On the Capacity of the Noncausal Relay Channel

Lele Wang and Mohammad Naghshvar
Department of Electrical and Computer Engineering
University of California San Diego
La Jolla, CA 92093 USA
{lew001, naghshvar}@ucsd.edu

Abstract

This paper studies the noncausal relay channel, also known as the relay channel with unlimited lookahead, introduced by El Gamal, Hassanpour, and Mammen. Unlike the standard relay channel model, where the relay encodes its signal based on the previous received output symbols, the relay in the noncausal relay channel encodes its signal as a function of the entire received sequence. In the existing coding schemes, the relay uses this noncausal information solely to recover the transmitted message and then cooperates with the sender to communicate this message to the receiver. However, it is shown in this paper that by applying the Gelfand–Pinsker coding scheme, the relay can take further advantage of the noncausally available information, which can achieve strictly higher rates than existing coding schemes. This paper also provides a new upper bound on the capacity of the noncausal relay that strictly improves upon the cutset bound. These new lower and upper bounds on the capacity coincide for the class of degraded noncausal relay channels and establish the capacity for this class.

I. INTRODUCTION

The relay channel was first introduced by van der Meulen [2]. In their classic paper [8], Cover and El Gamal established the cutset upper bound and the decode–forward, partial decode–forward, and compress–forward lower bounds for the relay channel. Furthermore, they established the capacity of the degraded and reversely degraded relay channels and relay channels with feedback.

The relay channel with lookahead was introduced by El Gamal, Hassanpour, and Mammen [3], who mainly studied the following two classes:

- Noncausal relay channel (also known as the relay channel with unlimited lookahead) in which the relay knows its entire received sequence in advance and hence the relaying functions can depend on the whole received block. Lower bounds on the capacity were established by extending (partial) decode–forward coding scheme to the noncausal case. The cutset upper bound for the noncausal relay channel was also established.

- Causal relay channel (also known as the relay-without-delay channel) in which the relay has access only to the past and present received sequence. A lower bound for the capacity of this channel was established by combining partial decode–forward and instantaneous relaying coding schemes. The cutset upper bound for the causal relay channel was also established.

The focus of this paper is on the noncausal relay channel. The existing lower bounds on the capacity of this channel are derived using the (partial) decode–forward coding scheme. In particular, the relay recovers (the part of) the transmitted message from the received sequence (available noncausally at the relay) and then cooperates with the sender to coherently
transmit this message to the receiver. Therefore, the noncausally available information is used solely to recover (the part of) the transmitted message at the relay. However, the channel conditional pmf can allow the relay to take further advantage of the received sequence by considering it as noncausal side information to help the relay’s communication to the receiver.

In this paper, we establish several improved lower bounds on the capacity of the noncausal relay channel based on this observation by combining the Gelfand–Pinsker coding scheme with (partial) decode–forward and compress–forward at the relay. Moreover, we establish a new upper bound on the capacity that improves upon the cutset bound \[1, \text{Theorem 17.6}\]. The new upper bound is shown to be optimal for the class of degraded noncausal relay channels and is achieved by the Gelfand–Pinsker decode–forward coding scheme.

The remainder of this paper is organized as follows. In Section \[\text{II}\] we formulate the problem and provide a brief overview of the existing results. In Section \[\text{III}\] we establish three improved lower bounds, the Gelfand–Pinsker decode–forward (GP-DF) lower bound, the Gelfand–Pinsker compress–forward lower bound, and the Gelfand–Pinsker partial decode–forward lower bound. We show through Example \[1\] that the GP-DF lower bound can be strictly tighter than the existing lower bound. In Section \[\text{IV}\] we establish an improved upper bound on the capacity, which is shown to strictly improve upon the cutset bound in Example \[5\]. The improved upper bound together with the GP-DF lower bound establish the capacity of the degraded noncausal relay channels.

Throughout the paper, we follow the notation in \[1\]. In particular, a random variable is denoted by an upper case letter (e.g., \(X, Y, Z\)) and its realization is denoted by a lower case letter (e.g., \(x, y, z\)). By convention, \(X = \emptyset\) means that \(X\) is a degenerate random variable (unspecified constant) regardless of its support. Let \(X^n_k = (X_{k1}, X_{k2}, \ldots, X_{kn})\). We say that \(X \rightarrow Y \rightarrow Z\) form a Markov chain if \(p(x, y, z) = p(x)p(y|x)p(z|y)\). For \(a \geq 0\), \([1 : 2^a] = \{1, 2, \ldots, 2^a\}\), where \([a]\) is the smallest integer greater than or equal to \(a\). For any set \(S\), \(|S|\) denotes its cardinality. The probability of an event \(A\) is denoted by \(P(A)\).

\section{Problem Formulation and Known Results}

\subsection{Noncausal Relay Channels}

Consider the 3-node point-to-point communication system with a relay depicted in Figure \[1\] The sender (node 1) wishes to communicate a message \(M\) to the receiver (node 3) with the help of the relay (node 2). The \textit{discrete memoryless (DM) relay channel with lookahead} is described as

\[(X_1, X_2, p(y_2|x_1)p(y_3|x_1, x_2, y_2), Y_2, Y_3, l)\]  \hspace{1cm} (1)

where the parameter \(l \in \mathbb{Z}\) specifies the amount of lookahead. The channel is memoryless in the sense that \(p(y_{2i}|x_1, y_{2i-1}, m) = p_{Y_2|X_1}(y_{2i}|x_{1i})\) and \(p(y_{3i}|x_{1i}, x_{2i}, y_{3i-1}, m) = p_{Y_3|X_1, X_2, Y_2}(y_{3i}|x_{1i}, x_{2i}, y_{2i})\).

A \((2^nR, n)\) code for the relay channel with lookahead consists of

- a message set \([1 : 2^{nR}]\),
- an encoder that assigns a codeword \(x^n_m(m)\) to each message \(m \in [1 : 2^{nR}]\),
- a relay encoder that assigns a symbol \(x_{2i}(y_{2i-1}^{i-1})\) to each sequence \(y_{2i-1}^{i+t}\) for \(i \in [1 : n]\), where the symbols that have nonpositive time indices or time indices greater than \(n\) are arbitrary, and
- a decoder that assigns an estimate \(\hat{M}(y^n_R)\) or an error message \(e\) to each received sequence \(y^n_R\).

We assume that the message \(M\) is uniformly distributed over \([1 : 2^{nR}]\). The average probability of error is defined as \(P^e = P(M \neq \hat{M})\). A rate \(R\) is said to be \textit{achievable} for the DM relay channel with lookahead if there exists a sequence.
of \((2^nR, n)\) codes such that \(\lim_{n \to \infty} \frac{n}{n} = 0\). The capacity \(C_l\) of the DM relay channel with lookahead is the supremum of all achievable rates.

The standard DM relay channel corresponds to lookahead parameter \(l = -1\), or equivalently, a delay of 1. The noncausal relay channel which we focus on in this paper is the case where \(l\) is unbounded, i.e., the relaying functions can depend on the entire received sequence \(y_2^n\). The purpose of studying this extreme case is to quantify the limit on the potential gain from relaying.

**B. Prior Work on the Noncausal Relay Channel**

The noncausal relay channel was initially studied by El Gamal, Hassanpour, and Mammen, who established the following lower bounds and cutset upper bound on the capacity \(C_\infty\).

- **Decode–forward (DF) lower bound:**
  \[
  C_\infty \geq R_{DF} = \max_{p(x_1, x_2)} \min \{I(X_1; Y_2), I(X_1, X_2; Y_3)\},
  \]

- **Partial decode–forward (PDF) lower bound:**
  \[
  C_\infty \geq R_{PDF} = \max_{p(v, x_1, x_2)} \min \{I(X_1, X_2; Y_3), I(V; Y_2) + I(X_1, Y_3 | X_2, V)\},
  \]

- **Cutset bound for noncausal relay channel:**
  \[
  C_\infty \leq R_{CS} = \max_{p(x_1)p(x_2|y_2)} \min \{I(X_1, X_2; Y_3), I(X_1; Y_2) + I(X_1; Y_3 | X_2, Y_2)\}.
  \]

**III. LOWER BOUNDS**

In this section, we establish three lower bounds by considering the received \(y_2^n\) sequence at the relay as noncausal side information to help communication. In Subsection III-A we first establish the Gelfand–Pinsker decode–forward (GP-DF) lower bound by incorporating Gelfand–Pinsker coding with the decode–forward coding scheme. Then, we show the GP-DF lower bound can be strictly tighter than the decode–forward lower bound and achieve the capacity in Example 1.

In Subsection III-B we establish the Gelfand–Pinsker compress–forward (GP-CF) lower bound via two different coding schemes. In Subsection III-C we further combine the the coding scheme for the GP-CF lower bound with partial decode–forward coding scheme.

---

1. Note that here we define the relay channel with lookahead as \(p(y_2|x_1)p(y_3|x_1, x_2, y_2)\), since the conditional pmf \(p(y_2, y_3|x_1, x_2)\) depends on the code due to the instantaneous or lookahead dependency of \(X_2\) on \(Y_2\).

2. There is a small typo in [3, Theorem 1] where the maximum is over \(p(x_1, x_2)\) instead of \(p(x_1)p(x_2|x_1, y_2)\).
A. Gelfand–Pinsker Decode–Forward Lower Bound

We first incorporate Gelfand–Pinsker coding with the decode–forward coding scheme.

**Theorem 1** (Gelfand–Pinsker decode–forward (GP-DF) lower bound). The capacity of the noncausal relay channel is lower bounded as

\[
C_\infty \geq R_{\text{GP-DF}} = \max \min \{ I(X_1; Y_2), \ I(X_1, U; Y_3) - I(U; Y_2|X_1) \},
\]

where the maximum is over all pmfs \( p(x_1)p(u|x_1, y_2) \) and functions \( x_2(u, x_1, y_2) \).

**Proof:** The GP-DF coding scheme uses multicoing and joint typicality encoding and decoding. For each message \( m \), we generate a \( x_1^n(m) \) sequence and a subcodebook \( C(m) \) of \( 2^{nR} \) \( u^n(l|m) \) sequences. To send message \( m \), the sender transmits \( x_1^n(m) \). Upon receiving \( y_2^n \) noncausally, the relay first finds a message estimate \( \hat{m} \). It then finds a \( u^n(l|\hat{m}) \in C(\hat{m}) \) that is jointly typical with \( (x_1^n(\hat{m}), y_2^n) \) and transmits \( x_2^n(\hat{m}) \). The receiver declares \( \hat{m} \) to be the message estimate if \( (x_1^n(\hat{m}), u^n(l|\hat{m}), y_3^n) \) are jointly typical for some \( u^n(l|\hat{m}) \in C(\hat{m}) \). We now provide the details of the proof.

**Codebook generation:** Fix \( p(x_1)p(u|x_1, y_2) \) and \( x_2(u, x_1, y_2) \) that attain the lower bound. Randomly and independently generate \( 2^{nR} \) sequences \( x_1^n(m) \), each according to \( \prod_{i=1}^{n} p(x_1(x_{1i})) \), \( m \in [1 : 2^{nR}] \). For each message \( m \in [1 : 2^{nR}] \), randomly and conditionally independently generate \( 2^{nR} \) sequences \( u^n(l|m) \), each according to \( \prod_{i=1}^{n} p(u_l|x_{1i}(m)) \), which form the subcodebook \( C(m) \). This defines the codebook \( C = \{(x_1^n(m), u^n(l|m), x_2^n(\hat{m}), y_3^n): m \in [1 : 2^{nR}], l \in [1 : 2^{nR}]\} \). The codebook is revealed to all parties.

**Encoding:** To send message \( m \), the encoder transmits \( x_1^n(m) \).

**Relay encoding:** Upon receiving \( y_2^n \) noncausally, the relay first finds the unique message \( \hat{m} \) such that \( (x_1^n(\hat{m}), y_2^n) \in T_1^{(n)} \). Then, it finds a sequence \( u^n(l|\hat{m}) \in C(\hat{m}) \) such that \( (u^n(l|\hat{m}), x_2^n(\hat{m}), y_3^n) \in T_2^{(n)} \). The relay transmits \( x_2 = x_2(u_{\hat{m}}(l|m), \hat{m}), y_2 \) at time \( i \in [1 : n] \).

**Decoding:** Let \( \epsilon > \epsilon' \). Upon receiving \( y_2^n \), the decoder declares that \( \hat{m} \in [1 : 2^{nR}] \) is sent if it is the unique message such that \( (x_1^n(\hat{m}), u^n(l|m), y_3^n) \in T_3^{(n)} \) for some \( u^n(l|m) \in C(\hat{m}) \); otherwise, it declares an error.

**Analysis of the probability of error:** We analyze the probability of decoding error averaged over codes. Assume without loss of generality that \( M = 1 \). Let \( \hat{M} \) be the relay’s message estimate and let \( L \) denote the index of the chosen \( U^n \) codeword for \( \hat{M} \) and \( Y_2^n \). The decoder makes an error only if one of the following events occur:

\[
\begin{align*}
\hat{E} &= \{ \hat{M} \neq 1 \}, \\
\hat{E}_1 &= \{(X_1^n(1), Y_2^n) \notin T_1^{(n)} \}, \\
\hat{E}_2 &= \{(X_1^n(m), Y_2^n) \in T_1^{(n)} \text{ for some } m \neq 1 \}, \\
\hat{E}_3 &= \{(U^n(l|\hat{M}), X_1^n(\hat{M}), Y_2^n) \notin T_2^{(n)} \text{ for all } U^n(l|\hat{M}) \in C(\hat{M}) \}, \\
\hat{E}_4 &= \{(X_1^n(1), U^n(L|1), Y_3^n) \notin T_3^{(n)} \}, \\
\hat{E}_5 &= \{(X_1^n(m), U^n(l|m), Y_3^n) \in T_3^{(n)} \text{ for some } m \neq 1, U^n(l|m) \in C(m) \}.
\end{align*}
\]
Thus, the probability of error is upper bounded as

\[
P(\mathcal{E}) = P\{\hat{M} \neq 1\}
\leq P(\mathcal{E} \cup \mathcal{E}_3 \cup \mathcal{E}_1 \cup \mathcal{E}_2)
\leq P(\mathcal{E}) + P(\mathcal{E}_3 \cap \mathcal{E}^c) + P(\mathcal{E}_1 \cap \mathcal{E}_c \cap \mathcal{E}_3^c) + P(\mathcal{E}_2)
\leq P(\mathcal{E}_1) + P(\mathcal{E}_2) + P(\mathcal{E}_3 \cap \mathcal{E}^c) + P(\mathcal{E}_1 \cap \mathcal{E}^c \cap \mathcal{E}_3^c) + P(\mathcal{E}_2).
\]

By the law of large numbers (LLN), the first term tends to zero as \( n \to \infty \). By the packing lemma \( [1] \) p. 3.18, the second term tends to zero as \( n \to \infty \) if \( R < I(X_1;Y_2) - \delta(e') \). Therefore, \( P(\mathcal{E}) \) tends to zero as \( n \to \infty \) if \( R < I(X_1;Y_2) - \delta(e') \). Given \( \mathcal{E}^c \), i.e. \( \{\hat{M} = 1\} \), by the covering lemma \( [1] \) p. 3.51, the third term tends to zero as \( n \to \infty \) if \( \tilde{R} > I(U;Y_2|X_1) + \delta(e') \).

By the conditional typicality lemma, the fourth term tends to zero as \( n \to \infty \). Finally, note that once \( m \) is wrong, \( U^n(l|m) \) is also wrong. By the packing lemma, the last term tends to zero as \( n \to \infty \) if \( R + \tilde{R} < I(X_1,U;Y_3) - \delta(e) \). Combining the bounds and eliminating \( \tilde{R} \), we have shown that \( P\{\hat{M} \neq 1\} \) tends to zero as \( n \to \infty \) if \( R < I(X_1;Y_2) - \delta(e') \) and \( R < I(X_1,U;Y_3) - I(U;Y_2|X_1) - \delta'(e) \) where \( \delta'(e) = \delta(e) + \delta(e') \). This completes the proof. \( \blacksquare \)

**Remark 1.** Unlike the coding schemes for the regular relay channel, we do not need block Markov coding for the noncausal relay channel for two reasons. First, from the channel statistics \( p(y_2|x_1) \), \( Y_2 \) does not depend on \( X_2 \) and hence there is no need to make \( x^n_1 \) correlated with the previous block \( x^n_2 \). Second, \( y^n_2 \) is available noncausally at the relay and hence the signals from the sender and the relay arrive at the receiver in the same block.

**Remark 2.** Taking \( U \) conditionally independent of \( Y_2 \) given \( X_1 \) and setting \( X_2 = U \), the GP-DF lower bound reduces to the DF lower bound in \( [2] \).

The GP-DF lower bound can be strictly tighter than the DF lower bound as shown in the following example.

**Example 1.** Consider a degraded noncausal relay channel \( p(y_2|x_1)p(y_3|x_2,x_2,y_2) = p(y_2|x_1)p(y_3|x_2,y_2) \) depicted in Figure 2. The channel from the sender to the relay is a BEC(1/2) channel, while the channel from the relay to the receiver is clean if \( Y_2 \in \{0,1\} \) and stuck at 1 if \( Y_2 \) is an erasure.

![Channel statistics of the degraded noncausal relay channel](image)

Note that the state of the channel from the relay to the receiver, namely, whether we get an erasure or not, is independent of \( X_1 \). The first term in both the DF lower bound and the GP-DF lower bound is easy to compute as

\[ \max_{p(x_1)} I(X_1;Y_2) = 1/2. \]

Consider the second term in the DF lower bound. Here \( X_2 \) is chosen such that \( Y_2 \to X_1 \to X_2 \) form a Markov chain. By carefully computing the conditional probability \( p(y_3|x_1,x_2) = \sum_{y_2} p(y_2|x_1)p(y_3|x_2,y_2) \) in this specific channel, we can
show that $X_1 \rightarrow X_2 \rightarrow Y_3$ form a Markov chain. Thus,

$$\max_{p(x_1, x_2)} I(X_1, X_2; Y_3) = \max_{p(x_2|x_1)} I(X_2; Y_3)$$

$$= \max_{p(x_2)} I(X_2; Y_3)$$

$$= H(1/5) - 2/5$$

$$= 0.3219,$$

where (a) follows since $X_1 \rightarrow X_2 \rightarrow Y_3$ form a Markov chain, (b) follows since $I(X_2; Y_3)$ is fully determined by the marginal distribution $p(x_2, y_3)$, and (c) follows since the channel from $X_2$ to $Y_3$ $p(y_3|x_2) = \sum_{y_2} p(y_3|x_2, y_2)p(y_2)$ is a $Z$ channel with crossover probability 1/2 regardless of $p(x_1)$. Thus,

$$R_{DF} = \min\{1/2, 0.3219\} = 0.3219.$$

Now consider the second term in the GP-DF lower bound [5]

$$\max_{p(x_1)p(y_2|x_1, y_2)} [I(X_1, U; Y_3) - I(U; Y_2|X_1)].$$

Let $U = X_2 = 1$, if $Y_2 = e$, and $U = X_2 = \text{Bern}(1/2)$, if $Y_2 = 0, 1$. Note that here we always have $Y_3 = X_2 = U$ and $X_1 \rightarrow Y_2 \rightarrow X_2$ form a Markov chain. Thus,

$$I(X_1, U; Y_3) - I(U; Y_2|X_1) = I(X_1, X_2; X_2) - I(X_2; Y_2|X_1)$$

$$= H(X_2) - H(X_2|X_1) + H(X_2|Y_2, X_1)$$

$$= I(X_1; X_2) + H(X_2|Y_2)$$

$$\geq H(X_2|Y_2)$$

$$= 1/2.$$

Therefore,

$$R_{GP-DF} = 1/2 > R_{DF} = 0.3219.$$

Moreover, it is easy to see from the cutset bound [4] that the rate 1/2 is also an upper bound, and hence $C_\infty = 1/2$.

B. Gelfand–Pinsker Compress–Forward Lower Bound

In this subsection, we first propose a two-stage coding scheme that incorporate Gelfand–Pinsker coding with the compress–forward coding scheme, then show an equivalent lower bound can be established directly by applying the recently developed hybrid coding scheme at the relay node.

**Theorem 2** (Gelfand–Pinsker compress–forward (GP-CF) lower bound). The capacity of the noncausal relay channel is lower bounded as

$$C_\infty \geq R_{GP-CF} = \max \min \{I(X_1; U, Y_3), I(X_1, U; Y_3) - I(U; Y_2|X_1)\},$$

where the maximum is over all pmfs $p(x_1)p(u|y_2)$ and functions $x_2(u, y_2)$. 


Outline of the proof: The coding scheme is illustrated in Figure 3. We use Wyner–Ziv binning, multicoding, and joint typicality encoding and decoding. A description \( \hat{y}_R^n \) of \( y_R^n \) is constructed at the relay. Since the receiver has side information \( y_R^n \) about \( \hat{y}_R^n \), we use binning as in Wyner–Ziv coding to reduce the rate necessary to send \( \hat{y}_R^n \). Since the relay has side information \( y_2^n \) of the channel \( p(y_2^n|x_1, x_2, y_2) \), we use multicoding as in Gelfand–Pinkevich coding to send the bin index of \( y_2^n \) via \( u^n \). The decoder first decode the bin index from \( u^n \). It then uses \( u^n \) and \( y_2^n \) to decode \( \hat{y}_R^n \) and \( x_1^n(m) \) simultaneously.

![Fig. 3. GP-CF coding scheme with binning and multicoding](image)

We now provide the details of the coding scheme.

Codebook generation: Fix \( p(x_1)p(u|y_2)p(\hat{y}_2|y_2)x_2(u, \hat{y}_2, y_2) \) that attains the lower bound. Randomly and independently generate \( 2^{nR} \) sequences \( x_1^n(m), m \in [1 : 2^{nR}] \), each according to \( \prod_{i=1}^n P_{X_i}(x_{1i}) \). Randomly and independently generate \( 2^{nR_2} \) sequences \( \hat{y}_R^n(k), k \in [1 : 2^{nR_2}] \), each according to \( \prod_{i=1}^n P_{Y_i}(\hat{y}_{2i}) \). Partition \( k \) into \( 2^{nR_2} \) bins \( B(l_m) \). For each \( l_m \), randomly and independently generate \( 2^{nR_2} \) sequences \( u^n(l|m), l \in [1 : 2^{nR_2}] \), each according to \( \prod_{i=1}^n P_{U_i}(u_i) \), which form subcodebook \( C(l_m) \). This defines the codebook \( C = \{ (x_1^n(m, \hat{y}_R^n(k), u^n(l|m), x_2^n(u, \hat{y}_R^n, y_2^n)) : m \in [1 : 2^{nR}], k \in [1 : 2^{nR_2}], l_m \in [1 : 2^{nR_2}] \}, l \in [1 : 2^{nR}] \}. \) The codebook is revealed to all parties.

Encoding: To send the message \( m \), the encoder transmits \( x_1^n(m) \).

Relay encoding and analysis of the probability of error: Upon receiving \( y_2^n \), the relay first finds the unique \( k \) such that \( (\hat{y}_R^n(k), y_2^n) \in T_{c}(\epsilon) \). This requires \( \hat{R}_2 > I(\hat{Y}_2; Y_2) + \delta(\epsilon) \) by the covering lemma. Upon getting the bin index \( l_m \) of \( k \), i.e., \( k \in B(l_m) \), the relay finds a sequence \( u^n(l|m) \in C(l_m) \) such that \( (u^n(l|m), y_2^n) \in T_{c}(\epsilon) \). This requires \( \hat{R}_2 > I(U; Y_2) + \delta(\epsilon) \) by the covering lemma. The relay transmits \( x_2(\hat{y}_2^n(k), u_i(l|m), y_2^n) \) at time \( i \in [1 : n] \).

Decoding and analysis of the probability of error: Let \( \epsilon > \epsilon' \). Upon receiving \( y_2^n \), the decoder finds the unique \( \hat{l}_m \) such that \( (u^n(l|m), y_2^n) \in T_{c}(\epsilon) \) for some \( u^n(l|m) \in C(l_m) \). This requires \( \hat{R}_2 + R_2 < I(U; Y_3) - \delta(\epsilon) \). The decoder then finds the unique message \( \hat{m} \) such that \( (x_1^n(\hat{m}), \hat{y}_R^n(k), y_2^n) \in T_{c}(\epsilon') \) for some \( k \in B(l_m) \). Let \( K \) be the chosen index for \( \hat{Y}_2^n \) at the relay. If \( \hat{m} \neq 1 \) but \( k = K \), this requires \( R < I(X_1; \hat{Y}_2, Y_3) - \delta(\epsilon) \). If \( \hat{m} \neq 1 \) and \( k \neq K \), this requires \( R + \hat{R}_2 - R_2 < I(X_1; \hat{Y}_2, Y_3) + I(\hat{Y}_2; X_1, Y_3) - \delta(\epsilon) \). Thus, we establish the following lower bound:

\[
C_\infty \geq R'_{GP-CF} = \max \{ I(X_1; \hat{Y}_2, Y_3), I(X_1, \hat{Y}_2; Y_3) - I(\hat{Y}_2; Y_2|X_1) + I(U; Y_3) - I(U; Y_2) \},
\]
where the maximum is over all pmfs \( p(x_1)p(u|y_2)p(\hat{y}_2|y_2) \) and functions \( x_2(u, \hat{y}_2, y_2) \).

Now we show the two lower bounds (7) and (6) are equivalent. Setting \( U = \emptyset \) in \( R_{\text{GP-CF}}' \) and relabeling \( \hat{Y}_2 \) as \( U \), \( R_{\text{GP-CF}}' \) reduces to \( R_{\text{GP-CF}} \). Thus,

\[
R_{\text{GP-CF}}' \geq R_{\text{GP-CF}}.
\]

(8)

On the other hand, letting \( U = (U, \hat{Y}_2) \) in \( R_{\text{GP-CF}} \), we have

\[
I(X_1, U, \hat{Y}_2; Y_3) - I(U, \hat{Y}_2; Y_2|X_1)
\]

\[
= I(X_1, \hat{Y}_2; Y_3) + I(U; Y_3|X_1, \hat{Y}_2) - I(\hat{Y}_2; Y_2|X_1) - I(U; Y_2|X_1, \hat{Y}_2)
\]

\[
= I(X_1, \hat{Y}_2; Y_3) - I(\hat{Y}_2; Y_2|X_1) + H(U|X_1, \hat{Y}_2) - H(U|X_1, Y_2) - H(U|X_1, \hat{Y}_2, Y_2)
\]

\[
\geq (a) I(X_1, \hat{Y}_2; Y_3) - I(\hat{Y}_2; Y_2|X_1) + H(U) - H(U|Y_3) - H(U) + H(U|X_1, \hat{Y}_2, Y_2)
\]

\[
(b) I(X_1, \hat{Y}_2; Y_3) - I(\hat{Y}_2; Y_2|X_1) + I(U; Y_3) - I(U; Y_2),
\]

where \((a)\) follows since conditioning reduces entropy and \((b)\) follows since \((X_1, \hat{Y}_2) \to Y_2 \to U\) form a Markov chain. Furthermore, since the maximum in \( R_{\text{GP-CF}} \) is over a larger set \( p(u, \hat{y}_2|y_2) \) than the set \( p(u|y_2)p(\hat{y}_2|y_2) \) in \( R_{\text{GP-CF}}' \),

\[
R_{\text{GP-CF}} \geq R_{\text{GP-CF}}'.
\]

(9)

Combining (8) and (9) establishes the equivalence.

Remark 3. Taking \( U \) independent of \( Y_2 \) and \( X_2 = U \) in (7), we establish the compress–forward lower bound without Gelfand–Pinsker coding as follows:

\[
C_{\infty} \geq R_{\text{CF}} = \max \min \{I(X_1; \hat{Y}_2, Y_3), I(X_1, \hat{Y}_2; Y_3) + I(X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1)\},
\]

where the maximum is over all pmfs \( p(x_1)p(x_2)p(\hat{y}_2|y_2) \).

In the analysis of the probability of error in Theorem 2 there is a technical subtlety in applying the standard packing lemma and joint typicality lemma, since the bin index \( L_m \), the compression index \( K \), and the multicoding index \( L \) all depend on the random codebook itself. In the following, we show the GP-CF lower bound (6) can be established directly by applying the recently developed hybrid coding scheme for joint source–channel coding by Lim, Minero, and Kim [5], [6], [9].

Proof of Theorem 2 via hybrid coding: In this coding scheme, we apply hybrid coding at the relay node as depicted

![Diagram](image_url)

Fig. 4. Hybrid coding interface at the relay. Illustration from Kim, Lim, and Minero [9]

in Figure [4]. The sequence \( y_2^n \) is mapped to one of \( 2^{nR} \) sequences \( u^n(l) \). The relay generates the codeword \( x_2^n \) through a symbol-by-symbol mapping \( x_2(u, y_2) \). The receiver declares \( \hat{m} \) to be the message estimate if \( (x_2^n(\hat{m}), u^n(l), y_2^n) \) are jointly typical for some \( l \in [1 : 2^{nR}] \). Similar to the hybrid coding scheme for joint source-channel coding [5], [6], [9], the precise
analysis of the probability of decoding error involves a technical subtlety. In particular, since $U^n(L)$ is used as a source codeword, the index $L$ depends on the entire codebook. This dependency issue is resolved by the technique developed in [5]. We now provide the details of the coding scheme.

**Codebook generation:** Fix $p(x_1)p(u_2)$ and $x_2(u_2, y_2)$ that attain the lower bound. Randomly and independently generate $2^{nR}$ sequences $x^n_1(m)$, $m \in [1 : 2^{nR}]$, each according to $\prod_{l=1}^{n} p_{X_1}(x_{1l})$. Randomly and independently generate $2^{n\tilde{R}}$ sequences $u^n(l), l \in [1 : 2^{n\tilde{R}}]$, each according to $\prod_{l=1}^{n} p_{U}(u_l)$. This defines the codebook $C = \{(x^n_1(m), u^n(l), x^n_2(u^n(l), y^n_2)) : m \in [1 : 2^{nR}], l \in [1 : 2^{n\tilde{R}}]\}$. The codebook is revealed to all parties.

**Encoding:** To send message $m$, the encoder transmits $x^n_1(m)$.

**Relay encoding:** Upon receiving $y^n_2$, the relay finds an index $l$ such that $(u^n(l), y^n_2) \in T_{c'}(n)$. If there is more than one such indices, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from $[1 : 2^{n\tilde{R}}]$. The relay then transmits $x_{2i}(u_i(l), y_{2i})$ at time $i \in [1 : n]$.

**Decoding:** Let $\epsilon > \epsilon'$. Upon receiving $y^n_2$, the decoder finds the unique message $\hat{m}$ such that $(x^n_1(\hat{m}), u^n(l), y^n_2) \in T_{\epsilon'}(n)$ for some $l \in [1 : 2^{n\tilde{R}}]$.

**Analysis of the probability of error:** We analyze the probability of decoding error averaged over codes. Let $L$ denote the index of the chosen $U^n$ codeword for $Y^n_2$. Assume without loss of generality that $M = 1$. The decoder makes an error only if one of the following events occur:

$$\tilde{E} = \{(U^n(l), Y^n_2) \notin T_{\epsilon'}(n) \text{ for all } l\},$$

$$E_1 = \{(X^n_1(1), U^n(L), Y^n_2) \notin T_{\epsilon'}(n)\},$$

$$E_2 = \{(X^n_1(m), U^n(L), Y^n_2) \in T_{\epsilon'}(n) \text{ for } m \neq 1\},$$

$$E_3 = \{(X^n_1(m), U^n(l), Y^n_2) \in T_{\epsilon'}(n) \text{ for } m \neq 1, l \neq L\}.$$

By the union of the events bound, the probability of error is upper bounded as

$$P(E) = P(\tilde{E} \cup E_1 \cup E_2 \cup E_3) \leq P(\tilde{E}) + P(E_1 \cap \tilde{E}^c) + P(E_2 \cap \tilde{E}^c) + P(E_3).$$

By the covering lemma, the first term tends to zero as $n \to \infty$ if $\tilde{R} > I(U; Y_2) + \delta(\epsilon')$. By the conditional typicality lemma, the second term tends to zero as $n \to \infty$. By the packing lemma, the third term tends to zero as $n \to \infty$ if $R < I(X_1; U, Y_3) - \delta(\epsilon)$.

The forth term requires special attention. Consider

$$P(E_3) = P\{(X^n_1(m), U^n(l), Y^n_3) \in T_{\epsilon'}(n) \text{ for } m \neq 1, l \neq L\} \leq \sum_{m=2}^{2^{nR}} \sum_{l=1}^{2^{n\tilde{R}}} P\{(X^n_1(m), U^n(l), Y^n_3) \in T_{\epsilon'}(n), l \neq L\} \leq \sum_{m=2}^{2^{nR}} \sum_{k=1}^{n} \sum_{y_2} P\{(X^n_1(m), U^n(l), Y^n_3) \in T_{\epsilon'}(n), l \neq L|Y^n_2 = y^n_2\} p(y^n_2) \leq \sum_{y_2} 2^{n\tilde{R}} \sum_{m=2}^{2^{nR}} P\{(X^n_1(1), U^n(1), Y^n_3) \in T_{\epsilon'}(n), L \neq 1|Y^n_2 = y^n_2\} p(y^n_2),$$

where the inequality $(b)$ follows from the packing lemma.
where (a) follows by the union of events bound and (b) follows by the symmetry of the codebook generation and relay encoding. Let $\tilde{C} = C \setminus \{(X^n_1(2), U^n(1), X^n_3(2), U^n(1))\}$. Then, for $n$ sufficiently large,

$$
\begin{align*}
\mathbb{P}(\{X^n_1(2), U^n(1), Y^n_3(2) \in T^{(n)}_c \mid L \neq 1, Y^n_2 = y^n_2\}) \\
\leq \mathbb{P}(\{X^n_1(2), U^n(1), Y^n_3(2) \in T^{(n)}_c \mid L \neq 1, Y^n_2 = y^n_2\}) \\
= \sum_{(x^n_1,u^n,y^n_3) \in T^{(n)}_c} \mathbb{P}(X^n_1(2) = x^n_1, U^n(1) = u^n, Y^n_3 = y^n_3 \mid L \neq 1, Y^n_2 = y^n_2) \\
= \sum_{(x^n_1,u^n,y^n_3) \in T^{(n)}_c} \sum_{\tilde{C}} \mathbb{P}(X^n_1(2) = x^n_1, U^n(1) = u^n, Y^n_3 = y^n_3 \mid L \neq 1, Y^n_2 = y^n_2, \tilde{C} = \tilde{c}) \mathbb{P}(\tilde{C} = \tilde{c} \mid L \neq 1, Y^n_2 = y^n_2) \\
= \sum_{(x^n_1,u^n,y^n_3) \in T^{(n)}_c} \sum_{\tilde{C}} \mathbb{P}(U^n(1) = u^n \mid L \neq 1, Y^n_2 = y^n_2, \tilde{C} = \tilde{c}) \mathbb{P}(X^n_1(2) = x^n_1 \mid L \neq 1, Y^n_2 = y^n_2, \tilde{C} = \tilde{c}) \mathbb{P}(\tilde{C} = \tilde{c} \mid L \neq 1, Y^n_2 = y^n_2) \\
= \sum_{(x^n_1,u^n,y^n_3) \in T^{(n)}_c} \sum_{\tilde{C}} 2 \mathbb{P}(U^n(1) = u^n) \mathbb{P}(X^n_1(2) = x^n_1) \mathbb{P}(Y^n_3 = y^n_3 \mid L \neq 1, Y^n_2 = y^n_2) \\
\leq \sum_{(x^n_1,u^n,y^n_3) \in T^{(n)}_c} \sum_{\tilde{C}} 2 \mathbb{P}(U^n(1) = u^n) \mathbb{P}(X^n_1(2) = x^n_1) \cdot 2 \mathbb{P}(Y^n_3 = y^n_3 \mid Y^n_2 = y^n_2), \quad (10)
\end{align*}
$$

where (a) follows since given $L \neq 1, U^n(1) \rightarrow (Y^n_2, \tilde{C}) \rightarrow (Y^n_3, X^n_1(2))$ form a Markov chain, (b) follows since for $n$ sufficiently large $\mathbb{P}(U^n(1) = u^n \mid K \neq 1, Y^n_2 = y^n_2, \tilde{C} = \tilde{c}) \leq 2 \mathbb{P}(U^n(1) = u^n)$ and $X^n_1(2)$ is independent of $(Y^n_2, Y^n_3, \tilde{C}, K)$, and (c) follows since for $n$ sufficiently large $\mathbb{P}(Y^n_3 = y^n_3 \mid L \neq 1, Y^n_2 = y^n_2) \leq 2 \mathbb{P}(Y^n_3 = y^n_3 \mid Y^n_2 = y^n_2)$.

The statements in (b) and (c) are established by Lim, Minero, and Kim in [8 Lemmas 1, 2]. Back to the upper bound on $\mathbb{P}(E_3)$, by the joint typicality lemma and (10), we have

$$
\mathbb{P}(E_3) = \mathbb{P}(\{(X^n_m, U^n_k, Y^n_l) \in T^{(n)}_c \mid m \neq 1, l \neq L\}) \\
\leq 4 \cdot 2^n R_2 + 2^n \sum_{y^n_2} p(y^n_2) \sum_{(x^n_1,u^n,y^n_3) \in T^{(n)}_c} \mathbb{P}(X^n_1(2) = x^n_1) \mathbb{P}(U^n(1) = u^n) p(y^n_3 \mid y^n_2) \\
= 4 \cdot 2^n R_2 + 2^n \sum_{(x^n_1,u^n,y^n_3) \in T^{(n)}_c} \mathbb{P}(X^n_1(2) = x^n_1) \mathbb{P}(U^n(1) = u^n) p(y^n_3) \\
\leq 4 \cdot 2^{n(R + \tilde{R} - I(X_1; Y_3) - I(U; X_1, Y_3) + (\epsilon))},
$$

which tends to zero as $n \rightarrow \infty$ if $R + \tilde{R} \leq I(X_1; Y_3) + I(U; X_1, Y_3) - (\epsilon)$. Eliminating $\tilde{R}$ and letting $n \rightarrow \infty$ completes the proof. □

C. Gelfand–Pinsker Partial Decode–Forward Compress–Forward Lower Bound

Finally, we further combine the hybrid coding scheme developed for the GP-CF lower bound with the partial decode–forward coding scheme by El Gamal, Hassanpour, and Mammen [3].
Theorem 3 (GP-PDF-CF lower bound). The capacity of noncausal relay channel is lower bounded as

\[ C_\infty \geq \max \min \{ I(V, U; Y_3) + I(X_1; U, Y_3 | V) - I(U; Y_2 | V), \]

\[ I(V; Y_2) + I(X_1; U, Y_3 | V), \]

\[ I(V; Y_2) + I(X_1; U, Y_3 | V) + I(U; Y_3 | V) - I(U; Y_2 | V), \]

where the maximum is over all pmfs \( p(v, x_1)p(u|v, y_2) \) and functions \( x_2(u, v, y_2) \).

Proof: In this coding scheme, message \( m \in [1:2^{nR}] \) is divided into two independent parts \( m' \) and \( m'' \) where \( m' \in [1:2^{nR'}], m'' \in [1:2^{nR''}] \), and \( R' + R'' = R \). For each message \( m = (m', m'') \), we generate a \( x_1^n(m''|m') \) sequence and a subcodebook \( C(m') \) of \( 2^{nR'} u^n(l|m') \) sequences. To send message \( m = (m', m'') \), the sender transmits \( x_1^n(m''|m') \).

Upon receiving \( y_2^n \) noncausally, the relay decodes the message \( \hat{m}' \), finds a \( u^n(l|\hat{m}') \in C(\hat{m}') \) that is jointly typical with \( (v^n(\hat{m}''), v_2^n) \), and transmits \( x_2^n(v^n(\hat{m}''), v^n(\hat{m}''), y_2^n) \). The receiver declares \( \hat{m} = (\hat{m}', \hat{m}'') \) to be the message estimate if \( (x_1^n(\hat{m}'', \hat{m}'), u^n(l|\hat{m}'), v^n(\hat{m}''), y_2^n) \) are jointly typical for some \( u^n(l|\hat{m}') \in C(\hat{m}') \).

We now provide the details of the coding scheme.

Codebook generation: Fix \( p(v, x_1)p(u|v, y_2)x_2(u, v, y_2) \) that attains the lower bound. Randomly and independently generate \( 2^{nR'} \) sequences \( e^n(m') \), \( m' \in [1:2^{nR'}] \), each according to \( \prod_{i=1}^n p_V(v_i) \). For each message \( m' \in [1:2^{nR'}] \), randomly and conditionally independently generate \( 2^{nR''} \) sequences \( \hat{x}_1^n(m''|m') \) and \( 2^{nR} u^n(l|m') \), each respectively according to \( \prod_{i=1}^n p_{X|V}(x_i|v_i(m')) \) and \( \prod_{i=1}^n p_{U|V}(u_i|v_i(m')) \), which form the subcodebook \( C(m') \). This defines the codebook \( C = \{ (v^n(m'), \hat{x}_1^n(m''|m'), u^n(l|m'), \hat{x}_2^n(u^n(l|m'), v^n(m'), y_2^n) ) : m' \in [1:2^{nR'}], m'' \in [1:2^{nR''}], l \in [1:2^{nR}] \} \). The codebook is revealed to all parties.

Encoding: To send message \( m = (m', m'') \), the encoder transmits \( x_1^n(m''|m') \).

Relay encoding: Upon receiving \( y_2^n \), the relay finds the unique \( \hat{m}' \) such that \( (v^n(\hat{m}''), v_2^n) \in \mathcal{T}_e^{(n)} \). Then, it finds the unique sequence \( u^n(l|\hat{m}') \in C(\hat{m}') \) such that \( (u^n(l|\hat{m}'), v^n(\hat{m}''), y_2^n) \in \mathcal{T}_e^{(n)} \). If there is more than one such index, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from \( [1:2^{nR_2}] \). The relay then transmits \( x_{2i} = x_2(u_i(l|\hat{m}'), v_i(\hat{m}''), y_{2i}) \) at time \( i \in [1:n] \).

Decoding: Let \( \epsilon > \epsilon' \). Upon receiving \( y_3^n \), the decoder declares that \( \hat{m} = (\hat{m}', \hat{m}'') \in [1:2^{nR}] \) is sent if it is the unique message such that \( (x_1^n(\hat{m}'', \hat{m}'), u^n(l|\hat{m}''), v^n(\hat{m}''), y_3^n) \in \mathcal{T}_e^{(n)} \) for some \( u^n(l|\hat{m}') \in C(\hat{m}') \); otherwise, it declares an error.

Analysis of the probability of error: We analyze the probability of error of message \( M \) averaged over codes. Assume without loss of generality that \( M = (M', M'') = (1, 1) \). Let \( \hat{M}' \) be the decoded message at the relay and let \( L \) denote the index of the chosen \( U^n \) codeword for \( \hat{M}' \). The decoder makes an error only if one of the following events occur:

\[ \tilde{E} = \{ \hat{M}' \neq 1 \}, \]

\[ \tilde{E}_1 = \{ (V^n(1), Y_2^n) \notin \mathcal{T}_e^{(n)} \}, \]

\[ \tilde{E}_2 = \{ (V^n(m'), Y_2^n) \in \mathcal{T}_e^{(n)} \text{ for some } m' \neq 1 \}, \]

\[ \tilde{E}_3 = \{ (U^n(l|\hat{M}'), V^n(\hat{M}'), Y_2^n) \notin \mathcal{T}_e^{(n)} \text{ for all } U^n(l|\hat{M}') \in C(\hat{M}') \}, \]

\[ \mathcal{E}_1 = \{ (X_1^n(1)|\hat{M}'), U^n(L|\hat{M}'), V^n(\hat{M}'), Y_3^n) \notin \mathcal{T}_e^{(n)} \}, \]

\[ \mathcal{E}_2 = \{ (X_1^n(m''|1), U^n(L|1), V^n(1), Y_3^n) \in \mathcal{T}_e^{(n)} \text{ for some } m'' \neq 1 \}, \]
\[ \mathcal{E}_3 = \{(X^n_l | m^n | l \neq L) \in \mathcal{T}^{(n)}_\epsilon \text{ for some } m^n \neq 1, l \neq L, \text{ and } U^n(l|1) \in \mathcal{C}(1)\}, \]
\[ \mathcal{E}_4 = \{(X^n_l | m^n | m' \neq 1, m^n \neq 1, U^n(l|1) \in \mathcal{C}(m')\}. \]

By the union of events bound, the probability of error is upper bounded as
\[
P(\mathcal{E}) = P(\hat{M} \neq 1) \leq P(\hat{M} \cup \mathcal{E}_3 \cup \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4)
\leq P(\hat{M}) + P(\mathcal{E}_3 \cap \hat{M}) + P(\mathcal{E}_2) + P(\mathcal{E}_3) + P(\mathcal{E}_4)
\leq P(\hat{M}) + P(\mathcal{E}_2) + P(\mathcal{E}_3 \cap \hat{M}) + P(\mathcal{E}_1 \cap \hat{M} \cap \mathcal{E}_3)
\leq P(\hat{M}) + P(\mathcal{E}_2) + P(\mathcal{E}_3) + P(\mathcal{E}_4).
\]

By the LLN, the first term tends to zero as \( n \to \infty \). By the packing lemma, the second term tends to zero as \( n \to \infty \) if \( R' < I(V; Y_2) - \delta(\epsilon') \). Therefore, \( P(\hat{M}) \) tends to zero as \( n \to \infty \) if \( R' < I(V; Y_2) - \delta(\epsilon') \). Given \( \hat{M} = 1 \), by the covering lemma, the third term tends to zero as \( n \to \infty \) if \( \hat{R} > I(U; Y_3) + \delta(\epsilon') \). By the conditional typicality lemma, the fourth term tends to zero as \( n \to \infty \) if \( R'' < I(X_1; U, Y_3 | V) - \delta(\epsilon) \). The last two terms require special attention because of the dependency between the index \( L \) and the codebook \( C = \{(V^n(m'), X^n_l(m') | m'), U^n(l|m'), X^n_l(U^n(l|m'), V^n(m'), Y^n_2(m'))\} \). With a similar argument as in the analysis for \( P(\mathcal{E}_3) \) in the proof via hybrid coding of Theorem 2 we can show the last two terms tend to zero as \( n \to \infty \) if \( R'' + \hat{R} < I(X_1; Y_3 | V) + I(U; X_1, Y_3 | V) - \delta(\epsilon) \) and \( R' + R'' + \hat{R} < I(V, X_1, U; Y_3) + I(X_1; U | V) - \delta(\epsilon) \) respectively. Eliminating \( R' \), \( R'' \), and \( \hat{R} \) and letting \( n \to \infty \) completes the proof.

Remark 4. Setting \( V = (V, X_2) \) and \( U = \emptyset \), the GP-PDF-CF lower bound reduces to the PDF lower bound \( \mathcal{E}_2 \). Note that the choice of \( X_2 \) gives the Markov chain \( X_2 \to V \to Y_2 \). Furthermore, setting \( V = \emptyset \) and \( U = \hat{Y}_2 \), the GP-PDF-CF lower bound reduces to the GP-CF lower bound \( \mathcal{E}_2 \). However, this lower bound does not recover the GP-DF lower bound \( \mathcal{E}_2 \) in Theorem 1.

IV. AN IMPROVED UPPER BOUND

In this section, we provide an improved upper bound on the capacity, which is tight for the class of degraded noncausal relay channels. We show through an example that the new upper bound can be strictly tighter than the cutset bound.

Theorem 4. The capacity of the noncausal relay channel is upper bounded as
\[ C_\infty \leq R_{\text{NUN}} = \max \min \{I(X_1; Y_2) + I(X_1; Y_3 | X_2, Y_2), I(X_1; U; Y_3) - I(Y_2; U | X_1)\}, \]
where the maximum is over all pmfs \( p(x_1)p(u|x_1, y_2) \) and functions \( x_2(u, x_1, y_2) \).

Proof: The first term in the upper bound follows from the cutset bound \( \mathcal{E}_2 \). To establish the second bound, identify \( U_i = (M, Y_{2,i+1}|Y_3) \). Let \( Q \sim \text{Unif}[1:n] \) be independent of \( (U^n, X^n_1, Y^n_2) \) and set \( U = (U_Q, Q), X_1 = X_{1Q}, Y_2 = Y_{2Q}, Y_3 = Y_{3Q} \). We have
\[ nR = H(M) \]
\[ \leq I(M; Y^n_3) + n\epsilon_n \]
\[ = \sum_{i=1}^{n} I(M; Y^n_3 | Y_{3,i-1}^{i-1}) + n\epsilon_n \]
\[
\leq \sum_{i=1}^{n} I(M, Y_3^{i-1}; Y_3) + n\epsilon_n \tag{12}
\]
\[
= \sum_{i=1}^{n} [I(M, Y_3^{i-1}; Y_3) - I(Y_3^{i-1}; Y_3| Y_3^i, M) + n\epsilon_n]
\]
\[
= \sum_{i=1}^{n} [I(M, Y_3^{i-1}; Y_3) - I(Y_3^{i-1}; Y_3| Y_3^i, M) + n\epsilon_n]
\]
\[
= \sum_{i=1}^{n} [I(X_1, M, Y_3^{i-1}; Y_3) - I(Y_3^{i-1}; Y_3| X_{i+1}, M, X_{i}) + n\epsilon_n]
\]
\[
= \sum_{i=1}^{n} [I(X_1, U; Y_3) - I(Y_2; U| X_1)] + n\epsilon_n
\]
\[
= n[I(X_1, U; Q; Y_3)| Q] - I(Y_2; U| X_1) + n\epsilon_n
\]
\[
= n[I(X_1, U; Q; Q)| Y_3] - I(Y_2; U| Q, X_1) + n\epsilon_n
\]
\[
= n[I(X_1, U; Y_3) - I(Y_2; U| X_1)] + n\epsilon_n,
\]
where (a) follows by Fano’s inequality, (b) follows by Csizsár sum identity, (c) follows since \( X_{i+1} \) is a function of \( M \), and (d) follows since the channel \( p(y_2|x_1) \) is memoryless and thus \( (Y_3^{i-1}, M) \to X_{i+1} \to Y_2 \) form a Markov chain.

Finally, we show that it suffices to maximize over \( p(x_1)p(u|x_1, y_2) \) and functions \( x_2(u, x_1, y_2) \). Consider a general pmf \( p(x_1)p(x_2, u|x_1, y_2) \), by the functional representation lemma \([\text{1}, \text{Appendix B}]\), there exists a random variable \( V \) independent of \( (U, X_1, Y_2) \) such that \( X_2 \) is a function of \( (U, X_1, Y_2, V) \). Now define \( \tilde{U} = (U, V) \). Then

\[
C_{\infty} \leq \max_{p(x_1)p(x_2, u|x_1, y_2)} \min\{I(X_1; Y_2) + I(X_1; X_2| Y_2) I(X_1, U; Y_3) - I(Y_2; U| X_1)\}
\]
\[
= \max_{p(x_1)p(u|x_1, y_2)} \min\{I(X_1; Y_2) + I(X_1; X_2| Y_2) I(X_1, U; V; Y_3) - I(Y_2; U, V| X_1)\}
\]
\[
\leq \max_{p(x_1)p(u|x_1, y_2)} \min\{I(X_1; Y_2) + I(X_1; Y_3| X_2, Y_2) I(X_1, \tilde{U}; Y_3) - I(Y_2; \tilde{U}| X_1)\}.
\]

Thus, there is no loss of generality in restricting \( X_2 \) to be a function of \((U, X_1, Y_2)\).

\[\square\]

**Remark 5.** This upper bound is always tighter than the cutset bound. To see this, note that the new upper bound is equivalent to expression \(12\) with all the remaining steps being equality. On the other hand, the cutset bound can be derived from \(12\) as

\[
nR \leq \sum_{i=1}^{n} I(M, Y_3^{i-1}; Y_3) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^{n} I(X_{i+1}, X_{i}; M, Y_3^{i-1}; Y_3) + n\epsilon_n
\]
\[
= \sum_{i=1}^{n} I(X_{i+1}, X_{i}; Y_3) + n\epsilon_n
\]
\[
= nI(X_1, X_2; Y_3) + n\epsilon_n,
\]
where (a) can be loose in general.
Theorem 5. The capacity of the degraded noncausal relay channel \( p(y_2|x_1)p(y_3|x_2, y_2) = p(y_2|x_1)p(y_3|x_2, y_2) \) is

\[
C_\infty = \max \min \{I(X_1; Y_2), I(X_1, U; Y_3) - I(Y_2; U|X_1)\},
\]

where the maximum is over all pmfs \( p(x_1)p(u|x_1, y_2) \) and functions \( x_2(u, x_1, y_2) \).

Proof: In the degraded case, we have \( I(X_1; Y_2) + I(X_1; Y_3|X_2, Y_2) = I(X_1; Y_2) \). Thus, the GP-DF lower bound in Theorem 1 and the improved upper bound in Theorem 4 coincide.

The improved upper bound on the capacity of the noncausal relay channel can be strictly tighter than the cutset bound. In the following, we provide an example, motivated by [4, Example 2], where \( R_{DF} < R_{GP-DF} = C_\infty = R_{NUB} < R_{CS} \).

Example 2. Consider a degraded noncausal relay channel \( p(y_2|x_1)p(y_3|x_2, y_2) = p(y_2|x_1)p(y_3|x_2, y_2) \) as depicted in Figure 5. The channel from the sender to the relay is BSC(\( p_1 \)), while the channel from the relay to the receiver is BSC(\( p_2 \)) if \( Y_2 = 0 \) and BSC(\( p_3 \)) if \( Y_2 = 1 \).

![Figure 5. Channel statistics of the degraded noncausal relay channel](image)

When \( p_1 = 0.2, p_2 = 0.1, \) and \( p_3 = 0.55 \), we have

\[
R_{DF} = 0.2203,
\]

\[
R_{CS} = 0.2566,
\]

\[
R_{GP-DF} = C_\infty = R_{NUB} = 0.2453.
\]

The DF lower bound and cutset bound expressions (2) and (4) contain no auxiliary random variable and thus can be computed easily. In the capacity expression (13) in Theorem 5, the maximum is attained by \( U \sim \text{Bern}(1/2) \) independent of \( (X_1, Y_2) \) and \( X_2 = U \oplus Y_2 \), which yields the capacity \( C_\infty = 0.2453 \).

We prove this via a symmetrization argument motivated by Nair [7]. Note that

\[
C_\infty = \max_{p(x_1)} \min_{p(x_2)} \left\{ I(X_1; Y_2), I(X_1, U; Y_3) - I(U; Y_2|X_1) \right\}
\]

\[
= \max_{p(x_1)} \min_{p(x_2)} \left\{ I(X_1; Y_2), \max_{p(u|x_1, y_2)} (I(X_1, U; Y_3) - I(U; Y_2|X_1)) \right\}. \tag{14}
\]

Consider the maximum in the second term for a fixed \( p(x_1) \). Assume without loss of generality that \( U = \{1, 2, \ldots, |U|\} \).

For any conditional pmf \( p_{U|X_1, y_2}(u|x_1, y_2) \) and function \( x_2(u, x_1, y_2) \), define \( \tilde{X}_2 \), \( \tilde{X}_3 \), and \( \tilde{Y}_3 \) as

\[
p_{\tilde{U}}(u) = p_{\tilde{U}}(-u) = \frac{1}{2} p_U(u), \quad u \in U,
\]

\[
p_{X_1, y_3|\tilde{U}}(x_1, y_2|u) = p_{X_1, y_3|\tilde{U}}(x_1, y_2|u) = p_{X_1, y_3|\tilde{U}}(x_1, y_2|u), \quad u \in U,
\]

\[
\tilde{x}_2(u, x_1, y_2) = 1 - \tilde{x}_2(-u, x_1, y_2) = x_2(u, x_1, y_2), \quad (u, x_2) \in U \times \{0, 1\}, \tag{15}
\]
\[ p_{Y_3|\hat{X}_2, Y_2}(y_3|\hat{x}_2, y_2) = p_{Y_3|X_2, Y_2}(y_3|\hat{x}_2, y_2), \quad y_3 \in \{0, 1\}. \]

Then for any \( u \in \mathcal{U} \),

\[
\begin{align*}
p_{U|X_1, Y_2}(u|x_1, y_2) &= p_{U|X_1, Y_2}(-u|x_1, y_2) = \frac{1}{2}p_{U|X_1, Y_2}(u|x_1, y_2), \\
p_{Y_3|X_1, U, \tilde{U}}(y_3|x_1, u) &= p_{Y_3|X_1, U, \tilde{U}}(y_3|x_1, -u) = p_{Y_3|X_1, U}(y_3|x_1, u), \\
p_{Y_3|X_1, U}(y_3|x_1, u) &= 1 - p_{Y_3|X_1, U}(y_3|x_1, -u) = p_{Y_3|X_1, U}(y_3|x_1, u).
\end{align*}
\]

Thus, \( H(Y_2|X_1, U = u) = H(Y_2|X_1, \tilde{U} = u) = H(Y_2|X_1, \tilde{U} = -u) \) for all \( u \in \mathcal{U} \), which implies that \( H(Y_2|X_1, U) = H(Y_2|X_1, \tilde{U}) \). Similarly, we can show \( H(Y_3|X_1, U) = H(\tilde{Y}_3|X_1, \tilde{U}) \). It can be also easily shown that for any \( y_2 \in \{0, 1\} \), \( p_{\hat{X}_2|Y_2}(0|\hat{y}_2) = 1/2 \), which implies that \( p_{\hat{Y}_3}(0) = 1/2 \) and \( H(\tilde{Y}_3) = 1 \). Hence,

\[
I(X_1; U; Y_3) - I(U; Y_2|X_1) = H(\tilde{Y}_3) - H(Y_3) - H(Y_3, U) - H(Y_2, X_1, U) \\
\leq H(\tilde{Y}_3) - H(\tilde{Y}_3, X_1, \tilde{U}) - H(Y_2, X_1) + H(Y_2, X_1, \tilde{U}) \\
= \sum_{u=1}^{\mid \mathcal{U} \mid} p_U(u) (H(Y_2|X_1, \tilde{U}, |\tilde{U}| = u) - H(\tilde{Y}_3|X_1, \tilde{U}, |\tilde{U}| = u)) + H(\tilde{Y}_3) + H(\tilde{Y}_3, X_1, \tilde{U}),
\]

where the last maximum is attained by \( p_{U}(U) = p_{U}(-U) = 1/2 \) for a single \( u \). Note that from our definiton of \( \tilde{U} \), this automatically guarantees the independence between \( \tilde{U} \) and \( (X_1, Y_2) \). Therefore, the maximum in the second term of (14) is attained by \( \tilde{U} \sim \text{Bern}(1/2) \) independent of \( (X_1, Y_2) \). Subsequently, we relabel \( \tilde{U} \) as \( U \) with alphabet \( \{0, 1\} \), \( \tilde{Y}_3 \) as \( Y_3 \), and \( \tilde{X}_2 \) as \( X_2 \).

Now we further optimize the second term in (14), which we have simplified as

\[
I(X_1, U; Y_3) - I(U; Y_2|X_1) \overset{(a)}{=} I(X_1, U; Y_3) \\
\overset{(b)}{=} 1 - H(Y_3|X_1, U),
\]

where \( (a) \) follows by the optimal choice of \( U \) independent of \( (X_1, Y_2) \) and \( (b) \) follows since \( H(Y_3) = 1 \). We maximize (18) over all functions \( \hat{x}_2(\hat{u}, x_1, y_2) \) satisfying \( x_2(0, x_1, y_2) = 1 - x_2(1, x_1, y_2) \) for all \( (x_1, y_2) \in \{0, 1\}^2 \). By the symmetry of \( U \) as described in (17), \( H(Y_3|X_1, U = 0) = H(Y_3|X_1, U = 1) \). Thus,

\[
H(Y_3|X_1, U) = p_U(0)H(Y_3|X_1, U = 0) + p_U(1)H(Y_3|X_1, U = 1) \\
= H(Y_3|X_1, U = 0) \\
= p_{X_1}(0)H(Y_3|X_1 = 0, U = 0) + p_{X_1}(1)H(Y_3|X_1 = 1, U = 0).
\]

By considering all four functions \( x_2(u = 0, x_1 = 0, y_2) \in \{\{0, 1\} \rightarrow \{0, 1\}\} \) and removing the redundant choices by the symmetry of the binary entropy function, we have

\[
H(Y_3|X_1 = 0, U = 0) \geq \min\{H(p_1\bar{p}_2 + \bar{p}_1\bar{p}_3), H(p_1\bar{p}_2 + \bar{p}_1p_3)\}.
\]

Similarly,

\[
H(Y_3|X_1 = 1, U = 0) \geq \min\{H(p_1\bar{p}_2 + \bar{p}_1p_3), H(\bar{p}_1\bar{p}_2 + p_1p_3)\}.
\]
When $p_1 = 0.2$, $p_2 = 0.1$, and $p_3 = 0.55$, the minimum is attained by $X_2 = U \oplus Y_2$ for both terms regardless of $p(x_1)$. Therefore the second term in (14) simplifies to

$$1 - p_{X_1}(0)H(p_1\bar{p}_2 + \bar{p}_1p_3) - p_{X_1}(1)H(\bar{p}_1\bar{p}_2 + p_1p_3).$$

Finally, maximizing

$$\min\{I(X_1;Y_2), 1 - p_{X_1}(0)H(p_1\bar{p}_2 + \bar{p}_1p_3) - p_{X_1}(1)H(\bar{p}_1\bar{p}_2 + p_1p_3)\}$$

over $p(x_1)$, we obtain the capacity $C_\infty = 0.2453$.

REFERENCES


