State-Dependent Relay Channel: Achievable Rate and Capacity of a Semideterministic Class

Majid Nasiri Khormuji, *Member, IEEE*, Abbas El Gamal, *Fellow, IEEE*, and Mikael Skoglund, *Senior Member, IEEE*

***Abstract*—This paper considers the problem of communicating over a relay channel with state when noncausal state information is partially available at the nodes. We first establish a lower bound on the achievable rates based on noisy network coding and Gelfand–Pinsker coding, and show that it provides an alternative characterization of a previously known bound. We then introduce the class of state-decoupled relay channels and show that our lower bound is tight for a subclass of semideterministic channels. We also compute the capacity for two specific examples of this subclass—a channel with multiplicative binary fading and a channel with additive Gaussian interference. These examples are not special cases of previous classes of semideterministic relay channels with known capacity.**

***Index Terms*—Capacity, channels with state, compress-and-forward (CF), noisy network coding (NNC), semideterministic relay channels.**

I. Introduction

The relay channel introduced by van der Meulen in 1972 is one of the main building blocks of network information theory [1]. The capacity, however, is known only for some special classes, including reversely degraded [2], degraded [2], and the channel with orthogonal transmit components [3] as well as the following semideterministic cases.

1) In [4], El Gamal and Aref showed that if $y_2 = f(x_1, x_2)$ (i.e., the received signal at the relay is a deterministic function of the symbols transmitted from the source and the relay), then the partial decode-and-forward scheme in [2] achieves the capacity.
2) In [5] and [6], Cover and Kim proved that if the relay has an orthogonal noiseless link to the destination and $y_2 = f(x_1, y_3)$ (i.e., the received signal at the relay is a deterministic function of the symbols transmitted from the source and the received signal at the destination), then either compress-and-forward (CF) [2] or hash-and-forward relaying [5], [6] achieves the capacity.

In this paper, we study the relay channel with state, which consists of a sender (i.e., source), a relay, and a destination. We assume that the state is determined by a random parameter, and that noncausal knowledge of the state is partially available at the nodes, cf., [7]–[9] for related work. We present a general lower bound on the achievable rates using the noisy network coding (NNC) scheme in [10] and the Gelfand–Pinsker multicoding scheme [11]. Although this achievable rate coincides with that in [9], it provides an alternative characterization that leads to a nontrivial optimality result. We then introduce the special class of state-decoupled relay channels and show that the general lower bound is tight for a semideterministic subclass of these channels. In particular, our results generalize the capacity results reported in [5, Sec. VIII]. We also compute the capacity for two specific examples, where we assume multiplicative binary fading and additive Gaussian interference, respectively. These examples are not special cases of any of the semideterministic relay channels studied in [4]–[6].

A. Organization

The remainder of this paper is organized as follows. Section II presents the main channel model and discuss an achievable rate. Section III introduces the state-decoupled relay channel and establishes a capacity result. Section IV quantifies the capacity of a state-decoupled relay channel with antipodal fading. Section V computes the capacity of a state-decoupled relay channel with additive Gaussian interference. Finally, Section VI concludes the paper.

II. Relay Channel With State

The discrete memoryless relay channel with random state and partial channel state information at the nodes

$$(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3|x_1, x_2, s), \mathcal{Y}_2 \times \mathcal{Y}_3) \cdot (\mathcal{S}, p(s_1, s_2, s_3|s)p(s), \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3)$$

is depicted in Fig. 1. The channel parameters are as follows.

1) $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$ denote the symbol transmitted from the encoder and the relay, respectively;
2) $y_2 \in \mathcal{Y}_2$ and $y_3 \in \mathcal{Y}_3$ are the symbol received at the relay and the decoder, respectively;
3) $s \in \mathcal{S}$ is the true random state of the channel and $s_1 \in \mathcal{S}_1$, $s_2 \in \mathcal{S}_2$, and $s_3 \in \mathcal{S}_3$ denote the partial knowledge of the channel state at the encoder, the relay, and the decoder, respectively;
4) $p(y_2, y_3|x_1, x_2, s)$ denotes the positive matrix factorization (pmf) modeling the interaction between the variables (x_1, x_2, s, y_2, y_3), whose n-extension is

$$p(y_{2,1}^n, y_{3,1}^n|x_{1,1}^n, x_{2,1}^n, s^n) = \prod_{i=1}^n p(y_{2,i}, y_{3,i}|x_{1,i}, x_{2,i}, s_i).$$

Manuscript received May 05, 2011; revised April 12, 2012; accepted December 15, 2012. Date of publication January 09, 2013; date of current version April 17, 2013. This work was supported in part by the Swedish Research Council. The material in this paper was presented in part at the 45th Annual Conference on Information Sciences and Systems, March 2011.

M. N. Khormuji and M. Skoglund are with the School of Electrical Engineering and the ACCESS Linnaeus Center, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (e-mail: khormuji@ee.kth.se; skoglund@ee.kth.se).

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

Communicated by E. Erkip, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2013.2238579

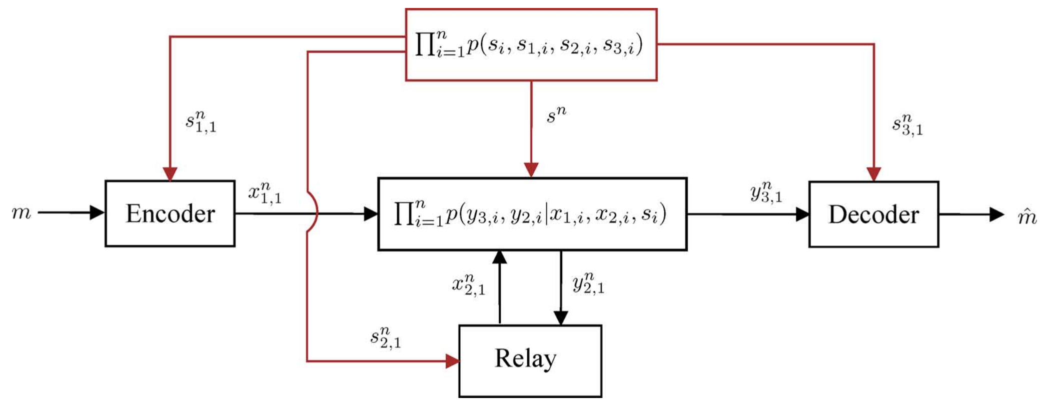

Fig. 1. Three-node relay channel with random state. Partial knowledge of the state is assumed to be noncausally known at the source, the relay, and the destination.

5) $p(s_1, s_2, s_3|s)p(s)$ denotes the pmf modeling the interaction between the true random state of the channel and its partial knowledge at the encoder, the relay, and the decoder. The n-extension is

$$p(s^n, s_{1,1}^n, s_{2,1}^n, s_{3,1}^n) = \prod_{i=1}^{n} p(s_{1,i}, s_{2,i}, s_{3,i}|s_i)p(s_i).$$

A $(2^{nR}, n)$ code for the relay channel with state consists of

1) an *encoder* that maps the message M uniformly drawn from the set $\mathcal{M} := [1 : 2^{nR}]$ to $X_{1,1}^n$ according to the mapping $\alpha : \mathcal{M} \times \mathcal{S}_{1,1}^n \to \mathcal{X}_{1,1}^n$ such that $X_{1,1}^n = \alpha(M, S_{1,1}^n)$. That is, the encoder incorporates the noncausal partial information available about the state.
2) a set of *relay functions*: $\{f_i\}_{i=1}^n$ such that $\forall i \in [1 : n]$, $X_{2,i} = f_i\left(Y_{2,1}^{i-1}, S_{2,1}^n\right)$. That is, the relay acts in a strictly causal manner on the received signals but it incorporates the noncausal knowledge of the channel state.
3) a *decoder* that maps the received signal $Y_{3,1}^n$ to an estimate of the transmitted message $\hat{M}$ according to the mapping $\beta : \mathcal{Y}_{3,1}^n \times \mathcal{S}_{3,1}^n \to \mathcal{M}$ where $\hat{M} = \beta(Y_{3,1}^n, S_{3,1}^n)$.

Remark 1: Since we assume block decoding at the destination, assuming causal knowledge of the channel state at the destination does not affect our results.

The rate R is said to be achievable if there exists a sequence of communication strategies

$$\mathfrak{S}_n := (\alpha^{(n)}, \{f_i\}_{i=1}^n, \beta^{(n)})$$

such that the average error probability at the decoder defined as $\mathbb{P}_e^{(n)} := \Pr\{\hat{M} \neq M\}$ goes to zero as $n \to \infty$. The capacity of the channel is defined as the supremum of all achievable rates.

In [9], it is shown that the following rate is achievable for the scenario in Fig. 1:

$$R < \sup\{I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2)\} \quad (1)$$
$$\text{s.t. } I(\hat{Y}_2; Y_2, S_2|U_2, Y_3, S_3) \leq I(U_2; Y_3, S_3) - I(U_2; S_2) \quad (2)$$

where the supremum is taken over pmfs of the form

$$p(x_1, u_1|s_1)p(x_2, u_2|s_2)p(\hat{y}_2|y_2, u_2, s_2). \quad (3)$$

Achievability of the rate in (1) is established using a Wyner–Ziv-based CF strategy combined with Gelfand–Pinsker coding.[1] Our first result is to establish the following equivalent characterization of the achievable rate.

Proposition 1: The rate $R < \sup \min\{R_1, R_2\}$ is achievable, where

$$R_1 = I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2) \quad (4)$$
$$R_2 = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) - I(U_1, U_2; S_1, S_2) \quad (5)$$

and the supremum is taken over the pmf given in (3).

The proof of this proposition follows by algebraic manipulation of the expressions in (1) and (2), in a similar approach to that in [12] and is given in Appendix A.

Remark 2 (An Alternative Scheme): In Appendix B, we describe an alternative coding scheme that also achieves the rate characterized in Proposition 1. Our new scheme is constructed using NNC and Gelfand–Pinsker coding and in contrast to that in [9] does not employ Wyner–Ziv coding. We also note that the compression of both the received signal Y_2 and knowledge of the channel state S_2 at the relay are considered in obtaining an achievable rate.

Remark 3 (Causal State Information): If the state is available causally at the source and the relay, i.e., $X_{1,i} = \alpha(M, S_{1,1}^i)$ and $X_{2,i} = f_i\left(Y_{2,1}^{i-1}, S_{2,1}^i\right)$, then the rate $R < \sup \min\{R_1, R_2\}$ is achievable where

$$R_1 = I(U_1; \hat{Y}_2, Y_3, S_3|U_2), \quad (6)$$
$$R_2 = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) \quad (7)$$

and the supremum is taken over pmfs of the form

$$p(u_1)p(x_1|u_1, s_1)p(u_2)p(x_2|u_2, s_2)p(\hat{y}_2|y_2, u_2, s_2).$$

Remark 4 (No State Information): If $S_1 = S_2 = S_3 = \emptyset$, then the expression in Proposition 1 simplifies to that in [10], where

$$R_1 = I(X_1; \hat{Y}_2, Y_3|X_2), \quad (8)$$
$$R_2 = I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_3) \quad (9)$$

[1]CF can actually obtain higher rate than the one given in [9]. This is because (30) in [9] should read $I(U_1; Y_3, \hat{Y}_2, U_2)$, since U_1 and U_2 are correlated, and S_1 and S_2 are correlated through S. Thus, the expression given by (1) in [9] changes to (1) in this paper.

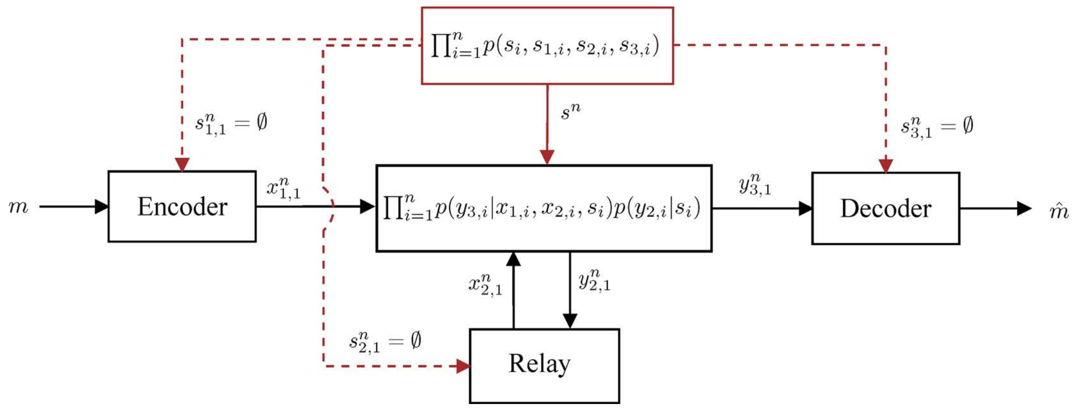

Fig. 2. State-decoupled relay channel.

and the supremum is taken over pmfs of the form

$$p(x_1)p(x_2)p(\hat{y}_2|y_2, x_2).$$

III. State-Decoupled Relay Channels

In this section, we specialize the general setup illustrated in Fig. 1 to the state-decoupled case and show that the bound in Proposition 1 is tight for a new semideterministic subclass of these channels.

We first discuss a motivating example. Consider a network where there is a node that interferes with the main source–destination pair. If another node (e.g., a relay) in the network overhears the communication, it can assist the destination by providing some information regarding the resulting interference. This scenario can be modeled by the discrete memoryless state-decoupled relay channel $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S}, p(y_3|x_1, x_2, s)p(y_2|s)p(s), \mathcal{Y}_2 \times \mathcal{Y}_3)$ depicted in Fig. 2. Note that this channel is a special case of that defined in Section II with $(X_{1,i}, X_{2,i}, Y_{3,i}) \to S_i \to Y_{2,i}$ forming a Markov chain.

In this section, we assume that the sender has *no* knowledge of the channel state and the relay and the destination are informed about the channel state *only* through Y_2 and Y_3, respectively. We now show that under this condition of state information availability, the achievable rate in Proposition 1 is optimal for the semideterministic special case with $Y_2 = f(X_1, X_2, Y_3)$.

Theorem 1: The capacity of the semideterministic state-decoupled relay channel with strictly causal relaying is

$$C = \max_{p(x_1)p(x_2)} \min\{I(X_1, X_2; Y_3), I(X_1; Y_3|X_2, Y_2)\}. \quad (10)$$

Proof: We first prove the positive part. By Remark 4, the capacity is bounded as

$$C \geq \max_{p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)} \min\{R_1, R_2\} \quad (11)$$

where

$$R_1 = I(X_1; Y_3, \hat{Y}_2|X_2) \quad (12)$$
$$R_2 = I(X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2|X_1, X_2, Y_3). \quad (13)$$

Now, let $\hat{Y}_2 = Y_2$. By the channel model assumption, $X_1(M) \to S \to Y_2 \to X_2$ form a Markov chain and M is independent of S; therefore

$$R_1 = I(X_1; Y_3, Y_2|X_2) = I(X_1; Y_3|X_2, Y_2). \quad (14)$$

Since $Y_2 = f(X_1, X_2, Y_3)$, we also have

$$R_2 = I(X_1, X_2; Y_3). \quad (15)$$

This completes the proof of achievability. The converse follows by the cutset bound and noting that the symbol $X_{1,i}(M)$ only depends on the message, which is independent of the symbol $X_{2,i}(Y_{2,1}^{i-1})$ that depends on the channel state. This completes the proof. ■

Remark 5: Theorem 1 subsumes Theorem 3 in [5, Sec. VIII]. This essentially follows by the fact that the channel model in Theorem 1 includes the channel model in [5, Sec. VIII] as a special case. To see this, let $Y_2 = S$, $Y_3 = (Y_{3r}, Y_{3d})$, and $p(y_3|x_1, x_2, s) = p(y_{3d}|x_1, s)p(y_{3r}|x_2)$ in the general state-decoupled relay channel. Here, Y_{3d} is the signal received over the direct link from the sender and Y_{3r} is the signal received from the relay over an orthogonal link. Then, without loss of generality, we can replace the link from the relay to the destination with a noiseless link with the rate given by $R_0 = \max_{p(x_2)} I(X_2; Y_{3r})$.

IV. Fading Relay Channel

In this section, we present an example of a semideterministic relay channel and establish its capacity with strictly causal, causal, and noncausal state information at the relay.

Consider the state-decoupled semideterministic relay channel with

$$Y_3 = SX_1 + X_2 \quad (16)$$
$$Y_2 = S \quad (17)$$

where $X_1, X_2, S \in \{+1, -1\}$ and

$$\Pr\{S = +1\} = \Pr\{S = -1\} = \frac{1}{2}.$$

This example models antipodal signaling with uniform phase fading at high signal-to-noise ratios. We assume that the sender has no knowledge of S and the relay knows S through Y_2. The knowledge of the channel state is transmitted from the relay to the destination over a common channel shared by the sender. That is, there is *no* orthogonal channel devoted to convey the channel state information to the destination. This is in contrast to previous examples in the literature in which knowledge of the channel state is conveyed to the destination over an orthogonal noisy, noiseless, or rate-limited link.

We first note that the direct-link transmission does not contribute to any positive reliable rate since

$$\begin{aligned} R &= \max_{p(x_1),x_2} I(X_1;Y_3|X_2=x_2) \\ &= \max_{p(x_1),x_2} \{H(Y_3|X_2=x_2) - H(Y_3|X_1,X_2=x_2)\} \\ &= \max_{p(x_1)}\{H(SX_1) - H(S)\} \\ &= 0. \end{aligned} \tag{18}$$

While the direct-link transmission fails, we next show that one can reliably transmit at positive rates by appropriately incorporating the relay.

A. Capacity With Strictly Causal Relaying

Proposition 2: The capacity of the relay channel described by (16) and (17) with strictly causal relaying, i.e., $X_{2,i}(S^{i-1})$ is $C = 0.5$ bits per transmission.

Proof: From (16) and (17), we have $Y_2 = (Y_3 - X_2)/X_1$. Therefore, the capacity of the channel is given by Theorem 1. In order to compute the capacity, let $\Pr\{X_2 = +1\} = p$, $\Pr\{X_2 = -1\} = 1-p$, $\Pr\{X_1 = +1\} = t$, $\Pr\{X_1 = -1\} = 1-t$. Then, consider

$$\begin{aligned} I(X_1;Y_3|X_2,Y_2) &= H(Y_3|X_2,S) - H(Y_3|X_1,X_2,S) \\ &= H(SX_1 + X_2|X_2,S) \\ &\quad - H(SX_1+X_2|X_1,X_2,S) \\ &= H(X_1) \\ &= \mathbb{H}_{\rm b}(t) \end{aligned} \tag{19}$$

where $\mathbb{H}_{\rm b}(t) := -t\log_2(t) - (1-t)\log_2(1-t)$ denotes the binary entropy function. Similarly, consider

$$\begin{aligned} I(X_1,X_2;Y_3) &= H(Y_3) - H(Y_3|X_1,X_2) \\ &= H(Y_3) - H(SX_1+X_2|X_1,X_2) \\ &= H(Y_3) - H(S) \\ &= \frac{1}{2}\mathbb{H}_{\rm b}(p) \end{aligned} \tag{20}$$

where the last equality follows because

$$\Pr\{Y_3=0\} = \frac{1}{2}, \Pr\{Y_3=2\} = \frac{p}{2}, \Pr\{Y_3=-2\} = \frac{1-p}{2}.$$

Combining (19) and (20) yields

$$C = \max_{t,p} \min\left\{\mathbb{H}_{\rm b}(t), \tfrac{1}{2}\mathbb{H}_{\rm b}(p)\right\} = 0.5. \tag{21}$$

This completes the proof. ■

TABLE I
INPUT AND OUTPUT SIGNALS USING THE CAUSAL RELAY MAPPING $X_{2i} = S_i$

X_1	S	$X_2 = S$	$Y_3 = S(X_1+1)$
+1	+1	+1	+2
+1	−1	−1	−2
−1	+1	+1	0
−1	−1	−1	0

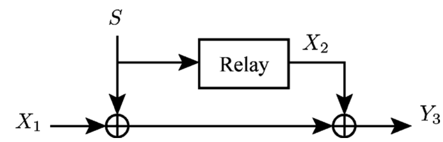

Fig. 3. Semideterministic relay channel with additive Gaussian interference.

B. Capacity With Causal and Noncausal Relaying

Now assume that the signal transmitted from the relay at time i can depend also on the current received channel state at the relay, i.e., $X_{2,i}(S^i)$. We next give a simple deterministic code that achieves the capacity of 1 bit per transmission using instantaneous relaying. Therefore, instantaneous relaying outperforms *strictly causal* relaying for this example.

Proposition 3: The capacity of the relay channel given by (16) and (17) where $X_{2,i}(S^l)$, $\forall l \in [i:n]$ is $C = 1$ bit per transmission.

Proof: The converse is immediate since $H(X_1) \le 1$. We next prove achievability using a simple scheme as an instance of the rate discussed in Remark 3. Let the source uniformly choose its symbol from the set $\{+1,-1\}$ and let the relay use the instantaneous mapping $X_{2,i} = S_i$. Then, the inputs and outputs of the channel are given in Table I. Now, it is easy to observe that the destination can recover X_1 from Y_3 without any error. Therefore, one error-free bit can be transmitted from the sender to the destination. By the upper bound, considering future received symbols at the relay, i.e., S_{i+1}^n, does not buy us any gain for this particular example. ■

V. Relay Channel With Additive Interference

We consider a second example of state-decoupled semideterministic relay channels and compute its capacity with strictly causal, causal, and noncausal relaying. We show that one can achieve higher rates by causal relaying in which the relay employs a nonlinear memoryless strategy.

Consider the state-decoupled relay channel with

$$Y_3 = X_1 + S + X_2 \tag{22}$$
$$Y_2 = S \tag{23}$$

where $S \sim \mathcal{N}(0,Q)$. Further assume that the sender and the relay operate under average power constraints: $\mathbb{E}X_1^2 \le P$ and $\mathbb{E}X_2^2 \le P_r$ (see Fig. 3 for an illustration). This example models an interference-limited scenario in which the additive noise at the relay and the destination is negligible. We assume that the sender does not know the interference, but that the relay perfectly knows it and what it transmits also creates an interference at the destination. A similar model but with an orthogonal link from the relay to the destination is investigated in [13, Sec. VI].

A. Capacity With Strictly Causal Relaying

Proposition 4: The capacity of the relay channel described by (22) and (23) with $S \sim \mathcal{N}(0,Q)$ and strictly causal relaying is

$$C = \frac{1}{2}\log_2\left(1 + \frac{P+P_r}{Q}\right). \quad (24)$$

Proof: We note that the received signal at the relay can be constructed from X_1, X_2, and Y_3. That is, $S = Y_3 - X_1 - X_2$. Thus, we can apply Theorem 1. The proof then follows by noting that (10) is optimized by choosing $X_1 \sim \mathcal{N}(0,P)$ and $X_2 \sim \mathcal{N}(0,P_r)$. ■

Remark 6: As one of the anonymous reviewers brought to our attention, the aforementioned capacity result can also be recovered from that reported in [14, Sec. II].

Remark 7: The relay channel shown in Fig. 3 can be generalized as follows. Let

$$Y_3 = f_1(X_1, X_2) + f_2(S) \quad (25)$$
$$Y_2 = f_3(S) \quad (26)$$

where f_1 and f_3 are arbitrary functions, f_2 is an invertible function, and S denotes the channel state with an arbitrary distribution. Note that the relay channel given in (25) and (26) is state decoupled and

$$Y_2 = f_3\left(f_2^{-1}\left(Y_3 - f_1(X_1, X_2)\right)\right) =: f(X_1, X_2, Y_3). \quad (27)$$

Thus, the capacity of the channel is achieved by NNC or CF and is given in (10).

B. Capacity With Causal and Noncausal Relaying

Proposition 5: The capacity of the relay channel described by (22) and (23) where $X_{2,i}(S^l), \forall l \in [i:n]$ is unbounded.

Proof: We prove the claim by constructing a *nonlinear instantaneous strategy*. In the following, let $X_{2,i} = f(S_i)$ where $f(\cdot)$ is a deterministic function. Note that one can always choose the function f such that $S + f(S) \in \mathcal{D}_s$, where $\mathcal{D}_s$ is a countable set whose elements are chosen from the real line. The smaller the power at the relay, the denser the set $\mathcal{D}_s$ should be chosen in order to meet the power constraint at the relay. By this choice of f, the received signal at the destination is given by

$$Y_3 = X_1 + S + X_2 = X_1 + Z_D$$

where $Z_D := S + X_2 \in \mathcal{D}_s$ denotes the equivalent additive *discrete* noise. Next, let X_1 be uniformly distributed over the interval $(-\frac{d}{2}, +\frac{d}{2})$, where $d \le d_{\min} := \min_{i \neq j}|d_i - d_j|$ for all $d_i, d_j \in \mathcal{D}_s$ and it also satisfies $\mathbb{E}X_1^2 \le P$. Because the effective interference is discrete, the destination can exactly recover X_1 from Y_3, and hence, an arbitrarily high transmission rate is achievable. This scheme can also be interpreted as having the relay transmit the error in quantizing S such that $\mathcal{D}_s$ is the set of the reconstruction points of the quantizer, cf., the approach proposed in [15] (see [15, Fig. 4]). ■

The channel shown in Fig. 3 is intimately related to the point-to-point *dirty tape channel*, in which the received signal is given by $Y_3 = X_1 + S + Z$, where X_1 is the transmitted signal, S is the additive interference, and Z is the additive noise at the receiver. The encoder is assumed to *causally* know S. Using the terminology of Costa in [16], the earlier suggested strategy for the noiseless case in fact organizes *the dirt* such that the encoder can *write* on the remaining clean space. Additionally, this strategy is similar to the "interference concentration" scheme suggested by Willems in [17]. The suggested strategy can also be considered as a sort of *interference alignment* invented by Maddah-Ali *et al.* [18]. In the language of interference alignment, the aforementioned scheme operates in a way that the effective interference is aligned on a countable subset of the real line and the remaining space is reserved for the transmission of the desired signal.

VI. Concluding Remarks

We studied the relay channel with state and presented a lower bound on the achievable rates based on NNC. We showed that NNC and CF achieve the same rates. We then considered the state-decoupled relay channel, established the capacity of a semideterministic class, and demonstrated that capacity is achieved by NNC. By constructing some examples, we also showed that one can increase the capacity by causal relaying as compared to that with strictly causal relaying.

Motivated by the new examples of state-decoupled relay channels discussed in this paper, we next present a conjecture.

Conjecture 1: The capacity of the state-decoupled relay channel is

$$C = \max_{p(q)p(x_1|q)p(x_2|q)p(\hat{y}_2|y_2,x_2,q)} \min\{R_1, R_2\} \quad (28)$$

where

$$R_1 = I(X_1; Y_3, \hat{Y}_2|X_2, Q) \quad (29)$$
$$R_2 = I(X_1, X_2; Y_3|Q) - I(Y_2; \hat{Y}_2|X_1, X_2, Y_3, Q) \quad (30)$$

and $|\hat{\mathcal{Y}}_2| \le |\mathcal{X}_2| \cdot |\mathcal{Y}_2| + 1$, $|\mathcal{Q}| \le 2$.

Here, time sharing is used since the objective function is not convex in general [12]. Conjecture 1 includes that of Han-Ahlswede and Han in [19, Sec. V] as a special case. This follows by a similar discussion as that in Remark 5.

Related to the aforementioned conjecture, Tandon and Ulukus in [20] have established a new upper bound on the capacity of the state-decoupled relay channel with a noiseless link from the relay to the destination, which is tighter than the cutset bound. We also remark that the channel studied by Aleksic *et al.* [21] is state decoupled and its capacity is achieved by CF. For this channel, the upper bound in [20] is tight. This channel, however, does not fall in the semideterministic classes studied in [4]–[6] and Theorem 1 and the capacity is yet achieved by CF.

Appendix A
Proof of Proposition 1

In order to proceed with the proof, we first present two lemmas.

Lemma 1:

$$\begin{aligned}&I(\hat{Y}_2;Y_2,S_2|U_2,Y_3,S_3)-[I(U_2;Y_3,S_3)-I(U_2;S_2)]\\&=\Big[I(U_1;\hat{Y}_2,Y_3,S_3|U_2)-I(U_1;S_1|U_2)\Big]\\&\quad-\Big[I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)\\&\quad-I(U_1,U_2;S_1,S_2)\Big].\end{aligned}\tag{31}$$

Proof: Consider the following series of equalities:

$$\begin{aligned}&I(U_1;\hat{Y}_2,Y_3,S_3|U_2)-I(U_1;S_1|U_2)\\&=H(\hat{Y}_2,Y_3,S_3|U_2)-H(\hat{Y}_2,Y_3,S_3|U_1,U_2)-I(U_1;S_1|U_2)\\&=H(Y_3,S_3|U_2)+H(\hat{Y}_2|U_2,Y_3,S_3)-H(Y_3,S_3|U_1,U_2)\\&\quad-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)-I(U_1;S_1|U_2)\\&=I(U_1;Y_3,S_3|U_2)+H(\hat{Y}_2|U_2,Y_3,S_3)\\&\quad-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)-I(U_1;S_1|U_2)\\&=I(U_1,U_2;Y_3,S_3)-I(U_2;Y_3,S_3)+H(\hat{Y}_2|U_2,Y_3,S_3)\\&\quad-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)-I(U_1;S_1|U_2)\\&=I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_2,Y_3,S_3)\\&\quad+H(\hat{Y}_2|U_2,Y_3,S_3)-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)\\&\quad-I(U_2;S_2)-I(U_1;S_1|U_2)+G\\&=I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_2,Y_3,S_3)\\&\quad+I(\hat{Y}_2;U_1|U_2,Y_3,S_3)-I(U_1,U_2;S_1,S_2)+G\\&=I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;S_2|U_2,Y_3,S_3)\\&\quad+I(\hat{Y}_2;U_1|U_2,Y_3,S_3)-I(\hat{Y}_2;Y_2|S_2,U_2,Y_3,S_3)\\&\quad-I(U_1,U_2;S_1,S_2)+G\\&=I(U_1,U_2;Y_3,S_3)+H(\hat{Y}_2|S_2,U_2,Y_3,S_3)\\&\quad-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)-I(\hat{Y}_2;Y_2|S_2,U_2,Y_3,S_3)\\&\quad-I(U_1,U_2;S_1,S_2)+G\\&=I(U_1,U_2;Y_3,S_3)-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)\\&\quad+H(\hat{Y}_2|Y_2,S_2,U_2,Y_3,S_3)-I(U_1,U_2;S_1,S_2)+G\\&=I(U_1,U_2;Y_3,S_3)-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)\\&\quad+H(\hat{Y}_2|Y_2,S_2,U_1,U_2,Y_3,S_3)-I(U_1,U_2;S_1,S_2)+G\\&=I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)\\&\quad-I(U_1,U_2;S_1,S_2)+G\end{aligned}\tag{32}$$

where

$$\begin{aligned}G:=\,&I(\hat{Y}_2;Y_2,S_2|U_2,Y_3,S_3)-I(U_2;Y_3,S_3)\\&+I(U_2;S_2).\end{aligned}\tag{33}$$

This completes the proof of the lemma. ■

Lemma 2: Let

$$\mathcal{P}^*=p(x_1,u_1|s_1)p(x_2,u_2|s_2)p(\hat{y}_2|y_2,u_2,s_2)\tag{34}$$

be the joint pmf that optimizes the rate $R=\sup\min\{R_1,R_2\}$ in Proposition 1. Then, for $\mathcal{P}^*$

$$\begin{aligned}&I(U_1;\hat{Y}_2,Y_3,S_3|U_2)-I(U_1;S_1|U_2)\le I(U_1,U_2;Y_3,S_3)\\&-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)-I(U_1,U_2;S_1,S_2).\end{aligned}\tag{35}$$

Proof: We give the proof by contradiction. We show that if

$$\begin{aligned}&I(U_1;\hat{Y}_2,Y_3,S_3|U_2)-I(U_1;S_1|U_2)>I(U_1,U_2;Y_3,S_3)\\&-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)-I(U_1,U_2;S_1,S_2)\end{aligned}\tag{36}$$

then there is another

$$\mathcal{P}'=p(x_1,u_1|s_1)p(x_2,u_2|s_2)p(\hat{y}_2'|y_2,u_2,s_2)\tag{37}$$

that attains a higher rate. Now, let $\hat{Y}_2'=\hat{Y}_2$ with probability p and $\hat{Y}_2'=\emptyset$ otherwise. We observe that both terms under min in Proposition 1 are continuous in p and the first term increases in p, while the second term decreases in p. Thus, there exists a p^* such that

$$\begin{aligned}&I(U_1;\hat{Y}_2',Y_3,S_3|U_2)-I(U_1;S_1|U_2)=I(U_1,U_2;Y_3,S_3)\\&-I(\hat{Y}_2';Y_2,S_2|U_1,U_2,Y_3,S_3)-I(U_1,U_2;S_1,S_2)\end{aligned}$$

which contradicts (36). This completes the proof of the lemma. ■

Using Lemma 2, the rate $R=\sup\min\{R_1,R_2\}$ can be written as

$$\sup R_1,\ \text{ s.t. } R_1\le R_2\tag{38}$$

and using the identity proved in Lemma 1, the rate in Proposition 1 is equivalent to that in (1) and (2). This completes the proof.

APPENDIX B
ALTERNATIVE CODING SCHEME

We next provide an alternative transmission scheme for the achievable rate in Proposition 1. Our scheme is constructed using the NNC strategy combined with Gelfand–Pinsker multicoding. In this scheme, the source transmits a message $m\in[1:2^{nbR}]$ in b blocks; i.e., repetition coding. We use binning to utilize the knowledge of the channel states at the source and the relay. The binning rates at the source and the relay are denoted by $\tilde{R}_1$ and $\tilde{R}_2$, respectively. The relay employs a compression codebook with rate $\hat{R}_2$ to transmit a coded compression index denoted by $l\in[1:2^{n\hat{R}}]$ to the destination. The compression index is binned against the knowledge of the channel state at the relay prior to its transmission. The relay compresses both its received noisy signal and partial channel state information. After receiving the signals over b blocks, the destination performs a joint simultaneous nonunique decoding to form an estimate of the transmitted message $\hat{m}$. We use strong typicality as defined in [22] for encoding and decoding. For brevity, we use the notation $\boldsymbol{u}_{ij}=[u_{i,(j-1)n+1},u_{i,(j-1)n+2},\ldots,u_{i,jn}]$ where $u_{i,(j-1)n+k}$ denotes the signal generated, received, or transmitted at kth channel use in jth block at node $i\in\{1,2,3\}$ where $k\in[1:n]$ and $j\in[1:b]$.

1) Codebook Generation: Fix the pmf

$$p(u_1|s_1)p(x_1|u_1,s_1)p(u_2|s_2)p(x_2|u_2,s_2)p(\hat{y}_2|y_2,u_2,s_2).$$

TABLE II
ILLUSTRATION OF THE COMMUNICATION OVER b BLOCKS

Block	1	2	$\cdots$	$b-1$	b
U_1	$\boldsymbol{u}_{11}(m,\tilde{l}_{11})$	$\boldsymbol{u}_{12}(m,\tilde{l}_{12})$	$\cdots$	$\boldsymbol{u}_{1,b-1}(m,\tilde{l}_{1,b-1})$	$\boldsymbol{u}_{1b}(m,\tilde{l}_{1b})$
U_2	$\boldsymbol{u}_{21}(1,\tilde{l}_{21})$	$\boldsymbol{u}_{22}(l_{21},\tilde{l}_{22})$	$\cdots$	$\boldsymbol{u}_{2,b-1}(l_{2,b-2},\tilde{l}_{2,b-1})$	$\boldsymbol{u}_{2b}(l_{2,b-1},\tilde{l}_{2b})$
Y_2	$\hat{\boldsymbol{y}}_{21}(l_{21}\|1,\tilde{l}_{21})$	$\hat{\boldsymbol{y}}_{22}(l_{22}\|l_{21},\tilde{l}_{22})$	$\cdots$	$\hat{\boldsymbol{y}}_{2,b-1}(l_{2,b-1}\|l_{2,b-2},\tilde{l}_{2,b-1})$	$\hat{\boldsymbol{y}}_{2b}(l_{2b}\|l_{2,b-1},\tilde{l}_{2b})$
Y_3	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	$\hat{m}$

Then, for each block $j \in [1:b]$, randomly and independently generate $2^{nbR} \times 2^{n\tilde{R}_1}$ sequences

$$\boldsymbol{u}_{1j}(m,\tilde{l}_{1j}) \sim \prod_{i=1}^{n} p_{U_1}(u_{1,(j-1)n+i})$$
$$\forall m \in [1:2^{nbR}],\ \ \forall \tilde{l}_{1j} \in [1:2^{n\tilde{R}_1}]. \qquad (39)$$

Similarly, randomly and independently generate $2^{n\hat{R}_2} \times 2^{n\tilde{R}_2}$ sequences

$$\boldsymbol{u}_{2j}(l_{2,j-1},\tilde{l}_{2j}) \sim \prod_{i=1}^{n} p_{U_2}(u_{2,(j-1)n+i})$$
$$\forall l_{2,j-1} \in [1:2^{n\hat{R}_2}],\ \ \forall \tilde{l}_{2j} \in [1:2^{n\tilde{R}_2}]. \qquad (40)$$

For each $\boldsymbol{u}_{2j}(l_{2,j-1},\tilde{l}_{2j})$, $l_{2,j-1} \in [1:2^{n\hat{R}_2}]$, $\tilde{l}_{2j} \in [1:2^{n\tilde{R}_2}]$, randomly and conditionally independent generate $2^{n\hat{R}_2}$ sequences

$$\hat{\boldsymbol{y}}_2(l_{2j}|l_{2,j-1},\tilde{l}_{2j}) \sim \prod_{i=1}^{n} p_{\hat{Y}_2|U_2}(\hat{y}_{2,(j-1)n+i}|u_{2,(j-1)n+i}(l_{2,j-1},\tilde{l}_{2j})). \quad (41)$$

2) Encoding: We next explain the encoding at the beginning of block $j \in [1:b]$. Let $m \in [1:2^{nbR}]$ be the message to be sent. The source node looks for the smallest index $\tilde{l}_{1j} \in [1:2^{n\tilde{R}_1}]$ such that

$$\left(\boldsymbol{u}_{1j}(m,\tilde{l}_{1j}),\boldsymbol{s}_{1j}\right) \in \mathcal{T}_{\epsilon_1}^{(n)}. \qquad (42)$$

If there is no such index, it picks one at random.

At the end of block $j-1$, the relay node knows $\boldsymbol{s}_{2,j-1}$, $\boldsymbol{s}_{2j}$, $l_{2,j-2}$, and $\tilde{l}_{2,j-1}$. It then looks for the smallest index such that

$$\left(\hat{\boldsymbol{y}}_{2,j-1}(l_{2,j-1}|l_{2,j-2},\tilde{l}_{2,j-1})\ ,\boldsymbol{y}_{2,j-1},\boldsymbol{s}_{2,j-1},\ \boldsymbol{u}_{2,j-1}(l_{2,j-2},\tilde{l}_{2,j-1})\right) \in \mathcal{T}_{\epsilon_2}^{(n)} \qquad (43)$$

where $l_{2,0}=1$ by convention. If there is no such index, it picks one at random.

Similarly, the relay node then looks for the smallest index $\tilde{l}_{2j} \in [1:2^{n\tilde{R}_2}]$ such that

$$(\boldsymbol{u}_{2j}(l_{2,j-1},\tilde{l}_{2j}),\boldsymbol{s}_{2j}) \in \mathcal{T}_{\epsilon_3}^{(n)}. \qquad (44)$$

If there is no such index, it picks one at random.

Having found $\boldsymbol{u}_{1j}(m,\tilde{l}_{1j}) = [u_{1,(j-1)n+1},\ldots,u_{1,jn}]$ and $\boldsymbol{u}_{2j}(l_{2,j-1},\tilde{l}_{2j})$, the source transmits $\boldsymbol{x}_{1j}$ with i.i.d. components

$$X_{1,(n-1)j+i} \sim p_{X_1|U_1,S_1}(x_{1,(n-1)j+i}|u_{1,(j-1)n+i}(m,\tilde{l}_{1j}),s_{1,(j-1)n+i}) \quad (45)$$

and the relay transmits $\boldsymbol{x}_{2j}$ with i.i.d. components

$$X_{2,(j-1)n+i} \sim p_{X_2|U_2,S_2}(x_{2,(j-1)n+i}|u_{2,(j-1)n+i}(l_{2,j-1},\tilde{l}_{2j}),s_{2,(j-1)n+i}) \quad (46)$$

for $i \in [1:n]$.

3) Decoding: Let $\epsilon > \max\{\epsilon_1,\epsilon_2,\epsilon_3\}$. The destination performs the decoding at the end of block b. The decoder looks for a unique index $\hat{m} \in [1:2^{nbR}]$ such that

$$\left(\boldsymbol{u}_{1j}(\hat{m},\hat{\tilde{l}}_{1j})\ ,\hat{\boldsymbol{y}}_{2j}(\hat{l}_{2j}|\hat{l}_{2,j-1},\hat{\tilde{l}}_{2j}),\ \boldsymbol{u}_{2j}(\hat{l}_{2,j-1},\hat{\tilde{l}}_{2j}),\boldsymbol{y}_{3j},\boldsymbol{s}_{3j}\right) \in \mathcal{T}_{\epsilon}^{(n)} \qquad (47)$$

and for some $\hat{\tilde{l}}_{1j}$, $\hat{l}_{2j}$, $\hat{\tilde{l}}_{2,j-1}$, and for all $j \in [1:b]$. (Table II summarizes the encoding and decoding over b blocks.)

4) Probability of Error: Let $M=1$ denote the message sent from the source node and L_{2j} denote indices chosen at the relay in the block $j \in [1:b]$. Now, define the following events:

$$\mathcal{E}_1 := \bigcup_{j=1}^{b}\left\{\left(\boldsymbol{u}_{1j}(1,\tilde{l}_{1j}),\boldsymbol{s}_{1j}\right) \notin \mathcal{T}_{\epsilon_1}^{(n)}\right\} \qquad (48)$$

$$\mathcal{E}_2 := \bigcup_{j=1}^{b}\left\{\left(\boldsymbol{u}_{2j}(L_{2,j-1},\tilde{l}_{2j}),\boldsymbol{s}_{2j}\right) \notin \mathcal{T}_{\epsilon_2}^{(n)}\right\} \qquad (49)$$

$$\mathcal{E}_3 := \bigcup_{j=1}^{b}\left\{\left(\hat{\boldsymbol{y}}_{2j}(l_{2,j}|L_{2,j-1},\tilde{L}_{2j}),\ \boldsymbol{y}_{2,j},\boldsymbol{s}_{2,j},\boldsymbol{u}_{2,j}(L_{2,j-1},\tilde{L}_{2j})\right) \notin \mathcal{T}_{\epsilon_3}^{(n)}\right\} \qquad (50)$$

$$\mathcal{E}_{4m} := \left\{\left(\boldsymbol{u}_{1j}(m,\hat{\tilde{l}}_{1j}),\hat{\boldsymbol{y}}_{2j}(\hat{l}_{2j}|\hat{l}_{2,j-1},\hat{\tilde{l}}_{2j}),\boldsymbol{u}_{2j}(\hat{l}_{2,j-1},\hat{\tilde{l}}_{2j}),\ \boldsymbol{y}_{3j},\boldsymbol{s}_{3j}\right) \in \mathcal{T}_{\epsilon}^{(n)}\ \text{ for some } \hat{\tilde{l}}_{1j},\hat{\tilde{l}}_{2j},\hat{l}_{2j}\right\}. \qquad (51)$$

Then, the probability of error can be bounded as

$$\mathbb{P}_e^{(n)} \le \mathbb{P}^{(n)}(\mathcal{E}_1)+\mathbb{P}^{(n)}(\mathcal{E}_2)+\mathbb{P}^{(n)}(\mathcal{E}_3)+ \mathbb{P}^{(n)}(\mathcal{E}_{41}^c\cap\mathcal{E}_1^c\cap\mathcal{E}_2^c\cap\mathcal{E}_3^c)+ \mathbb{P}^{(n)}(\cup_{m\neq 1}\mathcal{E}_{4m}). \qquad (52)$$

Using the covering lemma [22], we have

$$\mathbb{P}^{(n)}(\mathcal{E}_1) \to 0 \text{ as } n\to\infty,\ \text{if } \tilde{R}_1 > I(U_1;S_1)+\delta(\epsilon_1) \qquad (53)$$
$$\mathbb{P}^{(n)}(\mathcal{E}_2) \to 0 \text{ as } n\to\infty,\ \text{if } \tilde{R}_2 > I(U_2;S_2)+\delta(\epsilon_2) \qquad (54)$$
$$\mathbb{P}^{(n)}(\mathcal{E}_3) \to 0 \text{ as } n\to\infty,\ \text{if } \hat{R}_2 > I(\hat{Y}_2;Y_2,S_2|U_2)+\delta(\epsilon_3). \qquad (55)$$

By the conditional typicality lemma [22]

$$\mathbb{P}^{(n)}(\mathcal{E}_{41}^c \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3^c) \to 0 \text{ as } n \to \infty. \tag{56}$$

We next bound $\mathbb{P}^{(n)}(\cup_{m\neq 1}\mathcal{E}_{4m})$. Define

$$\mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j}) := \Big\{\Big(\boldsymbol{U}_{1j}(m,\tilde{l}_{1j}), \hat{\boldsymbol{Y}}_{2j}(l_{2j}|l_{2,j-1},\tilde{l}_{2j}),\boldsymbol{U}_{2j}(l_{2,j-1},\tilde{l}_{2j}),\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j}\Big) \in \mathcal{T}_\epsilon^{(n)}\Big\}. \tag{57}$$

Then, consider

$$\begin{aligned}\mathbb{P}^{(n)}(\mathcal{E}_{4m}) &= \mathbb{P}^{(n)}\Big(\cup_{\tilde{l}_1^b,l_2^b,\tilde{l}_2^b}\cap_{j=1}^b \mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j})\Big)\\ &\le \sum_{\tilde{l}_1^b,\tilde{l}_2^b}\sum_{l_2^b}\mathbb{P}^{(n)}\Big(\cap_{j=1}^b\mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j})\Big)\\ &= \sum_{\tilde{l}_1^b,\tilde{l}_2^b}\sum_{l_2^b}\prod_{j=1}^b\mathbb{P}^{(n)}\Big(\mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j})\Big)\\ &\le \sum_{\tilde{l}_1^b,\tilde{l}_2^b}\sum_{l_2^b}\prod_{j=2}^b\mathbb{P}^{(n)}\Big(\mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j})\Big). \end{aligned}\tag{58}$$

If $l_{2,j-1} = L_{2,j-1}$, $\tilde{l}_{2j} = \tilde{L}_{2j}$ but $m \neq 1$, then $\boldsymbol{U}_{lj}(m,\tilde{l}_{1j})$ is independent of $\Big(\hat{\boldsymbol{Y}}_{2j}(l_{2j}|l_{2,j-1},\tilde{l}_{2j}),\boldsymbol{U}_{2j}(l_{2,j-1},\tilde{l}_{2j}),\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j}\Big)$. Therefore, by the joint typicality lemma [22], we have

$$\begin{aligned}&\mathbb{P}^{(n)}\Big(\mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j})\,|\\ &\quad l_{2,j-1}=L_{2,j-1},\tilde{l}_{2j}=\tilde{L}_{2j},m\neq 1\Big)\\ &= \sum_{(\boldsymbol{U}_{1j},\hat{\boldsymbol{Y}}_{2j},\boldsymbol{U}_{2j},\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j})\in\mathcal{T}_\epsilon^{(n)}} p(\boldsymbol{U}_{1j})p(\hat{\boldsymbol{Y}}_{2j},\boldsymbol{U}_{2j},\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j})\\ &\le \sum_{(\boldsymbol{U}_{1j},\hat{\boldsymbol{Y}}_{2j},\boldsymbol{U}_{2j},\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j})\in\mathcal{T}_\epsilon^{(n)}} 2^{-n(1-\epsilon)H(U_1)}2^{-n(1-\epsilon)H(U_2,\hat{Y}_2,Y_3;S_3)}\\ &= \Big|\mathcal{T}_\epsilon^{(n)}\Big|\,2^{-n(1-\epsilon)H(U_1)}2^{-n(1-\epsilon)H(U_2,\hat{Y}_2,Y_3,S_3)}\\ &\le 2^{n(1-\epsilon)H(U_1,U_2,\hat{Y}_2,Y_3,S_3)}2^{-n(1-\epsilon)H(U_1)}\times\\ &\quad 2^{-n(1-\epsilon)H(U_2,\hat{Y}_2,Y_3,S_3)}\\ &= 2^{-n(I(U_1;U_2,\hat{Y}_2,Y_3,S_3)-\delta(\epsilon))}\\ &= 2^{-n(I(U_1;\hat{Y}_2,Y_3,S_3|U_2)+I(U_1;U_2)-\delta(\epsilon))}\\ &= 2^{-n(I_1-\delta(\epsilon))}\end{aligned}\tag{59}$$

where

$$I_1 := I(U_1;\hat{Y}_2,Y_3,S_3|U_2)+I(U_1;U_2). \tag{60}$$

Similarly

$$\begin{aligned}&\mathbb{P}^{(n)}\Big(\mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j})\,|\\ &\quad \{l_{2,j-1}\neq L_{2,j-1} \text{ or } \tilde{l}_{2j}\neq\tilde{L}_{2j}\},m\neq 1\Big)\\ &= \sum_{(\boldsymbol{U}_{1j},\hat{\boldsymbol{Y}}_{2j},\boldsymbol{U}_{2j},\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j})\in\mathcal{T}_\epsilon^{(n)}} p(\boldsymbol{U}_{1j})p(\hat{\boldsymbol{Y}}_{2j},\boldsymbol{U}_{2j})p(\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j})\\ &\le \sum_{(\boldsymbol{U}_{1j},\hat{\boldsymbol{Y}}_{2j},\boldsymbol{U}_{2j},\boldsymbol{Y}_{3j},\boldsymbol{S}_{3j})\in\mathcal{T}_\epsilon^{(n)}} 2^{-n(1-\epsilon)H(U_1)}2^{-n(1-\epsilon)H(U_2,\hat{Y}_2)}\\ &\quad\times 2^{-n(1-\epsilon)H(Y_3,S_3)}\\ &= \Big|\mathcal{T}_\epsilon^{(n)}\Big|\,2^{-n(1-\epsilon)H(U_1)}2^{-n(1-\epsilon)H(U_2,\hat{Y}_2)}2^{-n(1-\epsilon)H(Y_3,S_3)}\\ &\le 2^{n(1-\epsilon)H(U_1,U_2,\hat{Y}_2,Y_3,S_3)}2^{-n(1-\epsilon)H(U_1)}2^{-n(1-\epsilon)H(U_2,\hat{Y}_2)}\\ &\quad\times 2^{-n(1-\epsilon)H(Y_3,S_3)}\\ &= 2^{-n(I(U_1,U_2;Y_3,S_3)+I(\hat{Y}_2;U_1,Y_3,S_3|U_2)+I(U_1;U_2)-\delta(\epsilon))}\\ &=: 2^{-n(I_2-\delta(\epsilon))}\end{aligned}\tag{61}$$

where the last equality holds since

$$\begin{aligned}&H(U_1)+H(U_2,\hat{Y}_2)+H(Y_3,S_3)-\\ &H(U_1,U_2,\hat{Y}_2,Y_3,S_3)\\ &= H(U_1)+H(U_2)+H(\hat{Y}_2|U_2)+H(Y_3,S_3)\\ &\quad -H(U_1,U_2)-H(Y_3,S_3|U_1,U_2)\\ &\quad -H(\hat{Y}_2|U_1,U_2,Y_3,S_3)\\ &= [H(Y_3,S_3)-H(Y_3,S_3|U_1,U_2)]\\ &\quad +\Big[H(\hat{Y}_2|U_2)-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)\Big]\\ &\quad +[H(U_1)+H(U_1)-H(U_1;U_2)]\\ &= I(U_1,U_2;Y_3,S_3)+I(\hat{Y}_2;U_1,Y_3,S_3|U_2)+I(U_1;U_2)\\ &=: I_2.\end{aligned}\tag{62}$$

Thus

$$\begin{aligned}&\mathbb{P}^{(n)}(\mathcal{E}_{4m})\\ &\le \sum_{\tilde{l}_1^b}\sum_{l_2,\tilde{l}_2}\sum_{\tilde{l}_2^{b-1},l_2^{b-1}}\prod_{j=2}^b\mathbb{P}^{(n)}\Big(\mathcal{A}_j(m,\tilde{l}_{1j},l_{2,j-1},l_{2j},\tilde{l}_{2j})\Big)\\ &\le \sum_{\tilde{l}_1^b}\sum_{l_2,\tilde{l}_2}\sum_{k=0}^{b-1}\binom{b-1}{k}2^{-n(b-1-k)(\hat{R}_2+\tilde{R}_2)}\times\\ &\quad 2^{-n(kI_1+(b-1-k)I_2-(b-1)\delta(\epsilon))}\\ &\le \sum_{\tilde{l}_1^b}\sum_{l_2,\tilde{l}_2}\sum_{k=0}^{b-1}\binom{b-1}{k}2^{-n(b-1)(\min\{I_1,I_2-\hat{R}_2-\tilde{R}_2)\}-\delta(\epsilon))}\\ &= \sum_{\tilde{l}_1^b}2^{n(\hat{R}_2+\tilde{R}_2)}2^{b-1}2^{-n(b-1)(\min\{I_1,I_2-\hat{R}_2-\tilde{R}_2\}-\delta(\epsilon))}\\ &= 2^{nb\tilde{R}_1}2^{n(\hat{R}_2+\tilde{R}_2)}2^{b-1}2^{-n(b-1)(\min\{I_1,I_2-\hat{R}_2-\tilde{R}_2\}-\delta(\epsilon))}.\end{aligned}\tag{63}$$

Employing the union bound, we obtain

$$\mathbb{P}^{(n)}(\cup_{m\neq 1}\mathcal{E}_{4m}) \leq \sum_{m\neq 1}\mathbb{P}^{(n)}(\mathcal{E}_{4m}) \leq 2^{nb((R-\frac{b-1}{b}[\min\{I_1-\tilde{R}_1, I_2-\hat{R}_2-\tilde{R}_1-\tilde{R}_2\}-\delta(\epsilon)]+\frac{1}{b}(\hat{R}_2+\tilde{R}_2+\tilde{R}_1))}. \quad (64)$$

Thus, as $n\to\infty$, the probability of error goes to zero if

$$R < \frac{b-1}{b}\Big[\min\{I_1-\tilde{R}_1, I_2-\hat{R}_2-\tilde{R}_1-\tilde{R}_2\}-\delta(\epsilon)\Big] - \frac{1}{b}(\hat{R}_2+\tilde{R}_2+\tilde{R}_1).$$

We then simplify each term under the min. Consider

$$\begin{aligned} I_1-\tilde{R}_1 &= I(U_1;\hat{Y}_2,Y_3,S_3|U_2)+I(U_1;U_2) - I(U_1;S_1)-\delta(\epsilon_1)\\ &= I(U_1;\hat{Y}_2,Y_3,S_3|U_2)+H(U_1)-H(U_1|U_2) - H(U_1)+H(U_1|S_1)-\delta(\epsilon_1)\\ &= I(U_1;\hat{Y}_2,Y_3,S_3|U_2)-H(U_1|U_2)+ H(U_1|S_1,U_2)-\delta(\epsilon_1)\\ &= I(U_1;\hat{Y}_2,Y_3,S_3|U_2)-I(U_1;S_1|U_2)-\delta(\epsilon_1). \end{aligned} \quad (65)$$

Similarly, consider

$$\begin{aligned} &I_2-\hat{R}_2-\tilde{R}_1-\tilde{R}_2\\ &= I(U_1,U_2;Y_3,S_3)+I(\hat{Y}_2;U_1,Y_3,S_3|U_2)+I(U_1;U_2)\\ &\quad - I(\hat{Y}_2;Y_2,S_2|U_2)-I(U_1;S_1)-I(U_2;S_2)-\delta'(\epsilon)\\ &= I(U_1,U_2;Y_3,S_3)\\ &\quad + \Big[H(\hat{Y}_2|Y_2,S_2,U_2)-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)\Big]+\\ &\quad [-H(U_1|U_2)+H(U_1|S_1)-H(U_2)+H(U_2|S_2)]-\delta'(\epsilon)\\ &= I(U_1,U_2;Y_3,S_3)\\ &\quad + \Big[H(\hat{Y}_2|Y_2,S_2,U_2,U_1,Y_3,S_3)-H(\hat{Y}_2|U_1,U_2,Y_3,S_3)\Big]+\\ &\quad [-H(U_1|U_2)+H(U_1|S_1)-H(U_2)+H(U_2|S_2)]-\delta'(\epsilon)\\ &= I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)+\\ &\quad [-H(U_1,U_2)+H(U_1|S_1,S_2)+H(U_2|S_2,U_1,S_1)]-\delta'(\epsilon)\\ &= I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)\\ &\quad -[H(U_1,U_2)-H(U_1,U_2|S_1,S_2)]-\delta'(\epsilon)\\ &= I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)\\ &\quad -I(U_1,U_2;S_1,S_2)-\delta'(\epsilon) \end{aligned} \quad (66)$$

where $\delta'(\epsilon) := \delta(\epsilon_1)+\delta(\epsilon_2)+\delta(\epsilon_3)$.

Now, let $b\to\infty$ and $\{\epsilon,\epsilon_1,\epsilon_2,\epsilon_3\}\to 0$. Thus, the rate

$$\begin{aligned} R < \min\Big\{&I(U_1;\hat{Y}_2,Y_3,S_3|U_2)-I(U_1;S_1|U_2),\\ &I(U_1,U_2;Y_3,S_3)-I(\hat{Y}_2;Y_2,S_2|U_1,U_2,Y_3,S_3)\\ &-I(U_1,U_2;S_1,S_2)\Big\} \end{aligned} \quad (67)$$

is achievable. This completes the proof of the theorem.

ACKNOWLEDGMENT

This work was done in part while the first author was visiting the Information Systems Lab. at Stanford University. We also thank the anonymous reviewers for their constructive comments on an earlier version of the 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. 25, no. 5, pp. 572–584, Sep. 1979.

[3] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1815–1817, May 2005.

[4] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, p. 536, May 1982.

[5] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 591–595.

[6] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

[7] M. N. Khormuji and M. Skoglund, "On cooperative downlink transmission with frequency reuse," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun.–Jul. 28–3, 2009, pp. 849–853.

[8] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. V1andendorpe, "Cooperative relaying with state available non-causally at the relay," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2272–2298, May 2010.

[9] B. Akhbari, M. Mirmohseni, and M. R. Aref, "Compress-and-forward strategy for relay channel with causal and non-causal channel state information," *IET Commun.*, vol. 4, no. 10, pp. 1174–1186, Jul. 2010.

[10] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[11] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Inf. Control*, vol. 9, no. 1, pp. 19–31, 1980.

[12] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[13] M. N. Khormuji, A. Zaidi, and M. Skoglund, "Interference management using nonlinear relaying," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1924–1930, Jul. 2010.

[14] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally to one encoder," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 480–484.

[15] M. N. Khormuji and M. Skoglund, "On instantaneous relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3378–3394, Jul. 2010.

[16] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[17] F. M. J. Willems, "Signaling for the Gaussian channel with side information at the transmitter," presented at the IEEE Int. Symp. Inf. Theory, 2000.

[18] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[19] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 396–412, May 1983.

[20] R. Tandon and S. Ulukus, "A new upper bound on the capacity of a class of primitive relay channels," in *Proc. 46th Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, Sep. 2008, pp. 1562–1569.

[21] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 921–930, Mar. 2009.

[22] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

Majid Nasiri Khormuji (S'07–M'11) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran in 2004 and the M.Sc. in electrical engineering with a major in wireless systems and the Ph.D. degree in telecommunications, both from KTH-Royal Institute of Technology, Stockholm, Sweden in 2006 and 2011, respectively. He held a visiting position at Stanford University, Stanford, CA in 2011. He is currently a Postdoctoral Research Fellow at the School of Electrical Engineering and the ACCESS Linnaeus Center at KTH. His research interests include information theory, wireless communications, modulation and coding for cooperative communications and wireless sensor networks.

Abbas El Gamal (S'71–M'73–SM'83–F'00) received the B.Sc. (honors) degree in electrical engineering from Cairo University in 1972 and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1977 and 1978, respectively. From 1978 to 1980, he was an Assistant Professor in the Department of Electrical Engineering at the University of Southern California (USC). He has been on the Stanford faculty since 1981, where he is currently the Hitachi America Professor in the School of Engineering and Chair of the Department of Electrical Engineering. His research interests and contributions have spanned the areas of network information theory, wireless networks, imaging sensors and systems, and integrated circuits. He has authored or coauthored over 200 papers and 30 patents in these areas. He is coauthor of the book Network Information Theory (Cambridge Press 2011). He has won several honors and awards, including the 2012 Claude E. Shannon Award, the 2009 Padovani Lecture, and the 2004 Infocom best paper award. He served on the technical program committees of many international conferences, including ISIT, ISSCC, and the International Symposium on FPGAs. He has been on the Board of Governors of the IT Society and is currently its First Vice President. He has played key roles in several semiconductor, EDA, and biotechnology startup companies.

Mikael Skoglund (S'93–M'97–SM'04) received the Ph.D. degree in 1997 from Chalmers University of Technology, Sweden. In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed to the Chair in Communication Theory in 2003. At KTH, he heads the Communication Theory Division and he is the Assistant Dean for Electrical Engineering.

Dr. Skoglund has worked on problems in source-channel coding, coding and transmission for wireless communications, Shannon theory and statistical signal processing. He has authored and co-authored more than 300 scientific papers in these areas, and he holds six patents.

Dr. Skoglund has served on numerous technical program committees for IEEE sponsored conferences (including ISIT and ITW). During 2003–08 he was an associate editor with the IEEE TRANSACTIONS ON COMMUNICATIONS and during 2008–12 he was on the editorial board for IEEE TRANSACTIONS ON INFORMATION THEORY.