El Gamal, *Fellow, IEEE*, Navid Hassanpour, and James Mammen, *Student Member, IEEE*

***Abstract*—The paper investigates the effect of link delays on the capacity of relay networks. The relay-with-delay is defined as a relay channel with relay encoding delay of $d \in \mathbb{Z}$ units, or equivalently, a delay of d units on the link from the sender to the relay, zero delay on the links from the transmitter to the receiver and from the relay to the receiver, and zero relay encoding delay. Two special cases are studied. The first is the relay-with-unlimited lookahead, where each relay transmission can depend on its entire received sequence, and the second is the relay-without-delay, where the relay transmission can depend only on current and past received symbols, i.e., $d = 0$. Upper and lower bounds on capacity for these two channels that are tight in some cases are presented. It is shown that the cut-set bound for the classical relay channel, corresponding to the case where $d = 1$, does not hold for the relay-without-delay. Further, it is shown that instantaneous relaying can be optimal and can achieve higher rates than the classical cut-set bound. Capacity for the classes of degraded and semi-deterministic relay-with-unlimited-look-ahead and relay-without-delay are established. These results are then extended to the additive white Gaussian noise (AWGN) relay-with-delay case, where it is shown that for any $d \leq 0$, capacity is achieved using amplify-and-forward when the channel from the sender to the relay is sufficiently weaker than the other two channels. In addition, it is shown that a superposition of amplify-and-forward and decode-and-forward can achieve higher rates than the classical cut-set bound.**

The relay-with-delay model is then extended to feedforward relay networks. It is shown that capacity is determined only by the relative delays of paths from the sender to the receiver and not by their absolute delays. A new cut-set upper bound that generalizes both the classical cut-set bound for the classical relay and the upper bound for the relay-without-delay on capacity is established.

***Index Terms*—Capacity, cut-set bound, delay, relay channel, relay networks.**

I. INTRODUCTION

THIS paper is motivated by the general question of whether link delays can change the information-theoretic capacity of a communication network. The answer at first appears to be that delay should have no effect on capacity, because achieving capacity requires an arbitrarily long delay. As we shall see this is not always the case. Link delays can change the nature of cooperation in a network, and hence its capacity.

First consider a discrete-memoryless point-to-point channel (DMC) consisting of an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$, and a family of conditional probability mass functions (pmfs) $p(y|x)$ on $\mathcal{Y}$ for each $x \in \mathcal{X}$. If the sender X transmits x_t at time t, then the received symbol y_t at the receiver Y is chosen according to $p(y_t|x_t)$. Now, suppose there is a finite transmission delay of $d \in \mathbb{Z}$, then if x_t is transmitted, the received symbol y_{t+d} at time $t+d$ is chosen according to $p(y_{t+d}|x_t)$. We refer to this channel as DMC-with-delay and represent it graphically as shown in Fig. 1. Now, by applying a decoding delay of d, it is easy to see that the capacity of the DMC-with-delay is the same as the capacity with no delay. Similarly, it can be easily shown that finite link delays do not change the capacity region for the multiple-access channel or the broadcast channel.

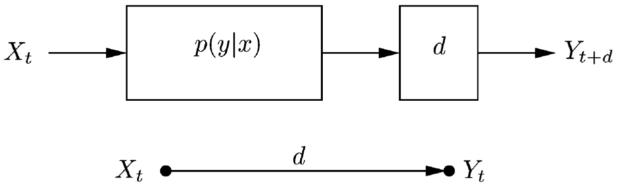

Fig. 1. DMC-with-delay and its graphical representation.

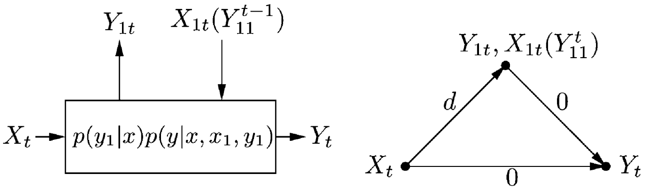

Fig. 2. (Left) Classical relay channel. (Right) Graphical representation for relay-with-delay; $d = 1$ corresponds to classical relay, $d = 0$ corresponds to relay-without-delay.

The story for the relay channel is quite different. Recall that in the *classical relay* channel model introduced by van der Meulen [1] and studied extensively in the literature (e.g., [2], [6], [10]), each transmitted relay symbol x_{1t} can depend only on its past received symbols y_{11}^{t-1} (see Fig. 2 (left)). This "encoding delay" can be equivalently represented by a link delay. By assuming zero encoding delay and adding a delay of 1 to the link from the sender X to the relay receiver Y_1 and a delay of zero to the other two links, we obtain the equivalent model in Fig. 2(right). If we now assume that the link from the sender to the relay receiver has a delay of 0 (instead of 1), we obtain the *relay-without-delay* model studied in [7], [8], which in turn can be viewed as a relay channel where each transmitted relay symbol depending on present as well as past received symbols, i.e., x_t depends on y_{11}^t. As shown in [7], [8], [11], this seemingly minor change to the channel model increases its capacity by allowing the sender and the relay sender to instantaneously cooperate.

In this paper, we provide a more unified and complete treatment of the results in these papers and in [9]. We introduce the more general *relay-with-delay* model, where there is a finite delay $d \in \mathbb{N}$ on the link from the sender X to the relay receiver Y_1 (see Fig. 2). As we shall see in Section VI, this is equivalent in capacity to having arbitrary delays on all three links. Thus

- $d = 1$ corresponds to the classical relay;

Manuscript received September 1, 2006; revised February 6, 2007.

A. El Gamal is with the Electrical Engineering Department and the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA (e-mail: abbas@ee.stanford.edu).

N. Hassanpour and J. Mammen are with the Electrical Engineering Department, Stanford University, Stanford, CA 94305-9510 USA (e-mail: navid@stanford.edu; james.mammen@gmail.com).

Communicated by M. Gastpar, Guest Editor for the Special Issue on Relaying and Cooperation.

Digital Object Identifier 10.1109/TIT.2007.904838

0018-9448/$25.00 © 2007 IEEE

- $d = 0$ corresponds to the relay-without-delay;
- $d > 1$ corresponds to relay encoding delay of $d > 1$; and
- $d < 0$ corresponds to a *look-ahead* of $-d$ at the relay encoder.

We also introduce the *relay-with-unlimited-look-ahead* model, where the relay knows its entire received sequence *noncausally* and thus each transmitted relay symbol can depend on the entire received sequence y_{11}^n. While this scenario does not currently have a clear practical motivation, it provides a limit on the extent to which relaying can help communication.

We present upper and lower bounds on capacity for the relay-with-unlimited-look-ahead and the relay-without-delay and show that they are tight in some cases. The lower bounds discussed are achieved using combinations of *coherent cooperation* strategies that depend on delay.

1) *Decode-and-forward*: Here the relay decodes part or all of the message and the sender and relay cooperate on sending the *previous* message [2]. This requires knowledge only of past received relay symbols and therefore is possible to implement for any finite d.
2) *Instantaneous relaying*: Here the relay sends a function only of its current received symbol. This is possible when the relay has access to the current received symbol, which is the case for any $d \leq 0$.
3) *Noncausal decode-and-forward*: This scheme is possible only when the relay has unlimited look-ahead. The relay predecodes part or all of the message before communication commences and cooperates with the sender to transmit the message to the receiver.

We then generalize the relay-with-delay model to relay networks with delays and present results on capacity including a new cut-set type upper bound.

Our results have several interesting implications.

- The well-known cut-set bound [2], [5] is not in general an upper bound on the capacity of a relay network with link delays.
- Instantaneous relaying alone can achieve higher rates than the cut-set bound.
- Amplify-and-forward can be optimal for the "full-duplex"additive white Gaussian noise (AWGN) relay-with-delay for $d \leq 0$. This is in contrast to the classical case where capacity is not known for *any* finite channel gain values.
- A mixture of cooperation strategies may be needed to achieve the capacity of a relay-with-delay.
- The capacity of a relay network with delays depends only on the relative path delays from the sender to the receiver, and not on absolute delays.

The following is an outline of the paper and the main results.

- Section II introduces needed definitions for the relay-with-delay and briefly reviews results on the capacity of the classical relay.
- Section III deals with the relay-with-unlimited-look-ahead. We discuss this case first because the upper bound on capacity provided is used in subsequent sections.
  — An upper bound on capacity, and thus on the capacity of the relay-with-delay for any d is established in Theorem 1.
  — Lower bounds on capacity based on "noncausal" decode-and-forward and partial decode-and-forward are provided in Propositions 1 and 3, respectively.
  — Capacity is established for the classes of degraded and semi-deterministic relay-with-unlimited-look-ahead in Propositions 2 and 4, respectively.
- Section IV deals with the relay-without-delay and has a parallel structure to Section III.
  — An upper bound on capacity, which is in general tighter than both the classical cut-set bound and the upper bound in Theorem 1, is established in Theorem 2.
  — A lower bound on capacity achieved by instantaneous relaying is provided in (15). It is shown through an example that this lower bound can be tight even for a relay-with-unlimited-delay, and can achieve higher rates than the classical cut-set bound.
  — A lower bound achieved by a superposition of instantaneous relaying and partial decode-and-forward is presented in Proposition 5.
  — The lower bound is shown to be optimal for degraded and semi-deterministic relay-without-delay in Propositions 6 and 7, respectively.
- Section V deals with the "full-duplex" AWGN relay-with-delay.
  — It is shown in Proposition 9 that when the channel from the sender to the relay is sufficiently weak, amplify-and-forward is optimal even for the unlimited look-ahead case.
  — It is shown that a superposition of amplify-and-forward and decode-and-forward can achieve higher rates than the classical cut-set bound.
  — The capacity of the AWGN relay-with-unlimited-look-ahead is established for the case when the channel from the relay to the receiver is sufficiently strong in Proposition 8.
- Section VI, deals with general feedforward relay networks with delays.
  — It is shown in Theorem 3 that two relay networks that differ only in their link delays have the same capacity if the relative delay of every path from the sender to the receiver (relative to the minimum delay path) is the same in both networks.
  — An upper bound that generalizes the classical cut-set bound and the bound for the relay-without-relay in Theorem 2 is given in Theorem 4. The bound involves the use of auxiliary random variables and multiple random variables per sender.
- Section VII discusses open questions and possible extensions of this work.

II. Preliminaries

The discrete-memoryless relay-with-delay channel consists of a sender alphabet $\mathcal{X}$, a receiver alphabet $\mathcal{Y}$, a relay sender alphabet $\mathcal{X}_1$, a relay receiver alphabet $\mathcal{Y}_1$, and a family of conditional pmfs $p(y_1|x)p(y|x, x_1, y_1)$[1] on $\mathcal{Y} \times \mathcal{Y}_1$, one for each $(x, x_1) \in \mathcal{X} \times \mathcal{X}_1$. We assume a delay of $d \in \mathbb{Z}$ on the link from the sender X to the relay receiver Y_1, and zero delay on the other two links (see Fig. 2). The channel is memoryless in the sense that for any block length $n = 1, 2, \ldots$

$$p\left(y_{11}^n|x^n\right) p\left(y^n|x^n, x_{11}^n, y_{11}^n\right) = \prod_{t=1}^{n} p(y_{1t}|x_{t-d}) p(y_t|x_t, x_{1t}, y_{1t})$$

where the pmfs with symbols that do not have positive time indices are arbitrary. Note that the common notation $y^t = (y_1, y_2, \ldots, y_t)$ and $y_{1s}^t = (y_{1s}, y_{1(s+1)}, \ldots, y_{1t})$ are used throughout.

A $(2^{nR}, n)$ code for the relay-with-delay consists of: i) a set of messages $\{1, 2^{nR}\}$, ii) an encoding function that maps each message $w \in \{1, 2^{nR}\}$ into a codeword $x^n(w)$ of length n, iii) relay encoding functions $x_{1t} = f_t(y_{11}, y_{12}, \ldots, y_{1(t-1)}, y_{1t}) = f_t\left(y_{11}^t\right)$, for $1 \le t \le n$, and iv) a decoding function that maps each received sequence y^n into an estimate $\hat{w}(y^n)$.

A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} = \mathrm{P}\{\hat{W} \neq W\} \to 0$, as $n \to \infty$. Channel capacity of a relay-with-delay, denoted by C_d, is the supremum over the set of achievable rates. Note that C_d is monotonically nonincreasing in d.

Thus, $d = 1$ corresponds to classical relay with capacity C_1 and $d = 0$ corresponds to relay-without-delay with capacity C_0.

To simplify notation for the discussion on the relay-with-delay, we use the following equivalent definition.

Look-Ahead Notation: We assume zero delay on all links and relay encoding delay of d, or equivalently a look-ahead of $-d$ for $d < 0$. Thus, the relaying functions for $1 \le t \le n$ are now of the form: $x_{1t} = f_t\left(y_{11}^{t-d}\right)$. In Section VI, we revert to the delay notation mentioned earlier to prove results for general networks.

We define the *relay-with-unlimited-look-ahead* as a relay channel where the relaying functions are of the form $x_{1t} = f_t\left(y_{11}^n\right)$ and denote its capacity by C^*.

Remarks:

1) Capacity of the relay-with-delay, C_d, is monotonically nonincreasing in d, as the dependency of relaying functions on more received symbols cannot hurt.
2) Capacity of the relay-with-delay, C_d, is not known in general for any finite d.
3) In this paper, we only consider the case of $d \le 0$. Although it is quite realistic to assume delay $d > 1$, we have no new results to report on this case beyond straightforward applications of known results for the classical relay.
4) A seemingly more general definition of relay-with-delay would be to have arbitrary delays $d(1,2)$ on the link from X to Y_1, $d(2,3)$ on the link from X_1 to Y, and $d(1,3)$ on the link from X to Y as shown in Fig. 9(a). As we show in Example 7 in Section VI, capacity in this case is the same as the capacity of the relay-without-delay as defined above with $d = d(1,2) + d(2,3) - d(1,3)$.
5) The definition of the relay-with-delay is extended to general feedforward relay networks with delays in Section VI.

A. Results for Classical Relay

In the classical relay, the transmitted relay symbol is allowed to depend only on past received symbols. This is equivalent to the case of $d = 1$. We shall refer to the following well-known results for this channel.

1) Classical cut-set bound [2]:

$$C_1 \le \max_{p(x,x_1)} \min\left\{I(X; Y, Y_1|X_1), I(X, X_1; Y)\right\}. \quad (1)$$

We shall refer to the first term inside the minimum as the *broadcast bound* and to the second as the multiple-access bound. Note that this bound is tight for all cases where capacity is known.

2) Partial decode-and-forward [2]:

$$C_1 \ge \max_{p(u,x,x_1)} \min\left\{ I(X, X_1; Y), I(X; Y|X_1, U) + I(U; Y_1|X_1) \right\}. \quad (2)$$

This is achieved using a block Markov coding scheme, where, in each block, the relay decodes part of the message represented by U and cooperates with the sender to help the receiver decode the message sent in the previous block. This scheme reduces to the decode-and-forward scheme where the relay decodes the entire message, which leads to the generally looser lower bound

$$C_1 \ge \max_{p(x,x_1)} \min\left\{I(X, X_1; Y), I(X; Y_1|X_1)\right\}. \quad (3)$$

Note that in partial decode-and-forward, the sender and relay coherently and noiselessly cooperate but with a very large delay.

3) Compress-and-forward with time-sharing [10]: [2]

$$C_1 \ge \max_{p(q)p(x|q)p(x_1|q)p(\hat{y}_1|y_1,q)} \min\left\{ I(X; Y, \hat{Y}_1|X_1, Q), I(X, X_1; Y|Q) - I(Y_1; \hat{Y}_1|X, X_1, Y, Q) \right\}. \quad (4)$$

4) Capacity theorems:

(a) Degraded relay channel [2]: Here $X \to (Y_1, X_1) \to Y$ form a Markov chain. Capacity is given by

$$C_{1,\,\mathrm{deg}} = \max_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y_1|X_1)\} \quad (5)$$

and is achieved using decode-and-forward.

(b) Semi-deterministic relay channel [6]: Here $Y_1 = g(X)$. Capacity is given by

$$C_{1,\,\mathrm{semi\!-\!-det}} = \max_{p(x,x_1)} \min\{I(X, X_1; Y),\ H(Y_1|X_1) + I(X; Y|X_1, Y_1)\} \quad (6)$$

and is achieved using partial decode-and-forward with $U = Y_1$.

[1] In the classical relay, the probability transition function is defined in a more general way as $p(y, y_1|x, x_1)$. The restriction in our definition is to avoid instantaneous feedback from X_1 to Y_1, which can occur because of the introduction of delay.

[2] The conditioning on Y in the last mutual information term was mistakenly dropped in [10].

III. Relay-With-Unlimited-Look-Ahead

In this section, we provide upper and lower bounds on the capacity C^* of the discrete-memoryless relay-with-unlimited-look-ahead, where the relay function at any time can depend on the entire received sequence y_{11}^n.

A. *Upper Bound on C^**

In the following, we provide an upper bound on C^*. Note that since $C_d \leq C^*$ for all d, this is an upper bound on C_d for any d. We show later that this bound can be strictly larger than the classical cut-set bound (1) for $d = 0$ (and thus for any $d \leq 0$). For $d \geq 1$, the classical cut-set bound applies and therefore this bound is uninteresting.

Theorem 1: The capacity of the discrete memoryless relay-with-unlimited-look-ahead is upper_bounded as follows:

$$C^* \leq \sup_{p(x,x_1)} \min\{I(X,X_1;Y), I(X;Y_1) + I(X;Y|X_1,Y_1)\}. \tag{7}$$

Proof: We prove a weak converse. Using Fano's inequality, it follows that

$$\begin{aligned} nR &= H(W) \\ &= I(W;Y^n) + H(W|Y^n) \leq I(W;Y^n) + n\epsilon_n \end{aligned}$$

for some $\epsilon_n \to 0$ as $n \to \infty$.

We bound the term $I(W;Y^n)$ in two ways. First consider

$$\begin{aligned} I(W;Y^n) &= \sum_{t=1}^n I\left(W;Y_t|Y^{t-1}\right) \\ &= \sum_{t=1}^n \left(H\left(Y_t|Y^{t-1}\right) - H\left(Y_t|W,Y^{t-1}\right)\right) \\ &\overset{(a)}{\leq} \sum_{t=1}^n \left(H(Y_t) - H\left(Y_t|X_t,X_{1t},W,Y^{t-1}\right)\right) \\ &\overset{(b)}{=} \sum_{t=1}^n (H(Y_t) - H(Y_t|X_t,X_{1t})) \\ &= \sum_{t=1}^n I(X_t,X_{1t};Y_t) \\ &= \sum_{t=1}^n I(X_t,X_{1t};Y_t|Q=t) \\ &= nI(X_Q,X_{1Q};Y_Q|Q) \overset{(c)}{\leq} nI(X,X_1;Y), \end{aligned}$$

where inequality (a) holds because conditioning reduces entropy and equality (b) holds because the channel is memoryless, Q is a "time-sharing" random variable taking values in $\{1,2,\ldots,n\}$, and is independent of X^n, X_{11}^n, Y^n, and y_{11}^n, and $X = X_Q$, $X_1 = X_{1Q}$, and $Y = Y_Q$. Inequality (c) holds by the concavity of mutual information.

Next, consider

$$\begin{aligned} I(W;Y^n) &\leq I\left(W;Y^n,Y_{11}^n\right) \\ &= I\left(W;Y_{11}^n\right) + I\left(W;Y^n|Y_{11}^n\right). \end{aligned} \tag{8}$$

The first term is upper-bounded as follows:

$$\begin{aligned} I\left(W;Y_{11}^n\right) &= \sum_{t=1}^n I\left(Y_{1t};W|Y_{11}^{t-1}\right) \\ &\leq \sum_{t=1}^n \left(H\left(Y_{1t}|Y_{11}^{t-1}\right) - H\left(Y_{1t}|W,Y_{11}^{t-1},X_t\right)\right) \\ &\overset{(a)}{=} \sum_{t=1}^n \left(H\left(Y_{1t}|Y_{11}^{t-1}\right) - H(Y_{1t}|X_t)\right) \\ &\leq \sum_{t=1}^n I(X_t;Y_{1t}|Q=t) \leq nI(X;Y_1). \end{aligned}$$

Equality (a) holds because the channel is memoryless.

Now, we bound the second term in (8)

$$\begin{aligned} I\left(W;Y^n|Y_{11}^n\right) &= \sum_{t=1}^n I\left(W;Y_t|Y^{t-1},Y_{11}^n\right) \\ &= \sum_{t=1}^n H\left(Y_t|Y^{t-1},Y_1^n\right) \\ &\quad - \sum_{t=1}^n H\left(Y_t|W,Y^{t-1},Y_1^n\right) \\ &\overset{(b)}{=} \sum_{t=1}^n H\left(Y_t|Y^{t-1},Y_{11}^n,X_{1t}\right) \\ &\quad - \sum_{t=1}^n H\left(Y_t|W,Y^{t-1},Y_1^n\right) \\ &\leq \sum_{t=1}^n H(Y_t|X_{1t},Y_{1t}) \\ &\quad - \sum_{t=1}^n H\left(Y_t|X_t,X_{1t},W,Y^{t-1},Y_1^n\right) \\ &\overset{(c)}{=} \sum_{t=1}^n H(Y_t|X_{1t},Y_{1t}) \\ &\quad - \sum_{t=1}^n H(Y_t|X_t,X_{1t},Y_{1t}) \\ &= \sum_{t=1}^n I(X_t;Y_t|X_{1t},Y_{1t},Q=t) \\ &\leq nI(X;Y|X_1,Y_1). \end{aligned}$$

Hence

$$I(W;Y^n) \leq n(I(X;Y_1) + I(X;Y|X_1,Y_1)).$$

Equality (b) holds because for relay-with-unlimited-look-ahead $x_{1i} = f_i\left(y_{11}^n\right)$, and (c) holds because the channel is memoryless. This completes the proof. □

Remark: Note that while the first bound in the above upper bound (7) coincides with the multiple-access bound in the classical cut-set bound (1), the second bound is increased over the broadcast bound by

$$I(X;Y_1) + I(X;Y|X_1,Y_1) - I(X;Y,Y_1|X_1) = I(X_1;Y_1).$$

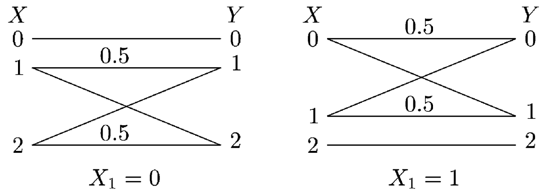

Fig. 3. Sato's example in which $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y} = \{0, 1, 2\}$, $\mathcal{X}_1 = \{0, 1\}$, and $Y_1 = X$.

This represents the potential rate gain achieved by using future and present in addition to past received relay symbols.

B. Noncausal Decode-and-Forward Lower Bound on C^*

Since the relay knows y_{11}^n before transmission commences, it can, assuming the rate is sufficiently low, decode part or all of the message. The sender and relay can then cooperate to send the message to the receiver. We refer to such a scheme as *noncausal decode-and-forward*. The following establishes a lower bound based on complete decodability of the message by the relay.

Proposition 1: The capacity of the relay-with-unlimited-look-ahead is lower-bounded as

$$C^* \geq \max_{p(x,x_1)} \min\{I(X; Y_1), I(X, X_1; Y)\}. \tag{9}$$

Proof: To prove achievability of the right-hand side expression, we use a random coding argument. We fix a joint pmf $p(x, x_1)$ and generate 2^{nR} codeword pairs $(x^n, x_{11}^n)(w)$ for $w \in [1, 2^{nR}]$ according to $\prod_{t=1}^n p(x_t, x_{1t})$. Since the relay knows y_{11}^n in advance, it decodes w before transmission commences. This can be achieved provided $R < I(X; Y_1)$. The sender and relay then cooperatively transmit $(x^n, x_{11}^n)(w)$. The receiver can reliably decode w provided $R < I(X, X_1; Y)$. Combining the two bounds completes the proof. □

We now show that this lower bound is tight for the degraded relay-with-unlimited-look-ahead.

Proposition 2: The capacity of the degraded relay-with-unlimited-look-ahead is given by

$$C_{\text{deg}}^* = \max_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y_1)\}. \tag{10}$$

Proof: Since for the degraded relay channel, $X \rightarrow (X_1, Y_1) \rightarrow Y$ form a Markov chain, the second term in (7) reduces to $I(X; Y_1) + I(X; Y|X_1, Y_1) = I(X; Y_1)$. This upper bound coincides with the lower bound in (9), which completes the proof. □

To illustrate this result, consider the following degraded relay-with-unlimited-look-ahead example.

Example 1: Consider the relay channel example introduced by Sato [14] in Fig. 3. In [14], Sato used first- and second-order Markov processing at the relay to find achievable rates for the classical case of 1.0437 and 1.0549 bits/transmission, respectively. By noting that this channel is physically degraded, in [2], it was found that the capacity for the classical case, $C_{1,\text{ Sato}} = 1.161878$ bits/transmission. This rate is achieved using decode-and-forward coding and coincides with the classical cut-set bound.

TABLE I
CAPACITY ACHIEVING INPUT DISTRIBUTION $p^*(x, x_1)$

	$X = 0$	$X = 1$	$X = 2$
$X_1 = 0$	$7/18$	$1/18$	$1/18$
$X_1 = 1$	$1/18$	$1/18$	$7/18$

We now show that the capacity of Sato's relay-with-unlimited-look-ahead is given by

$$C_{\text{Sato}}^* = \log(9/4) = 1.169925 \text{ bits/transmission.} \tag{11}$$

To prove this, we compute the capacity expression in (10). Let $p_{ij} = \mathrm{P}\{X_1 = i, X = j\}$ for $i = 0, 1$ and $j = 0, 1, 2$. Then the first term in (10) can be rewritten as

$$I(X, X_1; Y) = H(Y) - (p_{01} + p_{02} + p_{10} + p_{11}).$$

The second term is given by

$$I(X; Y_1) = H(X).$$

Both $H(X)$ and $H(Y)$ are concave functions of $\{p_{ij}\}$. Now, given a joint distribution $\{p_{ij}\}$, consider the joint distribution $p'_{00} = p'_{12} = \frac{1}{2}(p_{01} + p_{12})$ and $p'_{01} = p'_{02} = p'_{10} = p'_{11} = \frac{1}{4}(p_{01} + p_{02} + p_{10} + p_{11})$. It is easy to see that $H(X)$ and $H(Y)$ are higher for $\{p'_{ij}\}$ than for $\{p_{ij}\}$ and that $p_{01} + p_{02} + p_{10} + p_{11} = p'_{01} + p'_{02} + p'_{10} + p'_{11}$. Therefore, the maximizing input distribution $p^*(x, x_1)$ is obtained when $p_{10} = p_{02} = p_{10} = p_{11}$ and $p_{00} = p_{12}$. Maximizing the minimum of the two terms subject to these constraints yields $C_{\text{Sato}}^* = 1.169925$ and the maximizing input distribution in Table I.

Remarks:

1) Since for this example $C^* > C_1$ and C_1 coincides with the classical cut-set bound, we conclude that the upper bound in Theorem 1 can be strictly larger than the classical cut-set bound.
2) F. Willems has pointed out to the authors that any relay-without-delay for which $Y_1 = X$ is a special case of situation (3) in [17], which deals with the multiple-access channel with cribbing encoders. Note that for this case, the capacity region does not increase by having more than zero look-ahead. Using this observation, it also follows that

$$C_0 = \max_{p(x,x_1)} \min\{H(X), I(X, X_1; Y)\}.$$

Later, we show that the same rate can be achieved when $d = 0$, thus showing that

$$C_0 = C^* = \max_{p(x,x_1)} \min\{H(X), I(X, X_1; Y)\}.$$

We can generalize the above decode-and-forward scheme by having the relay decode only part of the message. In addition to cooperating with the sender to transmit this part of the message, the sender superimposes additional information destined only to the receiver. This scheme, which we refer to as partial noncausal decode-and-forward, yields the following lower bound on C^*.

Proposition 3: The capacity of the relay-with-unlimited-look-ahead is lower-bounded as

$$C^* \geq \max_{p(u,x_1)p(x|u)} \min\Big\{ I(X, X_1; Y), \\ I(U; Y_1) + I(X; Y | X_1, U)\Big\}. \quad (12)$$

Proof: Achievability is straightforward. Split the rate into cooperation rate R_0 and superposition rate R_1. The total rate is thus $R = R_0 + R_1$. For the scheme to work reliably, we must have

$$\begin{aligned} R_0 &< I(U; Y_1) \\ R_0 &< I(U, X_1; Y) \\ R_1 &< I(X; Y | X_1, U). \end{aligned}$$

The desired bound follows by adding the last inequality to each of the first and second inequalities. □

Note that this lower bound is tight for the class of semi-deterministic relay channels with unlimited look-ahead.

Proposition 4: The capacity of the semi-deterministic relay-with-unlimited-look-ahead channel is given by

$$C^*_{\text{semi-det}} = \max_{p(x,x_1)} \min\{I(X, X_1; Y), H(Y_1) \\ + I(X; Y | X_1, Y_1)\}. \quad (13)$$

Proof: Achievability follows from Proposition 3 by setting $U = Y_1$. The converse follows by noting that the right-hand side of (13) coincides with the upper bound of Theorem 1. □

IV. Relay-Without-Delay

Here we provide upper and lower bounds on the capacity C_0 of the discrete-memoryless relay-without-delay and show that they are tight in some cases. Further, we show that rates higher than the classical cut-set bound can be achieved.

A. Upper Bound on C_0

In the case of the relay-without-delay we obtain the following bound, which is in general tighter than the bound in (7).

Theorem 2: The capacity of the relay-without-delay channel is upper-bounded as follows:

$$C_0 \leq \max_{p(v,x), f(v,y_1)} \min\{I(V, X; Y), I(X; Y, Y_1 | V)\} \quad (14)$$

where $x_1 = f(v, y_1)$, and the cardinality of the auxiliary random variable V is upper-bounded as $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{X}_1| + 1$.

Proof: Assume there is a sequence of $(2^{nR}, n)$ codes for the relay-without-delay channel with $P_e^{(n)} \to 0$ as $n \to \infty$. From the structure of the channel and codes, the empirical probability distribution over $(W, X^n, Y_{11}^n, X_{11}^n, Y^n)$, for any n, is of the form

$$p(w)p(x^n|w)\prod_{t=1}^n p(y_{1t}|x_t)\prod_{t=1}^n p\left(x_{1t}|y_1^t\right)\prod_{t=1}^n p(y_t|x_t, x_{1t}, y_{1t}).$$

For $1 \leq t \leq n$, define the random variable $V_t = Y_{11}^{t-1}$. Note that $X_{1t} = f_t(V_t, Y_{1t})$. We show that $(W, Y^{t-1}) \to (X_t, V_t) \to (Y_t, Y_{1t})$ form a Markov chain as follows. Consider

$$\begin{aligned} p\left(w, x_t, y_{11}^t, y^t\right) &= p\left(w, y^{t-1}, x_t, y_{11}^{t-1}, y_{1t}, y_t\right) \\ &= p(w, y^{t-1}) p\left(x_t, y_{11}^{t-1} | w, y^{t-1}\right) \\ &\quad \cdot p\left(y_t, y_{1t} | x_t, y_{11}^{t-1}, w, y^{t-1}\right) \\ &= p(w, y^{t-1}) p\left(x_t, y_{11}^{t-1} | w, y^{t-1}\right) \\ &\quad \cdot p\left(y_{1t} | x_t, y_{11}^{t-1}, w, y^{t-1}\right) \\ &\quad \cdot p\left(y_t | x_t, y_{11}^{t-1}, w, y^{t-1}, y_{1t}\right) \\ &\overset{(a)}{=} p(w, y^{t-1}) p\left(x_t, y_{11}^{t-1} | w, y^{t-1}\right) \\ &\quad \cdot p\left(y_{1t} | x_t, y_{11}^{t-1}\right) p\left(y_t | x_t, y_{11}^{t-1}, y_{1t}\right) \\ &= p(w, y^{t-1}) p\left(x_t, y_{11}^{t-1} | w, y^{t-1}\right) \\ &\quad \cdot p\left(y_t, y_{1t} | x_t, y_{11}^{t-1}\right), \end{aligned}$$

where (a) follows by the memorylessness of the channel and the facts that i) Y_{1t} is conditionally independent of W, Y^{t-1}, Y_{11}^{t-1} given X_{1t}, and ii) Y_t is conditionally independent of W, Y^{t-1} given X_t, Y_{11}^t (because $X_{1t} = f_t(Y_t)$).

Now, using Fano's inequality

$$nR \leq I(W; Y^n) + n\epsilon_n, \qquad \epsilon_n \to 0 \text{ as } n \to \infty.$$

Again, we bound the mutual information term in two ways as follows. First consider

$$\begin{aligned} I(W; Y^n) &= \sum_{t=1}^n I\left(W; Y_t | Y^{t-1}\right) \\ &\leq \sum_{t=1}^n \left(H(Y_t) - H\left(Y_t | W, Y^{t-1}\right)\right) \\ &\leq \sum_{t=1}^n \left(H(Y_t) - H\left(Y_t | W, Y^{t-1}, Y_{11}^{t-1}, X_t\right)\right) \\ &\overset{(a)}{=} \sum_{t=1}^n \left(H(Y_t) - H\left(Y_t | Y_{11}^{t-1}, X_t\right)\right) \\ &= \sum_{t=1}^n I(X_t, V_t; Y_t) \\ &= nI(X_Q, V_Q; Y_Q | Q) \leq nI(X, V; Y) \end{aligned}$$

where equality (a) follows from the fact that $(W, Y^{i-1}) \to \left(X_i, Y_{11}^{i-1}\right) \to (Y_i, Y_{1i})$ form a Markov chain, Q is a time-sharing random variable, and $V = (V_Q, Q)$, $X = X_Q$, $Y = Y_Q$, and $Y_1 = Y_{1Q}$.

Next consider

$$\begin{aligned} I(W; Y^n) &\leq I\left(W; Y^n, Y_{11}^n\right) \\ &= \sum_{t=1}^n I\left(W; Y_t, Y_{1t} | Y^{t-1}, Y_{11}^{t-1}\right) \\ &= \sum_{t=1}^n H\left(Y_t, Y_{1t} | Y^{t-1}, Y_{11}^{t-1}\right) \\ &\quad - H\left(Y_t, Y_{1t} | Y^{t-1}, Y_{11}^{t-1}, W\right) \end{aligned}$$

$$\begin{aligned}&\le \sum_{t=1}^{n} H(Y_t, Y_{1t}|V_t)\\&\qquad\qquad - H\left(Y_t, Y_{1t}|Y^{t-1}, Y_{11}^{t-1}, W, X_t\right)\\&\overset{(b)}{=} \sum_{t=1}^{n} H(Y_t, Y_{1t}|V_t) - H(Y_t, Y_{1t}|V_t, X_t)\\&= \sum_{t=1}^{n} I(X_t; Y_t, Y_{1t}|V_t) \le nI(X; Y, Y_1|V),\end{aligned}$$

where equality (b) follows from the fact that $(W, Y^{t-1}) \to (X_t, V_t) \to (Y_t, Y_{1t})$ form a Markov chain.

The bound on cardinality can be proved using the same argument as in [4, p. 310]. This completes the proof. □

Remarks:

1) Note that a similar upper bound can be proved for the classical relay channel. However, since in this case X_{1t} is a function only of V_t, the bound readily reduces to the classical cut-set bound.
2) F. Willems pointed out to the authors that the above bound can be expressed as a cut-set bound for a relay channel with an appropriately defined relay sender. Consider a relay sender alphabet $\mathcal{X}_1'$ of cardinality $|\mathcal{X}_1|^{|\mathcal{Y}_1|}$, which consists of all mappings $X_1' : \mathcal{Y}_1 \to \mathcal{X}_1'$. Then bound (14) reduces to

$$C_0 \le \max_{p(x, x_1')} \min\left\{I(X; Y, Y_1|X_1'), I(X, X_1'; Y)\right\}.$$

Note that this is analogous to the Shannon expression of the capacity of the discrete-memoryless channel with state known causally at the encoder.

B. Instantaneous Relaying Lower Bound on C_0

Note that any lower bound on the capacity of the classical relay channel, e.g., using partial decode-and-forward (2) or compress-and-forward (4), is a lower bound on the capacity of the relay-with-delay. But, one expects that higher rates can be achieved using present and future received symbols in addition to past symbols. Here we present instantaneous relaying, which is the simplest such scheme, and show that this simple relaying scheme can be optimal.

In instantaneous relaying, the relay at time t transmits a function only of the received symbol y_{1t}. Note that such a scheme is feasible for any $d \le 0$, but not for classical relay, where $d = 1$.

Using this scheme, the relay-with-delay reduces to a point-to-point discrete-memoryless channel with capacity

$$c = \max_{p(x),\, f(y_1)} I(X; Y). \qquad (15)$$

This provides a lower bound on C_d for any $d \le 0$.

In the following example we show that instantaneous relay alone can be optimal.

Example 2: In Example 1, we showed that the capacity of the Sato relay-with-unlimited-look-ahead is given by $C^*_{\text{Sato}} = \log(9/4) = 1.169925$ bits/transmission. We now show that this capacity can be achieved using only instantaneous relaying. We consider instantaneous relaying with input X pmf $(3/9, 2/9, 4/9)$ and a mapping from X to X_1 of $0 \to 0, 1 \to 1, 2 \to 1$. It can be easily shown that these choices achieve 1.169925 bits/transmission. Thus, the capacity of the Sato relay-without-delay $C_{0,\text{ Sato}} = C^*_{\text{Sato}} = 1.169925$.

Remarks:

1) This result is not too surprising. Since the channel from the sender to the relay is noiseless, complete cooperation, which requires knowledge of the entire received sequence in advance, can be simply achieved using instantaneous relaying.
2) Since $C_{0,\text{ Sato}} > C_{1,\text{ Sato}}$ and $C_{1,\text{ Sato}}$ coincides with the classical cut-set bound, this result shows that instantaneous relaying can achieve higher rates than the classical cut-set bound.

C. Partial Decode-and-Forward and Instantaneous Relaying

As mentioned earlier, any coding scheme for classical relay, which uses only past received symbols, can be used for the relay-without-delay. Further, we have seen that instantaneous relaying, which uses only the present symbol, can achieve higher rates than any of these schemes. In general, an optimal coding scheme for the relay-without-delay may need to exploit both past and present received symbols. This can be done, for example, by combining instantaneous relaying with any known scheme for classical relay. The following lower bound on C_0 is obtained by a superposition of partial decode-and-forward and instantaneous relaying. We show later that this lower bound is tight for degraded and semi-deterministic relay-without-delay channels.

Proposition 5: The capacity of the relay-without-delay channel is lower-bounded as follows:

$$\begin{aligned}C_0 \ge \max_{p(u,v,x), f(v,y_1)} \min\Big\{&I(V, X; Y), I(U; Y_1|V)\\&+ I(X; Y|V, U)\Big\}. \qquad (16)\end{aligned}$$

Sketch of the Proof: Achievability of the above bound follows by combining the partial decode-and-forward scheme and instantaneous relaying. The auxiliary random variable U represents the information decoded by the relay in the partial decode-and-forward scheme and V represents the information sent cooperatively by both the sender and the relay to help the receiver decode the previous U. The proof of achievability follows the same lines as that for the partial decode-and-forward scheme in [2]. We therefore only provide an outline.

Random Code Generation: Fix a joint pmf $p(u, v, x)$ and a function $f(v, y_1)$. Each message $w \in [1, 2^{nR}]$ is represented by an index pair (j, k), where $j \in [1, 2^{nR_1}]$ and $k \in [1, 2^{nR_2}]$. Thus, $R = R_1 + R_2$. Generate 2^{nR_0} independent and identically distributed (i.i.d.) sequences $v^n(i)$ for $i \in [1, 2^{nR_0}]$ according to $\prod_{t=1}^{n} p(v_t)$. For each $v^n(i)$, generate 2^{nR_1} i.i.d. sequences $u^n(i, j)$ for $j \in [1, 2^{nR_1}]$ according to $\prod_{t=1}^{n} p(u_t|v_t)$. Randomly partition the set of indices $[1, 2^{nR_1}]$ into 2^{nR_0} bins. For each $(v^n(i), u^n(i, j))$, generate 2^{nR_2} i.i.d. sequences $x^n(i, j, k)$ for $k \in [1, 2^{nR_2}]$ according to $\prod_{t=1}^{n} p(x_t|v_t, u_t)$.

Encoding: A block Markov coding scheme is used as in the partial decode-and-forward scheme. To send w_b in block

$b = 1, 2, \ldots, B$, the sender sends $x^n(i_b, j_b, k_b)$, where i_b is the bin index for j_{b-1}. We assume that the relay has successfully decoded j_{b-1} at the end of block $b-1$ and thus knows its bin index i_b. At time $t = 1, 2, \ldots, n$, the relay sends $x_{1t} = f(v_{1t}(i_b), y_{1t})$.

Decoding: Upon receiving $y_1^n(b)$ at the end of block b, the relay decodes j_b. This can be reliably done provided

$$R_1 < I(U; Y_1|V).$$

The receiver first decodes i_b, which can be done provided

$$R_0 < I(V; Y).$$

The decoder then decodes j_{b-1}, which can be done provided

$$R_1 < I(U; Y|V) + R_0.$$

Finally, the decoder decodes k_{b-1} (and thus the message w_{b-1} sent in block $b-1$). This can be done provided

$$R_2 < I(X; Y|U, V).$$

Combining the above inequalities shows that any

$$R < \min\{I(V, X; Y), I(U; Y_1|V) + I(X; Y|V, U)\}$$

is achievable. This completes the outline of the proof of Proposition 5.

Remarks:

1) If $V = X_1$, the bound reduces to partial decode-and-forward. If $U = X$, the scheme reduces to a superposition of decode-and-forward and instantaneous relaying, which gives the looser bound

$$C_0 \geq \max_{p(v,x), f(v,y_1)} \min\{I(V, X; Y), I(X; Y_1|V)\}. \quad (17)$$

Also, if $U = V \in \emptyset$, the bound reduces to (15), which is achieved using instantaneous relaying only.

2) One can similarly combine compress-and-forward with time sharing for the classical relay channel [10] and instantaneous relaying to establish the lower bound on

$$C_0 \geq \max_{p(q)p(v,x|q)p(\hat{y}_1|y_1,v,q), f(y_1,v,q)} \min\{I(X; Y, \hat{Y}_1|V, Q), \\ I(X, V; Y|Q) - I(\hat{Y}_1; Y_1|V, X, Y, Q)\}. \quad (18)$$

D. Capacity Theorems for Relay-Without-Delay

We show that the lower bound obtained using superposition of instantaneous relaying and partial decode-and-forward is tight for degraded relay-without-delay channels.

Proposition 6: The capacity of the degraded relay-without-delay channel is given by

$$C_{0,\,\text{deg}} = \max_{p(v,x), f(v,y_1)} \min\{I(V, X; Y), I(X; Y_1|V)\} \quad (19)$$

where $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{X}_1| + 1$.

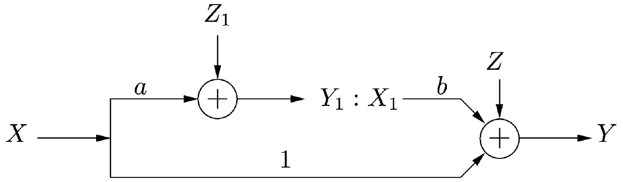

Fig. 4. AWGN relay channel.

Proof: Achievability follows from Proposition 5 by setting $U = X$, i.e., using decode-and-forward with instantaneous relaying. The converse follows from the upper bound of Theorem 2. Consider the second mutual information term in (14)

$$\begin{aligned} I(X; Y, Y_1|V) &= I(X; Y_1|V) + I(X; Y|Y_1, V) \\ &\overset{(a)}{=} I(X; Y_1|V). \end{aligned}$$

Equality (a) follows by the definition of degradedness, $(X \rightarrow (X_1, Y_1) \rightarrow Y)$, and the fact that $x_1 = f(v, y_1)$. This completes the proof of the proposition. □

The following shows that superposition of decode-and-forward and instantaneous relaying is also optimal for semi-deterministic relay-without-delay channels.

Proposition 7: The capacity of the semi-deterministic relay-without-delay channel is given by is

$$C_{0,\,\text{semi--det}} = \max_{p(v,x), f(v,y_1)} \min\{I(V, X; Y), I(X; Y, Y_1|V)\}.$$

Proof: Achievability follows from Proposition 5. To show this, set $U = Y_1$ in the achievable rate (16). Consider the second term under the $\min$

$$\begin{aligned} I(U; Y_1|V) + I(X; Y|U, V) &= H(Y_1|V) + I(X; Y|Y_1, V) \\ &= H(Y_1|V) + H(Y|Y_1, V) \\ &\quad - H(Y|Y_1, V, X) \\ &= H(Y, Y_1|V) - H(Y, Y_1|Y_1, V, X) \\ &\overset{(a)}{=} H(Y, Y_1|V) - H(Y, Y_1|V, X) \\ &= I(X; Y, Y_1|V). \end{aligned}$$

Step (a) follows from the fact that $Y_1 = g(X)$.

The converse follows from the upper bound of Theorem 2. This completes the proof of the proposition. □

V. AWGN Relay-With-Delay

Consider the general "full-duplex" AWGN relay channel model in Fig. 4. The parameters a and b are the path gains for the channels from X to Y_1 and X_1 to Y, respectively, normalized with respect to the gain of the channel from X to Y, which is set equal to 1. The AWGN components $Z_1 \sim \mathcal{N}(0, N)$ and $Z \sim \mathcal{N}(0, N)$ are assumed to be independent. Further, we assume an average power constraint P on each of the sender X and relay sender X_1.

To make our exposition self-contained, we first summarize relevant known results on the capacity of the classical case, where $d = 1$.

1) The capacity of the classical AWGN relay channel is not known for any nonzero a and b.

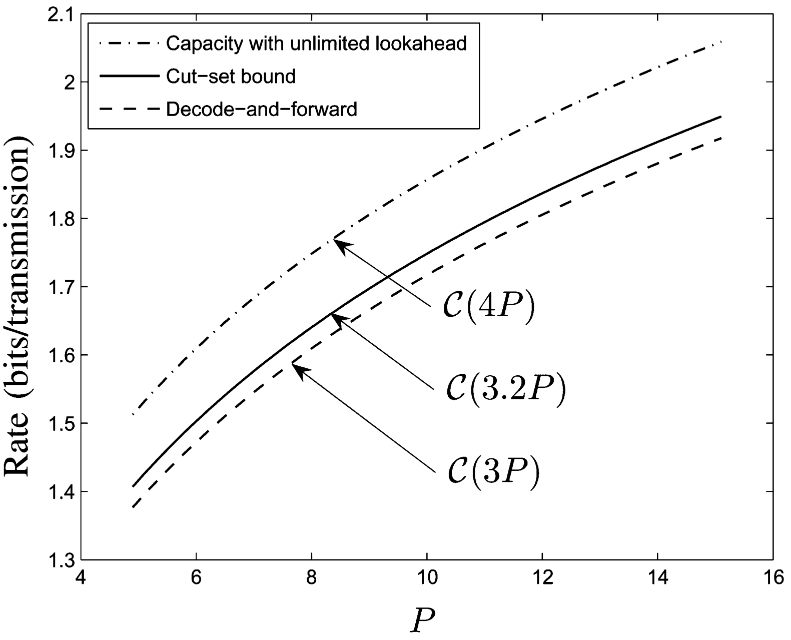

Fig. 5. Plot of capacity for unlimited look-ahead, the classical cut-set bound (20), and decode-and-forward versus P, for $N = 1$, $a = 2$, and $b = 1$.

2) Classical cut-set bound [10]:

$$C_{1,\text{ AWGN}} \le \begin{cases} \mathcal{C}\left(\frac{\left(ab+\sqrt{1+a^2-b^2}\right)^2 P}{(1+a^2)N}\right), & \text{if } a > b \\ \mathcal{C}\left(\frac{(1+a^2)P}{N}\right), & \text{otherwise} \end{cases} \tag{20}$$

where we use the conventional notation $\mathcal{C}(x) = 1/2\log_2(1+x)$.

3) Decode-and-forward [10]:

It was shown in [10] that partial decode-and-forward reduces to decode-and-forward and gives the lower bound

$$C_{1,\text{ AWGN}} \ge \begin{cases} \left(\frac{\left(b\sqrt{(a^2-1)}+\sqrt{a^2-b^2}\right)^2 P}{a^2N}\right), & \text{if } \frac{a^2}{1+b^2} \ge 1 \\ \mathcal{C}\left(\frac{\max\{1,a^2\}P}{N}\right), & \text{otherwise.} \end{cases} \tag{21}$$

Note that this bound is never equal to (20) for any finite a and b. Further the bound reduces to $\mathcal{C}(P/N)$, i.e., the capacity of the direct channel, when $a \le 1$.

Now, consider the AWGN relay-with-delay for $d \le 0$. It is straightforward to show that the upper bound on the capacity of the relay-with-unlimited-look-ahead (7) is achieved by Gaussian (X, X_1)[3] and thus reduces to

$$C^*_{\text{AWGN}} \le \max_{0\le\rho\le 1} \min\left\{\mathcal{C}\left(\frac{(1+b^2+2b\rho)P}{N}\right),\right.$$
$$\left.\mathcal{C}\left(\frac{\left(1+a^2-N\rho^2/(N+a^2P(1-\rho^2))\right)P}{N}\right)\right\} \tag{22}$$

where $\rho \in [0,1]$.

In Sections V-A and -B, we discuss lower bounds for the relay-with-unlimited-look-ahead and the relay-without-delay and show that they are tight in some cases.

A. AWGN Relay-With-Unlimited-Look-Ahead

Here we show that the lower bound in (10) coincides with the upper bound (22) when the channel from the sender to the relay is sufficiently stronger than the other two channels.

Proposition 8: The capacity of the AWGN relay-with-unlimited-look-ahead for $a \ge 1+b$ is given by

$$C^*_{\text{AWGN}} = \mathcal{C}\left(\frac{(1+b)^2P}{N}\right). \tag{23}$$

Proof: First we evaluate the lower bound in (10) for the AWGN relay channel using Gaussian (X, X_1) to obtain

$$C^*_{\text{AWGN}} \ge \min\left\{\mathcal{C}\left(\frac{(1+b)^2P}{N}\right), \mathcal{C}\left(\frac{a^2P}{N}\right)\right\}.$$

Note that for $a \le 1+b$, the second term is smaller than the first.

Next consider the upper bound (22). If $a > b$, the first term is smaller than the second and is maximized when $\rho = 1$. Thus, we have

$$C^*_{\text{AWGN}} \le \mathcal{C}\left(\frac{(1+b)^2)P}{N}\right).$$

This coincides with the lower bound, which completes the proof. □

Fig. 5 compares the capacity for unlimited look-ahead (where complete cooperation is achieved), decode-and-forward, where delayed cooperation is used, and the classical cut-set upper bound for $N = 1$, $a = 2$, and $b = 1$. The gap between decode-and-forward and the capacity in the unlimited look-ahead case (which is the maximum achievable rate for any d) represents the highest potential increase in rate by utilizing look-ahead.

[3] In contrast, the distribution on (V, X, X_1) that maximizes the tighter bound on the capacity of the relay-without-delay (14) is not known (and we do not believe that it is in general Gaussian).

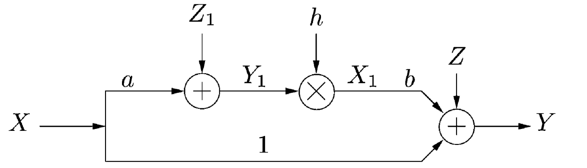

Fig. 6. AWGN relay-without-delay with amplify-and-forward.

B. AWGN Relay-Without-Delay

In the following, we show that the capacity of the AWGN relay-without-delay is known when the channel from the sender to the relay is sufficiently weak. Note that this is in contrast to AWGN classical relay, where capacity is not known for any finite a and b.

Proposition 9: The capacity of the AWGN relay-without-delay for $a^2 \le b^2 \min\{1, P/(a^2P+N)\}$ is given by

$$C_{0,\text{ AWGN}} = \mathcal{C}\left(\frac{(1+a^2)P}{N}\right). \tag{24}$$

Proof: Note that if $a \le b$, then the second term in (22) becomes smaller than the first. Further, it is easy to see that the difference between the first and second terms in this case is a strictly increasing function in $\rho \in [0,1]$. Thus, setting $\rho = 0$ maximizes the second term and we obtain

$$C_{0,\text{ AWGN}} \le \mathcal{C}\left(\frac{(1+a^2)P}{N}\right), \qquad \text{for } a \le b. \tag{25}$$

To obtain a lower bound on capacity, consider the following simple "amplify-and-forward" achievability scheme, which is a special case of instantaneous relaying (see Fig. 6). For time $t = 1, 2, \ldots$, set $x_{1t} = hy_{1t}$ for some constant gain parameter h. With this substitution, the channel reduces to the point-to-point AWGN channel

$$Y = bh(aX + Z_1) + X + Z. \tag{26}$$

Applying the average power constraint on X_1 (with equality) gives

$$h^2 = \frac{P}{a^2P+N}.$$

Substituting, in (26) and maximizing the capacity of the equivalent point-to-point AWGN channel gives

$$h^* = \frac{a}{b}, \qquad \text{for } a^2 \le b^2 \min \frac{P}{a^2P+N}.$$

Interestingly, the maximized capacity using h^* for $a^2 \le b^2(P/(a^2P+N)$ is given by $\mathcal{C}((1+a^2)P/N)$. If, in addition $a \le b$, i.e., $a^2 \le b^2 \min\{1, P/(a^2P+N)\}$, then this achievable rate coincides with the upper bound (25), which completes the proof. □

Remarks:

1) Because $C_{0,\text{ AWGN}}$ coincides with the upper bound on C^*_{AWGN} for $a^2 \le b^2 \min\{1, P/(a^2P+N)\}$, under this condition $C_{d,\text{ AWGN}} = C_{0,\text{ AWGN}}$ for all $d \le 0$.
2) Note that $C_{0,\text{ AWGN}}$ coincides with the classical cut-set bound (20) for $a^2 \le b^2 \min\{1, P/(a^2P+N)\}$. Thus, the classical cut-set bound can be achieved when $d \le 0$.
3) From the above result and Proposition 8, we know C^* for all parameter values except for the very small parameter range $b\sqrt{\min\{1, P/(a^2P+N)\}} < a < 1+b$.

We now show that the capacity of the AWGN relay-without-delay can be strictly higher than the classical cut-set bound.

Example 3: Consider the following special case of the scheme involving superposition of decode-and-forward encoding with instantaneous relaying, which achieves the lower bound (16). Let $U = X = V + X'$, where $V \sim \mathcal{N}(0, \alpha P)$ and $X' \sim \mathcal{N}(0, \overline{\alpha} P)$ are independent, where $0 \le \alpha \le 1$ and $\overline{\alpha} = 1 - \alpha$, and choose X_1 as a normalized convex combination of Y_1 and V of the form

$$X_1 = h(\beta Y_1 + \overline{\beta} V) = h((\overline{\beta} + a\beta)V + \beta(aX' + Z_1))$$

where $0 \le \beta \le 1$ and h is a normalizing parameter.

Using the power constraint on the relay sender X_1 with equality, we obtain

$$h^2 = \frac{P}{(\overline{\beta} + a\beta)^2 \alpha P + a^2\beta^2\overline{\alpha}P + \beta^2 N}.$$

Substituting the above choice of distribution on (V, X, X_1) into the expression (16) gives

$$C_{0,\text{ AWGN}} \ge \max_{\alpha,\,\beta} \min\left\{ \mathcal{C}\left(\frac{a^2\overline{\alpha}P}{N}\right), \mathcal{C}\left(\frac{\alpha P(bh(a\beta+\overline{\beta})+1)^2 + \overline{\alpha}P(bha\beta+1)^2}{(1+(\beta bh)^2)\,N}\right)\right\}. \tag{27}$$

Fig. 7 compares the above achievable rate to the classical cut-set bound (20) and the general upper bound on the capacity of relay-with-delay (22) for $N = 1$, $a = 2$, and $b = 1$.

Thus, the capacity of the AWGN relay-with-delay can be strictly larger than the capacity of its classical counterpart for $d \le 0$.

VI. Relay Networks With Delays

In this section, we investigate the effect of link delays on capacity for feedforward relay networks, i.e., relay networks with no feedback, which we refer to as relay networks with delays (see Fig. 8). We first introduce some graph theoretic notation needed for its definition. Consider a weighted, directed acyclic graph (DAG) $(\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{1, 2, \ldots, K\}$ is the set of nodes (vertices) and $\mathcal{E}$ is the set of directed edges. We assume that node 1 has only outgoing edges and node K has only incoming edges. To prevent trivialities, we further assume that each node lies on a path from 1 to K. The weight of an edge $(i,j) \in \mathcal{E}$ is denoted by $d(i,j) \in \{0, 1, \ldots\}$. We need the following definitions:

1) $\mathcal{N}_i = \{j \in \mathcal{N} : (j,i) \in \mathcal{E}\}$ for $i \in \mathcal{N}$, i.e., $\mathcal{N}_i$ is the set of nodes with edges incident on i;
2) $x_t(\mathcal{N}_i) = \{x_{j(t-d(j,i))} : j \in \mathcal{N}_i\}$;
3) $d(i)$ is the weight of a minimum-weight-path from node 1 to node i for $i \in \mathcal{N}$; and

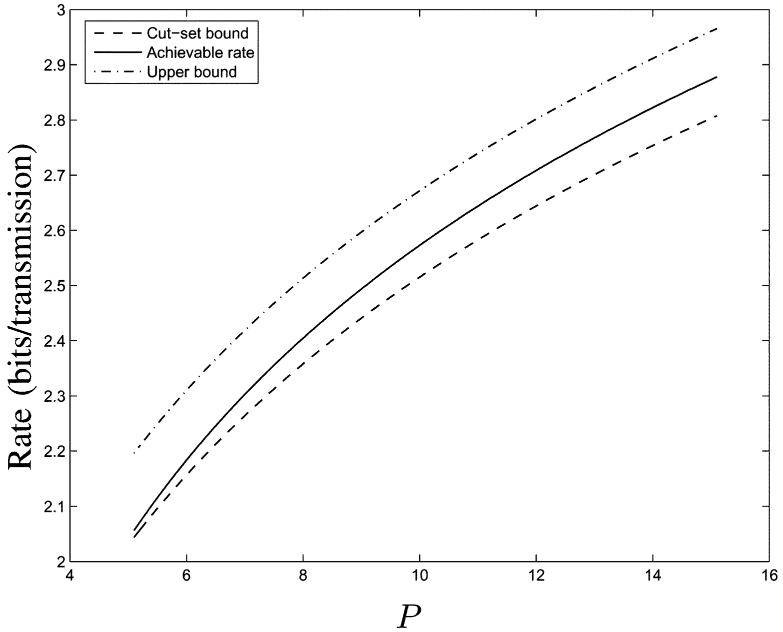

Fig. 7. Plot of the achievable rate for the AWGN relay-without-delay (27), the classical cut-set bound (20), and the upper bound on the capacity of the AWGN relay-with-unlimited-look-ahead (22) versus P, for $N = 1$, $a = 2$, and $b = 1$.

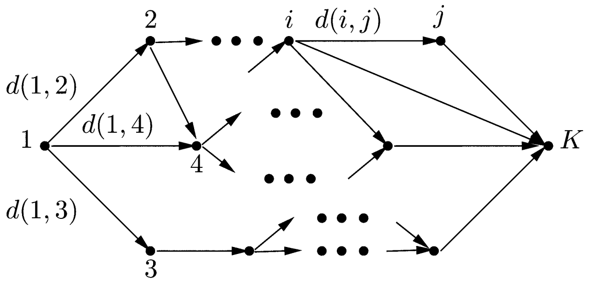

Fig. 8. Relay network with delays.

4) $D(i)$ is the weight of a maximum-weight-path from node 1 to node i for $i \in \mathcal{N}$.

In the definitions that follow, an edge corresponds to a communication link, a graph corresponds to a communication network, and the weight of an edge corresponds to the delay on that communication link. Hence, we refer to the weight of a path as the delay of a path.

A discrete-memoryless relay network with delays consists of: i) a weighted DAG $(\mathcal{N}, \mathcal{E})$, $\mathcal{N} = \{1, 2, \dots, K\}$, where node 1 is the sender and node K is the receiver and the rest are relay nodes, ii) a set of symbols taking values in finite alphabets associated with each node, where $x_1 \in \mathcal{X}_1$ is associated with the sender node 1, $y_K \in \mathcal{Y}_K$ is associated with the receiver node K, and $x_i \in \mathcal{X}_i$, $y_i \in \mathcal{Y}_i$ are associated with relay sender–receiver pair i, and iii) a family of conditional pmfs $\{p(y_i|\{x_k : k \in \mathcal{N}_i\}), i = 2, 3, \dots, K\}$.[4]

The weight $d(j,i)$ assigned to edge (j,i) corresponds to the delay on that link in the sense that the received symbol at node i at time t depends on the transmitted symbol at node j at time $t - d(j,i)$. The pmf of the received symbol at node i at time t depends only on the transmitted symbols at the nodes with edges incident on node i and is given by $p(y_{it}|\{x_{j(t-d(j,i))} : j \in \mathcal{N}_i\})$, which can be written as $p(y_{it}|x_t(\mathcal{N}_i))$ using the shorthand notation defined earlier.

The network structure results in a factorization of the conditional pmf of the received symbols at any set of nodes $S \subset \mathcal{N}$ and time indexes $\{t_i \geq 1 : i \in S\}$ of the form

$$p\left(\left\{y_{i(t_i)} : i \in S\right\} \middle| \cup_{i \in S} x_{t_i}(\mathcal{N}_i)\right) = \prod_{i \in S} p\left(y_{i(t_i)} \middle| x_{t_i}(\mathcal{N}_i)\right). \tag{28}$$

The network is memoryless in the sense that at any node i and for any $n \geq D(i)$

$$p\left(y^n_{i(D(i)+1)} \middle| \cup^n_{t=D(i)+1} x_t(\mathcal{N}_i)\right) = \prod_{t=D(i)+1}^{n} p\left(y_{it} \middle| x_t(\mathcal{N}_i)\right). \tag{29}$$

The received symbol, y_{it}, is considered to be arbitrary for $t \leq D(i)$.

A $(2^{nR}, n)$ code for the discrete-memoryless relay network with delays consists of: i) a set of messages $\{1, 2, \dots, 2^{nR}\}$, ii) an encoding function that maps each message w into a codeword $x^n_{11}(w)$ of length n, iii) relay encoding functions $x_{it} = f_{it}\left(y^t_{i(D(i)+1)}\right)$ for $i = 2, \dots, K-1$ and $t \geq D(i) + 1$, and iv) a decoding function that maps each received sequence $y^{d(K)+n}_{K(D(K)+1)}$ into an estimate $\hat{w}\left(y^{d(K)+n}_{K(D(K)+1)}\right)$.

A rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes with $P^{(n)}_e = \mathrm{P}\{\hat{W} \neq W\} \to 0$ as $n \to \infty$. The network capacity, C, is defined as the supremum of achievable rates.

Remark: Note that there is no loss of generality in assuming d to be nonnegative, since by Theorem 3 below, only relative path delays matter for capacity. Thus, we can always convert a network with arbitrary link delays to one with the same capacity that has only nonnegative delays.

[4] Unlike our definition for the relay-with-delay, here we do not condition on any y variables for simplicity of exposition.

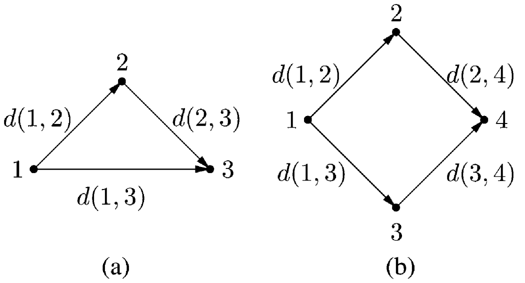

Fig. 9. Example relay network with delays. (a) Relay-with-delay. (b) Two-relay network with delays.

To help understand these definitions, consider the following illustrative examples:

Example 4 (Relay-Without-Delay): Here $\mathcal{N} = \{1,2,3\}$ and $\mathcal{E} = \{(1,2),(2,3),(1,3)\}$ as in Fig. 9(a). The delays are $d(1,2) = d(2,3) = d(1,3) = 0$ and thus, $d(2) = d(3) = 0$. Furthermore, $D(2) = D(3) = 0$. The conditional pmfs are of the form

$$p(y_{2t}y_{3t}|x_{1t}x_{2t}) = p(y_{2t}|x_{1t})p(y_{3t}|x_{1t}x_{2t}).$$

Example 5 (Classical Relay): Here $\mathcal{N} = \{1,2,3\}$ and $\mathcal{E} = \{(1,2),(2,3),(1,3)\}$ as in Fig. 9(a). The delays are $d(1,2) = 1$, $d(2,3) = d(1,3) = 0$, and thus, $d(2) = 1$ and $d(3) = 0$. Furthermore, $D(2) = D(3) = 1$.

The conditional pmfs are of the form

$$p(y_{2(t+1)}y_{3t}|x_{1t},x_{2t}) = p(y_{2(t+1)}|x_{1t})p(y_{3t}|x_{1t}x_{2t}).$$

Example 6 (Two-Relay Network): Fig. 9(b) depicts the two-relay network. Here $\mathcal{N} = \{1,2,3,4\}$ and $\mathcal{E} = \{(1,2),(1,3),(2,4),(3,4)\}$. The delays are $d(1,2) = 2$, $d(3,4) = 1$, $d(1,3) = d(2,4) = 0$, and thus, $d(2) = 2$, $d(3) = 0$, and $d(4) = 1$. Furthermore, $D(2) = 2$, $D(3) = 0$, and $D(4) = 2$.

The conditional pmfs are of the form

$$\begin{aligned} &p(y_{2(t+2)}y_{3t}y_{4(t+1)}|x_{1t}x_{2(t+1)}x_{3t}) \\ &\quad = p(y_{2(t+2)}|x_{1t})p(y_{3t}|x_{1t})p(y_{4(t+1)}|x_{2(t+1)}x_{3t}). \end{aligned}$$

In Sections VI-A and -B we present two results. The first concerns the question of when does delay not change the capacity of a network. The second result is an upper bound on capacity that generalizes both the cut-set bound for relay networks [5] and the upper bound in Theorem 2 for the relay-without-delay. This bound involves the use of auxiliary random variables and multiple random variables per sender.

A. Only Relative Delays Matter

In this subsection, we show that as far as network capacity is concerned, only relative path delays matter.

Theorem 3: Consider two relay networks with delays with the same directed acyclic graph (DAG) $(\mathcal{N},\mathcal{E})$ and the same associated sets of symbols and conditional pmfs but different link delays $\{d_1(i,j)\}$ and $\{d_2(i,j)\}$. Let $w_i(p)$, $i = 1,2$, be the delay of path p from the sender to the receiver. If there exists an integer a such that $w_1(p) - w_2(p) = a$ for every path p, i.e., if all paths in both networks have the same relative delays, then the two networks have the same capacity.

Proof: We use $\mathcal{R}_1$ and $\mathcal{R}_2$ to refer to the two relay networks. We denote the transmitted and received symbols in $\mathcal{R}_1$ by x and y, respectively, and those in $\mathcal{R}_2$ by $\tilde{x}$ and $\tilde{y}$. We use the subscript $s = 1,2$ to refer to probabilities and quantities such as $d(i)$, $D(i)$ in network $\mathcal{R}_s$.

Suppose we are given a $(2^{nR}, n)$ code for network $\mathcal{R}_1$. This includes i) the encoding function $x_1^n : \{1,\ldots,2^{nR}\} \to \mathcal{X}_1^n$, ii) relay encoding functions $x_{it} = f_{it}\left(y_{i1}^t\right)$ for $i \in \{2,\ldots,K-1\}$ and all t with the convention that the function f_{it} takes an arbitrary value when $t < D_1(i)+1$, and iii) a decoding function

$$\hat{w} : \mathcal{Y}_K^{n+d_1(K)-D_1(K)} \to \{1,\ldots,2^{nR}\}$$

that uses the received sequence $y_{K(D_1(K)+1)}^{d_1(K)+n}$.

The decoder uses $y_{K(D_1(K)+1)}^{d_1(K)+n}$ to decode the index that was sent using the codeword x_1^n sent from node 1. Hence, for this code, the probability of error in network $\mathcal{R}_1$ depends only on the following conditional probability:

$$p_1\left(y_{K(D_1(K)+1)}^{d_1(K)+n}|x_{11}^n\right).$$

Let $l(i)$ be the minimum of the delays on paths from node i to node K and define $l(K) = 0$. Then using (45) in the Appendix, the above conditional probability can be written as follows. For network $\mathcal{R}_s$, $s = 1,2$, see (30) at the bottom of the page.

For $t = 1,2,\ldots$, let

$$\tilde{x}_{i(t+D_2(i))} = x_{i(t+D_1(i))} \tag{31}$$

$$\tilde{y}_{i(t+D_2(i))} = y_{i(t+D_1(i))}. \tag{32}$$

Next we construct a $(2^{nR}, n)$ code for network $\mathcal{R}_2$ using the one for $\mathcal{R}_1$ with the same probability of error. This code for $\mathcal{R}_2$ uses the same encoding function as that for $\mathcal{R}_1$. The decoding function is also the same except that it uses the received sequence $\tilde{y}_{K(D_2(K)+1)}^{d_2(K)+n}$. The relay encoding function at node i at time t is denoted by f_{it} in $\mathcal{R}_1$ and by $\tilde{f}_{it}$ in $\mathcal{R}_2$. The relay encoding functions for $t = 1,2,\ldots$ in $\mathcal{R}_2$ are chosen as follows:

$$\tilde{f}_{i(D_2(i)+t)}\left(\tilde{y}_{i(D_2(i)+1)}^{D_2(i)+t}\right) = f_{i(D_1(i)+t)}\left(y_{i(D_1(i)+1)}^{D_1(i)+t}\right) \tag{33}$$

and both functions take the same arbitrary value when $t < 1$.

$$p_s\left(y_{K(D_s(K)+1)}^{d_s(K)+n}|x_{11}^n\right) = \sum_{\{x_i,y_i,i=2,\ldots,K-1\}} \prod_{i\in\mathcal{N}-\{1\}} \prod_{t=D_s(i)+1}^{d_s(K)+n-l_s(i)} p_s(y_{it}|x_t(\mathcal{N}_i))p_s\left(x_{it}|y_{i1}^t\right). \tag{30}$$

Let $I(A)$ denote the indicator function of event A. Consider any node $i \in \{2,\ldots,K-1\}$. Then

$$\begin{aligned} p_1&\left(x_{i(D_1(i)+t)}\middle| y_i^{D_1(i)+t}\right)\\ &= I\left(f_{i(D_1(i)+t)}\left(y_{i(D_1(i)+1)}^{D_1(i)+t}\right) = x_{i(D_1(i)+t)}\right)\\ &= I\left(\tilde{f}_{i(D_2(i)+t)}\left(\tilde{y}_{i(D_2(i)+1)}^{D_2(i)+t}\right) = \tilde{x}_{i(D_2(i)+t)}\right) \quad (34)\\ &= p_2\left(\tilde{x}_{i(D_2(i)+t)}\middle| \tilde{y}_i^{D_2(i)+t}\right). \quad (35) \end{aligned}$$

Equality holds in (34) because of (31) and the relationship between the relay encoding functions in $\mathcal{R}_1$ and $\mathcal{R}_2$ given by (33).

This shows that the terms of the form $p\left(x_{it}|y_i^t\right)$ in (30) are equal for both networks $\mathcal{R}_1$ and $\mathcal{R}_2$. Next we show that terms of the form $p(y_{it}|x_t(\mathcal{N}_i))$ are also equal for both networks. To show this, we proceed as follows. Let $f_1(i)$ and $g_1(i)$ be the delays of two paths from node 1 to node i in network $\mathcal{R}_1$ and let $f_2(i)$ and $g_2(i)$ be the corresponding quantities in $\mathcal{R}_2$. We claim that

$$f_1(i) - g_1(i) = f_2(i) - g_2(i). \quad (36)$$

If node i has only one incoming edge then the above equation is trivially true since both sides are equal to 0. Suppose node i has two or more incoming edges and that the above equation does not hold. Then there are two distinct paths from node 1 to node i such that the $f_1(i) - f_2(i) \neq g_1(i) - g_2(i)$. Since each node lies on a path from node 1 to node K, it follows that there are two distinct paths from node 1 to node K such that they do not have the same relative delay in both networks. This contradicts our hypothesis and hence the claim in (36) holds.

Observe that $D_1(j) + d_1(j,i)$ is the delay on a path from node 1 to node i in $\mathcal{R}_1$ and $D_2(j) + d_2(j,i)$ is the delay of the same path in $\mathcal{R}_2$. Further, due to our hypothesis, the maximum delay path from node 1 to any node i is the same in both networks. Hence, it follows from (36) that

$$D_1(i) - D_1(j) - d_1(j,i) = D_2(i) - D_2(j) - d_2(j,i). \quad (37)$$

Now for any node i and $t = 1,2,\ldots$, we have

$$\begin{aligned} p_1&\left(y_{i(D_1(i)+t)}\middle| x_t(\mathcal{N}_i)\right)\\ &= \mathrm{P}\Big(Y_{i(D_1(i)+t)} = y_{i(D_1(i)+t)}\Big|\\ &\qquad \{X_{j(D_1(i)+t-d_1(j,i))} = x_{j(D_1(i)+t-d_1(j,i))} : j \in \mathcal{N}_i\}\Big)\\ &= \mathrm{P}\Big(Y_{i(D_1(i)+t)} = y_{i(D_1(i)+t)}\Big|\\ &\qquad \{X_{j(D_1(i)+t-d_1(j,i))} = x_{j(D_1(i)+t-d_1(j,i))} : j \in \mathcal{N}_i\}\Big)\\ &= \mathrm{P}\Big(\tilde{Y}_{i(D_2(i)+t)} = \tilde{y}_{i(D_2(i)+t)}\Big|\\ &\qquad \{\tilde{X}_{j(D_1(i)+t-d_1(j,i))}\\ &\qquad = \tilde{x}_{j(D_1(i)+t-d_1(j,i)+D_2(j)-D_1(j))} : j \in \mathcal{N}_i\}\Big) \quad (38)\\ &= \mathrm{P}\Big(\tilde{Y}_{i(D_2(i)+t)} = \tilde{y}_{i(D_2(i)+t)}\Big|\\ &\qquad \{\tilde{X}_{j(D_1(i)+t-d_1(j,i))} = \tilde{x}_{j(D_2(i)+t-d_2(j,i))} : j \in \mathcal{N}_i\}\Big) \quad (39)\\ &= p_2\left(\tilde{y}_{i(D_2(i)+t)}\middle| \tilde{x}_t(\mathcal{N}_i)\right) \quad (40) \end{aligned}$$

where (38) follows from (31) and (32), and (39) holds due to (37).

Using (35) and (40) for the factorization as in (30) for $\mathcal{R}_1$ and $\mathcal{R}_2$, it follows that

$$p_1\left(y_{K(D_1(K)+1)}^{d_1(K)+n}|x_{11}^n\right) = p_2\left(\tilde{y}_{K(D_2(K)+1)}^{d_2(K)+n}|\tilde{x}_{11}^n\right).$$

Thus, we have shown that using the code that achieves rate R in $\mathcal{R}_1$, we can obtain a code for $\mathcal{R}_2$ that achieves the same rate. The same argument implies that any rate that can be achieved in $\mathcal{R}_2$ can also be achieved in $\mathcal{R}_1$. As a result both networks have the same capacity. □

Remark: The condition in the theorem can be checked in $O(|\mathcal{E}|)$ time, even though the theorem appears to require checking *all* path delays from node 1 to node K. We provide a brief description of an algorithm for checking this condition and leave it to the reader to verify the details.

Note that since the relay network with delays consists of a DAG, there exists an ordering of the nodes, $i_1 = 1, i_2, \ldots, i_K = K$, such that $\mathcal{N}_{i_j} \subset \{i_1,\ldots,i_{j-1}\}$. In other words, there exists an ordering such that all incoming edges to any node are from previous nodes in the ordering. Consider the subset of nodes that have more than one incoming edge. If this subset is empty there is only one path in the network. Otherwise, consider each vertex i in this subset according to the above ordering and check if the relative delay of the shortest path from node 1 through each incoming edge is the same in both networks. It can be verified that the condition in the theorem is satisfied if and only if the above check is satisfied at each node of this subset.

We now demonstrate the implications of the above theorem for some simple networks.

Example 7: Consider the relay channel in Fig. 9(a). There are two paths from the sender to the receiver. The direct path has delay $d(1,3)$ and the other path has delay $d(1,2) + d(2,3)$. Hence, by Theorem 3, the capacity of this network remains unchanged if $d(1,2)$ is replaced by $d = d(1,2) + d(2,3) - d(1,3)$ and $d(2,3)$ and $d(1,3)$ are set to 0.

In particular, if $d(1,2) = d(2,3) = 1$, $d(2,3) = 2$, the relative delays are the same as that for the relay-without-delay, in which there are two paths each with delay 0. The above result implies that both these networks have the same capacity.

Example 8: Consider two two-relay networks $\mathcal{R}_1$ and $\mathcal{R}_2$ having the same DAG and conditional pmfs as shown in Fig. 9(b). Let the delays for $\mathcal{R}_1$ be $d(1,2) = 0$, $d(1,3) = 1$, $d(2,4) = 0$, $d(3,4) = 2$, and for $\mathcal{R}_2$ be $d(1,2) = 1$, $d(1,3) = 4$, $d(2,4) = 1$, $d(3,4) = 1$. Both networks have two paths from the sender to the receiver with path delays $(0,3)$ in $\mathcal{R}_1$ and $(2,5)$ in $\mathcal{R}_2$, respectively. Since the relative delays in both networks are the same, the above result implies that $\mathcal{R}_1$ and $\mathcal{R}_2$ have the same capacity.

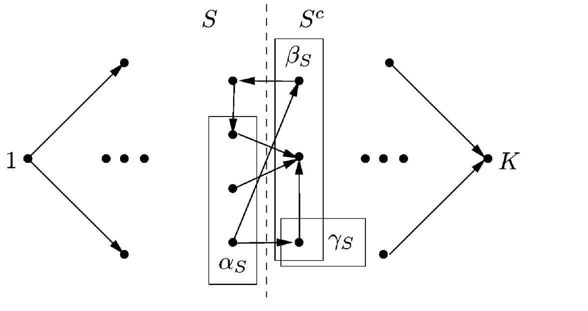

Fig. 10. An illustration of the sets $\alpha_S, \beta_S, \gamma_S$ for some $S \in \mathbb{S}$. Set α consists of nodes in S with outgoing edges to nodes in S^c. Set β consists of nodes in S^c with incoming edges from nodes in S. Set γ consists of nodes in β with outgoing edges to other nodes in β.

B. Cut-Set Bound for Relay Networks With Delays

The cut-set bound in [2] provides an upper bound on the capacity of the classical relay channel. Theorem 2 provides an upper bound on the capacity of the relay-without-delay. The following upper bound on the capacity of a relay network with delays is a generalization of both these upper bounds.

Recall that the classical cut-set bound has terms of the form $I(X(S); Y(S^c)|X(S^c))$, where S is a subset of $\mathcal{N}$ that contains node 1 and does not contain node K. Our graphical structure allows a refinement so that $X(S)$ is replaced by $X(\alpha_S)$, $Y(S^c)$ is replaced by $Y(\beta_S)$, and $X(S^c)$ is replaced by $X(\gamma_S)$, where the sets $\alpha_S, \beta_S, \gamma_S$ are as shown in Fig. 10. The introduction of arbitrary delays complicates the expression so that the X variables for some nodes have to be replaced by auxiliary variables called U. This leads to a cut-set bound with terms of the form shown in Theorem 4. To state the theorem precisely, we begin with some needed notation.

1) Define $\mathbb{S}$ to be a set of subsets defined as $\mathbb{S} = \{S \subset \mathcal{N} : 1 \in S,\ K \in S^c\}$.
   (a) Given $S \in \mathbb{S}$, define $\alpha_S = \{i : (i,j) \in \mathcal{E}, i \in S, j \in S^c\}$, i.e., α_S is the set of nodes in S with outgoing edges to nodes in S^c.
   (b) Similarly, define $\beta_S = \{j : (i,j) \in \mathcal{E}, i \in S, j \in S^c\}$, i.e., β_S is the set of nodes in S^c with incoming edges from nodes in S.
   (c) Define $\gamma_S = \{i : (i,j) \in \mathcal{E},\ i, j \in \beta_S\}$. That is, γ_S is the set of nodes in β_S with outgoing edges to other nodes in β_S.

   Note that $\alpha_S \cap \beta_S = \emptyset$ and that $\gamma_S \subset \beta_S$. For typographical ease, we drop the dependence of $\alpha_S, \beta_S, \gamma_S$ on S and instead write α, β, γ, since it does not result in any ambiguity. See Fig. 10 for an illustration.

2) Let $A_{ij} = 1$ if $(i,j) \in \mathcal{E}$ is on a shortest path from node 1 to node j and there exists some $S \in \mathbb{S}$ for which $i, j \in \beta_S$; otherwise, let $A_{ij} = 0$. Define the set

$$A = \{i : A_{ij} = 1, \text{ for some } j \in \mathcal{N}\}.$$

   As we shall see, each node $i \in A$ has a corresponding auxiliary random variable U_i instead of X_i in the expression for the upper bound.

3) Let $\tau = \max\{d(i,j) - d(j) + 1 : (i,j) \in \mathcal{E}\}$. This is used to ensure that all subscripts (time indices) are positive in Theorem 4 stated below.

We define the following shorthand notation for representing groups of random variables for a set $B \subset \mathcal{N}$ with reference to $S \in \mathbb{S}$:

$$\begin{aligned} X_t(B) &= \{X_{i(d(j)-d(i,j)+t)} : i \in B,\ j \in \beta\} \\ U_t(B) &= \{U_{i(d(i)+t)} : i \in B\} \\ Y_t(B) &= \{Y_{j(d(j)+t)} : j \in B\}. \end{aligned}$$

We also use the following natural extensions to handle multiple time indices:

$$\begin{aligned} X^t(B) &= \cup_{s=1}^{t} X_s(B) \\ U^t(B) &= \cup_{s=1}^{t} U_s(B) \\ Y^t(B) &= \cup_{s=1}^{t} Y_s(B). \end{aligned}$$

The following theorem provides a "single-letter" upper bound on the capacity of the relay network with delays.

Theorem 4: The capacity of a discrete memoryless relay network with delays is upper-bounded by

$$C \le \sup \min_{S \in \mathbb{S}} \left\{ I(X_\tau(\alpha \cap A^c), U_\tau(\alpha \cap A); Y_\tau(\beta) | X_\tau(\gamma \cap A^c), U_\tau(\gamma \cap A) \right\}$$

where the supremum is over all joint distributions of the random variables constituting $X_\tau(A^c), U_\tau(A)$, and over all functions such that $X_{it} = f_i(U_{it}, Y_{it})$ for all $i \in A$. The cardinality of any U_{it} is upper-bounded by $\prod_{i=1}^{K-1} |\mathcal{X}_i| + 2^{K-2} - 1$.

Remarks:

1) As we show in the examples that follow, the above bound reduces to the classical cut-set bound and feedforward relay network. Further, it reduces to the upper bound in (14) for the relay-without-delay.
2) The cut-set bound in [5] for classical relay networks has terms of the form $I(X(S); Y(S^c)|X(S^c))$. Note that our bound uses the structure of the network to pare off unnecessary random variables by replacing S with α and S^c with β and γ.
3) Our bound differs from the classical cut-set bound in two ways. First, multiple X random variables with different time indices can arise for the same node. These are the nodes in A^c. Second, auxiliary random variables, U are assigned to some nodes instead of X. These are the nodes in A. It is easy to see from the definition of $U_t(A)$ that each node in A has only one U variable.

To illustrate the application of Theorem 4, we consider some examples before proceeding to the proof.

Example 9 (Classical Relay Channel): In this case $\tau = 1$, $\mathbb{S} = \{\{1\}, \{1,2\}\}$, and $A_{12} = A_{13} = A_{23} = 0$, and $A = \emptyset$.

For $S = \{1\}$, we obtain $\alpha = \{1\}$, $\beta = \{2,3\}$, $\gamma = \{2\}$. Hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{11}\}, \qquad X_\tau(\gamma \cap A^c) = \{X_{21}\}, \\ U_\tau(\alpha \cap A) &= U_\tau(\gamma \cap A) = \emptyset, \quad Y_\tau(\beta) = \{Y_{22}, Y_{31}\}. \end{aligned}$$

Similarly, for $S = \{1, 2\}$, we obtain $\alpha = \{1, 2\}$, $\beta = \{3\}$, $\gamma = \emptyset$, and hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{11}, X_{21}\}, \quad X_\tau(\gamma \cap A^c) = \emptyset, \\ U_\tau(\alpha \cap A) &= U_\tau(\gamma \cap A) = \emptyset, \qquad Y_\tau(\beta) = \{Y_{31}\}. \end{aligned}$$

As a result, the upper bound on capacity is given by

$$\sup_{p(x_{11}, x_{21})} \min\{I(X_{11}; Y_{22}, Y_{31} | X_{21}), I(X_{11}, X_{21}; Y_{31})\}.$$

This is exactly the same as the classical cut-set upper bound (1), although the notation is different since time indices are also included. Note that $A = \emptyset$ and hence there are no auxiliary random variables, i.e., $U(\alpha \cap A) = U(\gamma \cap A) = \emptyset$.

Example 10 (Classical Feedforward Relay Network): The classical feedforward relay network consists of relay nodes that have a delay of one unit at each relay encoder. In our notation this means that the graph has nodes $1, \ldots, K$ with weights $d(1, i) = 0$ for $i = 2, \ldots, K$ and $d(j, k) = 1$ for $j = 2, \ldots, K-1$ and $k = j+1, \ldots, K$.

In this case, $\tau = 2$ and $A = \emptyset$, since the shortest path to any node from node 1 is the edge from 1 with weight 0. Consider the $K-1$ cuts that partition $\{1, \ldots, K\}$ into $S_i = \{1, \ldots, i\}$ and $S_i^c = \{i+1, \ldots, K\}$ for $i = 1, \ldots, K-1$. Given S_i, $\alpha = \{1, \ldots, i\}$, $\beta = \{i+1, \ldots, K\}$, and $\gamma = \{i+1, \ldots, K-1\}$. Hence, it follows that the capacity is upper-bounded by

$$\sup_{p(x_{12}, x_{21} \ldots, x_{(K-1)1})} \min_{i=1,\ldots,K-1} \{I(X_{12}, X_{21}, \ldots, X_{i1}; \\ Y_{(i+1)2}, \ldots, Y_{K2} | X_{(i+1)1}, \ldots, X_{(K-1)1})\}.$$

Since each variable appears only with one time index, we can rewrite this upper bound as

$$\sup_{p(x_1, x_2 \ldots, x_{K-1})} \min_{i=1,\ldots,K-1} \{I(X_1, X_2, \ldots, X_i; \\ Y_{i+1}, \ldots, Y_K | X_{i+1}, \ldots, X_{K-1})\}.$$

Note that this is the same as the classical cut-set bound [5]. In this example, we have used only $K-1$ sets instead of all sets in $\mathbb{S}$ as in Theorem 4.

Example 11 (Relay-Without-Delay): In this case, $\mathbb{S}$ and τ are the same as for the classical relay channel discussed above and $A_{13} = A_{12} = 0$, $A_{23} = 1$, and $A = \{2\}$.

For $S = \{1\}$, we obtain $\alpha = \{1\}$, $\beta = \{2, 3\}$, $\gamma = \{2\}$. Hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{11}\}, \quad X_\tau(\gamma \cap A^c) = U_\tau(\alpha \cap A) = \emptyset, \\ U_\tau(\gamma \cap A) &= \{U_{21}\}, \quad Y(\beta) = \{Y_{21}, Y_{31}\}. \end{aligned}$$

Similarly, for $S = \{1, 2\}$, we obtain $\alpha = \{1, 2\}$, $\beta = \{3\}$, $\gamma = \emptyset$, and hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{11}\}, \; X_\tau(\gamma \cap A^c) = \emptyset, \\ U_\tau(\alpha \cap A) &= \{U_{21}\}, \quad U_\tau(\gamma \cap A) = \emptyset, \quad Y_\tau(\beta) = \{Y_{31}\}. \end{aligned}$$

As a result, the upper bound on capacity is given by

$$\sup_{p(x_{11}, u_{21}), f} \min\{I(X_{11}; Y_{21}, Y_{31} | U_{21}), I(X_{11}, U_{21}; Y_{31})\}$$

where $x_{21} = f(y_{21}, u_{21})$.

This is exactly the same as the upper bound (14) for the relay-without-delay channel, but with different notation.

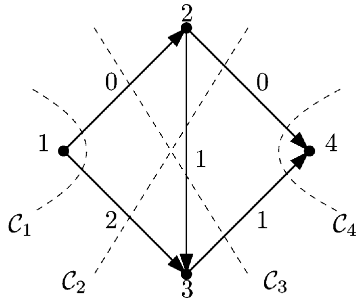

Fig. 11. Two-relay network with four cuts.

Note that in the relay-without-delay, node 2 is on the shortest path from 1 to 3 but this is not the case in the classical relay. In general, U_i replaces X_i when node i is on the shortest path to some node $j \in \beta$ and there exists some $S \in \mathbb{S}$ such that i, $j \in \beta$.

Example 12 (Two-Relay Network): Consider the two-relay network in Fig. 11.

In this case, $\mathbb{S} = \{\{1\}, \{1, 2\}, \{1, 3\}, \{1, 2, 3\}\}$, $\tau = 2$, $A_{12} = A_{13} = A_{34} = 0$, $A_{23} = A_{24} = 1$, and $A = \{2\}$. The four elements of $\mathbb{S}$ correspond to the four cuts $\mathcal{C}_1, \ldots, \mathcal{C}_4$ shown in Fig. 11.

For $S = \{1\}$, we have $\alpha = \{1\}$, $\beta = \{2, 3\}$, $\gamma = \{2\}$. Hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{11}, X_{12}\}, \quad U_\tau(\alpha \cap A) = \emptyset, \\ X_\tau(\gamma \cap A^c) &= \emptyset, \quad U_\tau(\gamma \cap A) = \{U_{22}\}, \\ Y_\tau(\beta) &= \{Y_{22}, Y_{33}\}. \end{aligned}$$

For $S = \{1, 2\}$, we have $\alpha = \{1, 2\}$, $\beta = \{3, 4\}$, $\gamma = \{3\}$. Hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{11}\}, \quad U_\tau(\alpha \cap A) = \{U_{22}\}, \\ X_\tau(\gamma \cap A^c) &= \{X_{31}\}, \quad U_\tau(\gamma \cap A) = \emptyset, \\ Y_\tau(\beta) &= \{Y_{33}, Y_{42}\}. \end{aligned}$$

For $S = \{1, 3\}$, we have $\alpha = \{1, 3\}$, $\beta = \{2, 4\}$, $\gamma = \{2\}$. Hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{12}, X_{31}\}, \quad U_\tau(\alpha \cap A) = \emptyset, \\ X_\tau(\gamma \cap A^c) &= \emptyset, \quad U_\tau(\gamma \cap A) = \{U_{22}\}, \\ Y_\tau(\beta) &= \{Y_{22}, Y_{42}\}. \end{aligned}$$

For $S = \{1, 2, 3\}$, we have $\alpha = \{2, 3\}$, $\beta = \{4\}$, $\gamma = \emptyset$. Hence

$$\begin{aligned} X_\tau(\alpha \cap A^c) &= \{X_{31}\}, \quad U_\tau(\alpha \cap A) = \{U_{22}\}, \\ X_\tau(\gamma \cap A^c) &= \emptyset, \quad U_\tau(\gamma \cap A) = \emptyset, \quad Y_\tau(\beta) = \{Y_{42}\}. \end{aligned}$$

As a result, the upper bound on capacity is given by

$$\sup_{p(x_{11}, x_{12}, x_{31}, u_{22}), f} \min\{I_1, I_2, I_3, I_4\}$$

where $x_{22} = f(y_{22}, u_{22})$ and

$$\begin{aligned} I_1 &= I(X_{11}, X_{12}; Y_{22}, Y_{33} | U_{22}) \\ I_2 &= I(X_{11}, U_{22}; Y_{33}, Y_{42} | X_{31}) \\ I_3 &= I(X_{12}, X_{31}; Y_{22}, Y_{42} | U_{22}) \end{aligned}$$

$$I_4 = I(X_{31}, U_{22}; Y_{42}).$$

In this example, note the following differences from the classical cut-set bound.

1) Variables X_{11}, X_{12} corresponding to node 1 with different time indices appear in the upper bound. Consider cut $\mathcal{C}_1$. Note that $Y_{3(t+1)}$ depends on $X_{1(t-1)}$, while Y_{2t} depends on X_{1t}. Since there is an edge from node 2 to node 3, $Y_{3(t+1)}$ also depends indirectly on X_{1t}. This creates the need to include two different random variables (corresponding to two time indices) for X_1 in the inequality corresponding to this cut.
2) Auxiliary random variables U replace X for nodes that lie on shortest paths. Node 2 is on the shortest path from 1 to 3 and 4 and hence the upper bound has U_2 instead of X_2.

The proof of Theorem 4 requires the following lemmas.

Lemma 1: For any $S \in \mathbb{S}$ corresponding to a relay network with delays

$$W \to \left\{Y_{i(D(i)+1)}^{n-l(i)} : i \in \beta\right\} \to Y_{K(D(K)+1)}^{n}.$$

The proof of this lemma is provided in the Appendix.

Lemma 2: For any $S \in \mathbb{S}$ corresponding to a relay network with delays, $X_t(\gamma \cap A^c)$ is a function of $Y^{t-1}(\beta)$.

Proof: Recall that

$$X_t(\gamma \cap A^c) = \{X_{i(d(j)-d(i,j)+t)} : i \in \gamma \cap A^c, j \in \beta\}.$$

Therefore, if $X_{i(d(j)-d(i,j)+t)} \in X_t(\gamma \cap A^c)$ then $A_{ij} \neq 1$, which means that node i does not lie on the shortest path from 1 to j. As a result, $d(j) - d(i,j) \leq d(i) - 1$, which implies that the time index of X_i is no more than $d(i) + t - 1$ and, hence, is a function of $Y_i^{d(i)+t-1}$, which is contained in $Y^{t-1}(\beta)$. □

Lemma 3: For any $S \in \mathbb{S}$ corresponding to a relay network with delays

$$\begin{aligned} &W, Y^{t-1}(\beta - (\gamma \cap A)) \to X_t(\gamma \cap A^c), \\ &\quad Y^{t-1}(\gamma \cap A), \quad X_t(\alpha \cap A^c), \quad Y^{t-1}(\alpha \cap A) \to Y_t(\beta). \end{aligned}$$

The main idea in the proof of this lemma is similar to that in the proof of Lemma 1, but the proof is significantly more cumbersome, and hence is not provided.

We are now ready to prove Theorem 4.

Proof of Theorem 4: Consider some $S \in \mathbb{S}$

$$\begin{aligned} nR &\overset{(a)}{\leq} I\left(W; Y_{K(D(K)+1)}^{d(K)+n}\right) + n\epsilon_n \\ &\overset{(b)}{\leq} I\left(W; Y_{K(D(K)+1)}^{d(K)+n}, \left\{Y_{i(D(i)+1)}^{d(K)+n-l(i)} : i \in \beta\right\}\right) + n\epsilon_n \\ &\overset{(c)}{=} I\left(W; \left\{Y_{i(D(i)+1)}^{d(K)+n-l(i)} : i \in \beta\right\}\right) + n\epsilon_n \\ &\overset{(d)}{\leq} I\left(W; Y^n(\beta)\right) + n\epsilon_n \end{aligned} \tag{41}$$

where (a) holds due to Fano's inequality, (b) is due to the nonnegativity of conditional mutual information, and (c) follows from Lemma 1. Inequality (d) holds due to the nonnegativity of conditional mutual information and the fact that

$$\left\{Y_{i(D(i)+1)}^{d(K)+n-l(i)} : j \in \beta\right\} \subset Y^n(\beta)$$

which is easy to verify.

Next we obtain

$$\begin{aligned} I\left(W; Y^n(\beta)\right) &\overset{(e)}{=} \sum_{t=1}^{n} H\left(Y_t(\beta)|Y^{t-1}(\beta)\right) \\ &\quad - H\left(Y_t(\beta)|Y^{t-1}(\beta), W\right) \\ &\overset{(f)}{=} \sum_{t=1}^{n} H\left(Y_t(\beta)|Y^{t-1}(\beta), X_t(\gamma \cap A^c)\right) \\ &\quad - H\left(Y_t(\beta)|Y^{t-1}(\beta), W\right) \\ &\overset{(g)}{\leq} \sum_{t=1}^{n} H\left(Y_t(\beta)|Y^{t-1}(\beta), X_t(\gamma \cap A^c)\right) \\ &\quad - H\Big(Y_t(\beta)|Y^{t-1}(\beta), W, X_t(\gamma \cap A^c), \\ &\qquad X_t(\alpha \cap A^c), Y^{t-1}(\alpha \cap A)\Big) \\ &\overset{(h)}{=} \sum_{t=1}^{n} H\left(Y_t(\beta)|Y^{t-1}(\beta), X_t(\gamma \cap A^c)\right) \\ &\quad - H\Big(Y_t(\beta)|Y^{t-1}(\gamma \cap A), X_t(\alpha \cap A^c), \\ &\qquad X_t(\gamma \cap A^c), Y^{t-1}(\alpha \cap A)\Big) \\ &\overset{(i)}{\leq} \sum_{t=1}^{n} H\left(Y_t(\beta)|Y^{t-1}(\gamma \cap A), X_t(\gamma \cap A^c)\right) \\ &\quad - H\Big(Y_t(\beta)|Y^{t-1}(\gamma \cap A), X_t(\alpha \cap A^c), \\ &\qquad X_t(\gamma \cap A^c), Y^{t-1}(\alpha \cap A)\Big) \\ &= \sum_{t=1}^{n} I\Big(X_t(\alpha \cap A^c), Y^{t-1}(\alpha \cap A); \\ &\qquad Y_t(\beta)|Y^{t-1}(\gamma \cap A), X_t(\gamma \cap A^c)\Big) \\ &\overset{(j)}{=} \sum_{t=1}^{n} I\Big(X_t(\alpha \cap A^c), U_t(\alpha \cap A); Y_t(\beta)| \\ &\qquad U_t(\gamma \cap A), X_t(\gamma \cap A^c)\Big). \end{aligned} \tag{42}$$

Equality (e) follows by the chain rule and (f) holds by Lemma 2. Inequalities (g) and (i) hold since conditioning reduces entropy. Lemma 3 results in (h), and (j) follows due to renaming.

The rest of the proof, without the cardinality bound, follows by using a standard time-sharing argument. It is also easy to verify that the choice to τ leads to positive time indices for all variables in such a way that at least one of the variables has time index 1. The cardinality bound on U_{it} follows by using the same argument as in the proof of Lemma 3.5 in [4] and by noting that $|\mathbb{S}| = 2^{K-2}$. □

VII. Conclusion

The paper investigated the effect of link delays on the capacity of relay networks. Although achieving capacity requires arbitrarily long coding delay, we showed that finite link delays can change the nature of cooperation in a network, and thus change its capacity.

We first studied the relay-with-delay channel. We presented upper and lower bounds on the capacity of the relay-with-unlimited-look-ahead and the relay-without-delay. The bounds were shown to be tight for the following.

1) Sato's example for any $d \leq 0$.
2) Degraded and semi-deterministic relay-with-unlimited-look-ahead.
3) Degraded and semi-deterministic relay-without-delay, where $d = 0$.
4) AWGN relay-with-delay for any $d \leq 0$ when $a^2 \leq b^2 \min\{1, P/(a^2P + N)\}$.
5) AWGN relay-with-unlimited-delay when $a > b + 1$.

In addition it is shown that the classical cut-set bound does not hold for relay-with-delay when $d \leq 0$.

The lower bounds are achieved using different cooperation strategies that depend on delay. The capacity of the Sato relay-without-delay is achieved using instantaneous relaying. Capacity for the degraded and semi-deterministic relay-with-unlimited-delay is achieved by noncausal decode-and-forward. Capacity of the AWGN relay-without-delay when $a^2 \leq b^2 \min\{1, P/(a^2P + N)\}$ is achieved using amplify-and-forward. Furthermore, we showed that achieving capacity may require a mixture of different cooperation strategies. For example, to achieve the capacity of the degraded or semi-deterministic relay-without-delay, a superposition of instantaneous relaying and decode-and-forward is needed.

We then defined the relay network with delays, which includes the classical relay networks and the relay-with-delay as special cases. We showed that only relative paths delays matter to the capacity of a network. We then provided a new cut-set upper bound that generalizes the classical cut-set bound and the upper bound for the relay-without-delay in Theorem 2.

Many open questions remain.

1) We did not evaluate the upper bound on the relay-without-delay (14) for the AWGN case. We know that in general it is not maximized by Gaussian (X, V) and linear relaying functions of the form $X_1 = \alpha Y_1 + \beta V$. It would be interesting to find the optimal choice of the distribution on (X, V) and function $f(V, Y_1)$ for this case.
2) Fig. 7 provides an example of the potential rate increase achievable by exploiting delay. It would be interesting to investigate how fast in terms of increase in negative delay can this limit be achieved.
3) Our lower bounds are developed for two extreme cases; the relay-without-delay (where $d = 0$) and the relay-with-unlimited-look-ahead, where the relay knows its entire received sequence Y_{11}^n noncausally. Theorem 4 provides a general upper bound on capacity as a function of link delays. It would be interesting to develop lower bounds that are functions of d. We know how to benefit from past and present symbols, and unlimited look-ahead. How do we use finite look-ahead (where $d < 0$)? Can the upper bound in Theorem 1 be achieved for $d < 0$?
4) It would be interesting to investigate the case of delay $d \geq 2$. For example, is $C_2 < C_1$ in general?
5) The following questions arise regarding network capacity.
   (a) Theorem 3 provides a sufficient condition for two networks to have the same capacity. What are the necessary conditions for capacity to be equal?
   (b) For a given network, how does change in delay on one of the links affect capacity? In some special cases such as the relay-with-delay channel, Theorem 3 can be used to answer this question.
   (c) The proofs suggest that the results derived for relay networks with delay in this paper would hold even if the graph has cycles, provided that the length of each cycle is strictly positive. This would be an interesting generalization.
   (d) The classical cut-set bound for networks in [5] is generalized to multiuser networks in [3]. It would be interesting to generalize the new cut-set bounds to multiuser networks.

Appendix
Markov Properties

The probabilistic structure of the relay network allows the joint distribution of the X, Y variables corresponding to various nodes to be factorized by nodes as in (28) and the memorylessness property allows further factorization by time slots as in (29). In order to state some general forms of these factorizations, we need the following definitions. These are in addition to the definitions of $\mathcal{N}_i$, $x_t(\mathcal{N}_i)$, $d(i)$, $D(i)$, $\mathbb{S}$, α_S, and β_S defined in Section VI.

1) Let $l(i)$ be the minimum of the delays on paths from node i to node K and define $l(K) = 0$.
2) Any node that lies on a path from node 1 to node j is said to be downstream of j. Let $F(j)$ denote the set of nodes that are strictly downstream of node j. That is, $F(j)$ is the set of nodes that lie on some path from node 1 to node j and does not include j. For $V \subset \mathcal{N}$, define $F(V) = \cup_{j \in V} F(j)$.
3) Given A, B that are disjoint subsets of $\mathcal{N}$, a direct path from A to B is any path from a node in A to a node in B that does not pass through another node in A. For $S \in \mathbb{S}$, we define $U(S)$ to be the set of nodes that lie on direct paths from S to $\{K\}$, not including the nodes in S. That is, $U(S)$ is the set of nodes that are strictly upstream of S.

From the above definitions, it is easy to see that

$$\mathcal{N} = U(\beta) \cup F(\beta) \cup [\beta - F(\beta)] = U(\beta) \cup \beta \cup F(\beta) \quad (43)$$

and that $U(\beta) \cap (\beta \cup F(\beta)) = \emptyset$. However, $\beta \cap F(\beta)$ need not be empty.

The graphical structure of the relay network with delays provides the conditional pmf of Y_{it} for any $i \in \mathcal{N}$ given $X_t(\mathcal{N}_i)$. Further, x_{it} is a function of $y_{i(D(i)+1)}^t$. For typographical ease, more generally, we write $p\left(x_{it} | y_{i1}^t\right)$ instead of $p(x_{it} | y_{i(D(i)+1)}^t)$ in what follows. Starting at node K and proceeding recursively, the joint distribution of all variables corresponding to nodes that are downstream of K with the appropriate time indices can be factorized. Observe that only time indices ranging from $D(i)+1$ to $t-l(i)$ of x_i and y_i can affect y_{Kt}. From this, it is easy to see that the joint distribution of $y^n_{K(d(K)+1)}$ and $\{x_{it}, y_{it} : i \in \mathcal{N} - \{K\}, D(i)+1 \le t \le n-l(i)\}$ is given by

$$\begin{aligned} p\Bigg(&y^n_{K(D(K)+1)}, \{x_{it}, y_{it} : i \in \mathcal{N} - \{1,K\}, \\ &\qquad D(i)+1 \le t \le n-l(i)\}, x_{11}^{n-l(1)}\Bigg) \\ &= \prod_{t=D(K)+1}^{n} p(y_{Kt}|x_t(\mathcal{N}_K)) \\ &\quad \cdot \prod_{i\in\mathcal{N}-\{1,K\}} \prod_{s=D(i)+1}^{n-l(i)} p(y_{is}|x_s(\mathcal{N}_i)) p\left(x_{is}|y^s_{i1}\right) \\ &\quad \cdot p\left(x_{11}^{n-l(1)}\right). \end{aligned} \tag{44}$$

Using the convention that $p\left(x_{Kt}|y^t_{K1}\right) \equiv 1$, and $Y_{1t} = X_{1t}$ (or, equivalently, that $p\left(y_{1t}|x_t(\mathcal{N}_1)\right) p\left(x_{1t}|y^t_{11}\right) = p(x_{1t})$), we can rewrite (44) as

$$\begin{aligned} p&\left(\{x_{it}, y_{it} : i \in \mathcal{N}, D(i)+1 \le t \le n-l(i)\}\right) \\ &= \prod_{i\in\mathcal{N}} \prod_{t=D(i)+1}^{n-l(i)} p(y_{it}|x_t(\mathcal{N}_i)) p\left(x_{it}|y^t_{i1}\right). \end{aligned} \tag{45}$$

The relay encoding function at node i at time t determines x_{it} based on y^t_i deterministically. Hence, the pmf $p\left(x_{it}|y^t_{i1}\right)$ is 1 if x_{it} is the value of the relay encoding function at y^t_{i1} and 0 otherwise. Using this and the directed acyclic graphical structure it is easy to see that we can eliminate the xs determined by the relay encoding functions from (45) to obtain

$$\begin{aligned} p&\left(\{y_{it} : i \in \mathcal{N}, D(i)+1 \le t \le n-l(i)\}\right) \\ &= \prod_{i\in\mathcal{N}} \prod_{t=D(i)+1}^{n-l(i)} p(y_{it}|\{y_{js} : 1 \le s \le t-d(k,i)\}). \end{aligned} \tag{46}$$

Note that our convention, $Y_{1t} = X_{1t}$, allows us to express the above factorization in terms of only the y variables. Akin to $x_t(\mathcal{N}_i)$, we define $y^t(\mathcal{N}_i) = \{y_{js} : j \in \mathcal{N}_i, 1 \le s \le t - d(j,i)\}$. This allows us to rewrite (46) as

$$\begin{aligned} p&\left(\{y_{it} : i \in \mathcal{N}, D(i)+1 \le t \le n-l(i)\}\right) \\ &= \prod_{i\in\mathcal{N}} p\left(y^{n-l(i)}_{i(D(i)+1)}|y^{n-l(i)}(\mathcal{N}_i)\right). \end{aligned} \tag{47}$$

Above, we factorized the joint distribution of variables corresponding to node K and all its downstream nodes. In the same manner, we can factorize the joint distribution of variables corresponding to a set of nodes $V \subset \mathcal{N}$ (instead of a single node K) and its downstream nodes $F(V)$. In particular, for $\beta (= \beta_S)$ corresponding to $S \in \mathbb{S}$, as defined earlier

$$\begin{aligned} p&\left(\left\{y^{n-l(i)}_{i(D(i)+1)} : i \in \beta \cup F(\beta)\right\}\right) \\ &= \prod_{i\in\beta\cup F(\beta)} p\left(y^{n-l(i)}_{i(D(i)+1)}|y^{n-l(i)}(\mathcal{N}_i)\right). \end{aligned} \tag{48}$$

Proof: of Lemma 1: Let $p(\mathcal{N})$ denote the joint pmf of $\left\{Y^{n-l(i)}_{i(D(i)+1)} : i \in \mathcal{N}\right\}$, then from (43) and (47), we obtain

$$p(\mathcal{N}) = \prod_{i\in U(\beta)} p(\cdot|\cdot) \prod_{i\in\beta\cup F(\beta)} p(\cdot|\cdot), \tag{49}$$

where $p(\cdot|\cdot) = p\left(y^{n-l(i)}_{i(D(i)+1)}|y^{n-l(i)}(\mathcal{N}_i)\right)$. Recall that we are using the notation $Y_{1t} = X_{1t}$ in (49).

Denote the marginal distribution of the Y random variables corresponding to the nodes in $\{1\} \cup \beta \cup \{K\}$ by $p(1,\beta,K)$. Using the joint distribution in (49), this marginal distribution can be written as

$$p(1,\beta,K) = \sum_{y\in\mathcal{N}-\{1,K\}-\beta} \prod_{i\in U(\beta)} p(\cdot|\cdot) \prod_{i\in\beta\cup F(\beta)} p(\cdot|\cdot) \tag{50}$$

where by $y \in B \subset \mathcal{N}$ we mean $\{y_{it} : i \in B, D(i)+1 \le t \le n-l(i)\}$. For $i \in U(\beta)$, $\mathcal{N}_i$ only contains nodes in $U(\beta)\cup\beta$, and similarly, for $i \in \beta\cup F(\beta)$, $\mathcal{N}_i$ only contains nodes in $\beta\cup F(\beta)$. As a result, (50) can be rewritten as

$$\begin{aligned} &p(1,\beta,K) \\ &= \sum_{y\in U(\beta)-\{K\}} \prod_{i\in U(\beta)} p(\cdot|\cdot) \sum_{y\in F(\beta)-\beta-\{1\}} \prod_{i\in\beta\cup F(\beta)} p(\cdot|\cdot). \end{aligned} \tag{51}$$

Denote the marginal pmf of the Y's corresponding to the nodes in $\beta \cup \{1\}$ by $p(1,\beta)$. From (48)

$$p(1,\beta) = \sum_{y\in F(\beta)-\beta-\{1\}} \prod_{i\in\beta\cup F(\beta)} p(\cdot|\cdot). \tag{52}$$

Using (52) above, we can rewrite (51) as

$$p(1,\beta,K) = \left(\sum_{y\in U(\beta)-\{K\}} \prod_{i\in U(\beta)} p(\cdot|\cdot)\right) p(1,\beta). \tag{53}$$

Let $p(\beta)$ be the marginal pmf of the Y's corresponding to nodes in β. From (48)

$$p(\beta) = \sum_{y\in F(\beta)-\beta} \prod_{i\in\beta\cup F(\beta)} p(\cdot|\cdot). \tag{54}$$

Now consider the joint pmf of the Y variables corresponding to nodes in $\{K\} \cup \beta$, denoted by $p(\beta,K)$. It can be written as

$$\begin{aligned} p(\beta,K) &= \sum_{y\in\mathcal{N}-\{1,K\}-\beta} \prod_{i\in U(\beta)} p(\cdot|\cdot) \prod_{i\in\beta\cup F(\beta)} p(\cdot|\cdot) \\ &= \sum_{y\in U(\beta)-\{K\}} \prod_{i\in U(\beta)} p(\cdot|\cdot) \sum_{y\in F(\beta)-\beta} \prod_{i\in\beta\cup F(\beta)} p(\cdot|\cdot) \qquad (55) \\ &= \left(\sum_{y\in U(\beta)-\{K\}} \prod_{i\in U(\beta)} p(\cdot|\cdot)\right) p(\beta) \qquad (56) \end{aligned}$$

where (55) holds for the same reason as (51) and (56) holds due to (54).

From (53) and (56), it is clear that

$$p(1,\beta,K)/p(1,\beta)=p(\beta,K)/p(\beta)$$

if $p(1,\beta)\neq 0$ and $p(\beta)\neq 0$. Once more recall that we defined $Y_{1t}=X_{1t}$ and hence this proves the desired Markov property stated in Lemma 1. □

ACKNOWLEDGMENT

The authors wish to thank Young-Han Kim and the reviewers for many comments that have significantly improved the content and presentation of this paper.

REFERENCES

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

[2] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

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

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

[5] A. El Gamal, "On information flow in relay networks," in *Proc. IEEE Nat. Telecommunications Conf.*, Nov. 1981, vol. 2, pp. D4.1.1–D4.1.4.

[6] A. El Gamal and M. Aref, "The capacity of the semi-deterministic relay channel," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 3, pp. 536–536, May 1982.

[7] A. El Gamal and N. Hassanpour, "Relay without delay," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 1078–1080.

[8] A. El Gamal and N. Hassanpour, "Capacity theorems for the relay without delay channel," in *Proc. Allerton Conf. Communication, Control and Computation*, Monticello, IL, Sep. 2005.

[9] A. El Gamal and J. Mammen, "Relay networks with delay," presented at the UCSD Workshop on Information Theory and Its Applications, San Diego, CA, Feb. 2006.

[10] A. El Gamal, M. Mohseni, and S. Zahedi, "On reliable communication over additive white Gaussian noise relay channels," *IEEE Trans. Inf. Theory*, submitted for publication.

[11] N. Hassanpour, "Relay Without Delay," Ph.D. dissertation, Dep. Elec. Eng., Stanford Univ., Stanford, CA, 2006.

[12] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channel," *IEEE Trans. Inf. Theory*, submitted for publication.

[13] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[14] H. Sato, "Information Transmission Through a Channel With Relay," The Aloha System, Univ. Hawaii, Honolulu, 1976, Tech. Rep. B76-7.

[15] B. Schein and R. G. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jul. 2000, p. 22.

[16] F. M. J. Willems, "The multiple-access channel with cribbing encoders revisited," presented at the Workshop on Mathematics of Relaying and Cooperation in Communication Networks, Berkeley, CA, Apr. 2006.

[17] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.