The Arbitrarily Varying Relay Channel

Uzi Pereg and Yossef Steinberg Department of Electrical Engineering
Technion, Haifa 32000, Israel.
Email: `uzipereg@campus.technion.ac.il, ysteinbe@ee.technion.ac.il`

Abstract

We study the arbitrarily varying relay channel, and establish the cutset bound and partial decode-forward bound on the random code capacity. We further determine the random code capacity for special cases. Then, we consider conditions under which the deterministic code capacity is determined as well.

Index Terms

Arbitrarily varying channel, relay channel, decode-forward, Markov block code, minimax theorem, deterministic code, random code, symmetrizability.

I. INTRODUCTION

The relay channel was first introduced by van der Meulen [13] to describe point to point communication with the help of a relay, which receives a noisy version of the transmitter signal and transmits a signal of its own to the destination receiver, in a strictly causal manner. The capacity of the relay channel is not known in general, however, Cover and El Gamal established the cutset upper bound, the decode-forward lower bound, and the partial decode-forward lower bound [6]. It was also shown in [6] that for the reversely degraded relay channel, direct transmission is capacity achieving. For the degraded relay channel, the decode-forward bound and the cutset bound coincide, thus characterizing the capacity for this model [6]. In general, the partial decode-forward lower bound is tighter than both direct transmission and decode-forward lower bounds. El Gamal and Zahedi [11] determined the capacity of the relay channel with orthogonal sender components, by showing that the partial decode-forward bound and cutset bound coincide.

In practice, the channel statistics are not necessarily known in exact, and they may even change over time. The arbitrarily varying channel (AVC) is an appropriate model to describe such a situation [4]. Considering the AVC without a relay, Blackwell *et al.* determined the random code capacity [4], *i.e.* the capacity achieved by stochastic-encoder stochastic-decoder coding schemes with common randomness. It was also demonstrated in [4] that the random code capacity is not necessarily achievable using deterministic codes. A well-known result by Ahlswede [1] is the dichotomy property presented by the AVC. Specifically, the deterministic code capacity either equals the random code capacity or else, it is zero. Subsequently, Ericson [10] and Csiszár and Narayan [8] established a simple single-letter condition, namely non-symmetrizability, which is both necessary and sufficient for the capacity to be positive.

In this work, we study the arbitrarily varying relay channel (AVRC), which combines the previous models, *i.e.* the relay channel and the AVC. In the analysis, we incorporate the block Markov coding schemes of [6] in Ahlswede's Robustification and Elimination Techniques [1, 2]. We establish the cutset upper bound and the full/partial decode-forward lower bound on the random code capacity of the AVRC. We determine the random code capacity for special cases of the degraded AVRC, the reversely degraded AVRC, and the AVRC with orthogonal sender components. Then, we give extended symmetrizability conditions under which the deterministic code capacity coincides with the random code capacity. We show by example that the deterministic code capacity can be strictly lower than the random code capacity of the AVRC. We also give generalized symmetrizability conditions under which the deterministic code capacity coincides with the random code capacity, and conditions under which it is zero.

II. DEFINITIONS

A. Notation

We use the following notation conventions throughout. Calligraphic letters $\mathcal{X}, \mathcal{S}, \mathcal{Y}, ...$ are used for finite sets. Lowercase letters $x, s, y, \ldots$ stand for constants and values of random variables, and uppercase letters $X, S, Y, \ldots$ stand for random variables. The distribution of a random variable X is specified by a probability mass function (pmf) $P_X(x) = p(x)$ over a finite set $\mathcal{X}$. The set of all pmfs over $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. We use $x^j = (x_1, x_2, \ldots, x_j)$ to denote a sequence of letters from $\mathcal{X}$. A random sequence X^n and its distribution $P_{X^n}(x^n) = p(x^n)$ are defined accordingly. For a pair of integers i and j, $1 \leq i \leq j$, we define the discrete interval $[i : j] = \{i, i+1, \ldots, j\}$. The notation $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is used when it is understood from the context that the length of the sequence is n, and the ℓ^2-norm of $\mathbf{x}$ is denoted by $\|\mathbf{x}\|$.

B. Channel Description

A state-dependent discrete memoryless relay channel $(\mathcal{X}, \mathcal{X}_1, \mathcal{S}, W_{Y,Y_1|X,X_1,S}, \mathcal{Y}, \mathcal{Y}_1)$ consists of five sets, $\mathcal{X}, \mathcal{X}_1, \mathcal{S}, \mathcal{Y}$ and $\mathcal{Y}_1$, and a collection of conditional pmfs $W_{Y,Y_1|X,X_1,S}$. The sets stand for the input alphabet, the relay transmission alphabet, the state alphabet, the output alphabet, and the relay input alphabet, respectively. The alphabets are assumed to be finite, unless explicitly said otherwise. The channel is memoryless without feedback, and therefore

$$W_{Y^n,Y_1^n|X^n,X_1^n,S^n}(y^n, y_1^n|x^n, x_1^n, s^n) = \prod_{i=1}^{n} W_{Y,Y_1|X,X_1,S}(y_i, y_{1,i}|x_i, x_{1,i}, s_i). \tag{1}$$

Communication over a relay channel is depicted in Figure 1. Following [14], a relay channel $W_{Y,Y_1|X,X_1,S}$ is called degraded if the channel can be expressed as

$$W_{Y,Y_1|X,X_1,S}(y, y_1|x, x_1, s) = W_{Y_1|X,X_1,S}(y_1|x, x_1, s)p(y|y_1, x_1, s), \tag{2}$$

and it is called reversely degraded if

$$W_{Y,Y_1|X,X_1,S}(y, y_1|x, x_1, s) = W_{Y|X,X_1,S}(y|x, x_1, s)p(y_1|y, x_1, s). \tag{3}$$

The *arbitrarily varying relay channel* (AVRC) is a discrete memoryless relay channel $(\mathcal{X}, \mathcal{X}_1, \mathcal{S}, W_{Y,Y_1|X,X_1,S}, \mathcal{Y}, \mathcal{Y}_1)$ with a state sequence of unknown distribution, not necessarily independent nor stationary. That is, $S^n \sim q(s^n)$ with an unknown joint pmf $q(s^n)$ over $\mathcal{S}^n$. In particular, $q(s^n)$ can give mass 1 to some state sequence s^n. We use the shorthand notation $\mathcal{L} = \{W_{Y,Y_1|X,X_1,S}\}$ for the AVRC, where the alphabets are understood from the context.

To analyze the AVRC, we consider the *compound relay channel*. Different models of compound relay channels have been considered in the literature [15, 3]. Here, we define the compound relay channel as a discrete memoryless relay channel $(\mathcal{X}, \mathcal{X}_1, \mathcal{S}, W_{Y,Y_1|X,X_1,S}, \mathcal{Y}, \mathcal{Y}_1)$ with a discrete memoryless state, where the state distribution $q(s)$ is not known in exact, but rather belongs to a family of distributions $\mathcal{Q}$, with $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$. That is, $S^n \sim \prod_{i=1}^{n} q(s_i)$, with an unknown pmf $q \in \mathcal{Q}$ over $\mathcal{S}$. We use the shorthand notation $\mathcal{L}^{\mathcal{Q}}$ for the compound relay channel, where the transition probability $W_{Y,Y_1|X,X_1,S}$ and the alphabets are understood from the context.

In the analysis, we also use the following model. Suppose that the user transmits $B > 0$ blocks of length n, and the jammer is entitled to use a different state distribution $q_b(s) \in \mathcal{Q}$ for every block $b \in [1 : B]$, while the encoder, relay and receiver are aware of this jamming scheme. In other words, every block is governed by a different memoryless state. We refer to this channel as the block-compound relay channel, denoted by $\mathcal{L}^{\mathcal{Q} \times B}$. Although this is a toy model, it is a useful tool for the analysis of the AVRC.
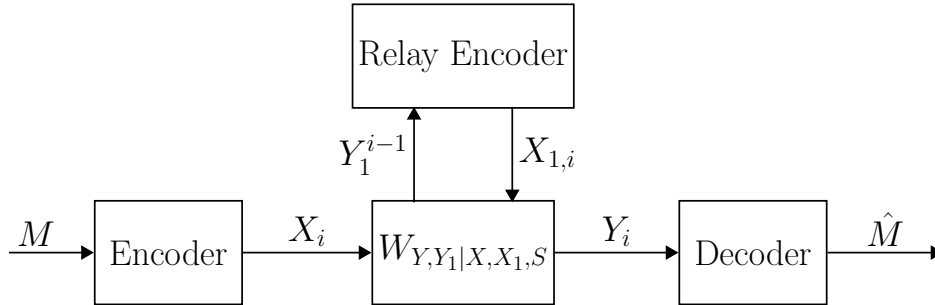

Fig. 1. Communication over the AVRC $\mathcal{L} = \{W_{Y,Y_1|X,X_1,S}\}$. Given a message M, the encoder transmits $X^n = f(M)$. At time $i \in [1 : n]$, the relay transmits $X_{1,i}$ based on all the symbols of the past Y_1^{i-1} and then receives a new symbol $Y_{1,i}$. The decoder receives the output sequence Y^n and finds an estimate of the message $\hat{M} = g(Y^n)$.

C. Coding

We introduce some preliminary definitions, starting with the definitions of a deterministic code and a random code for the AVRC $\mathcal{L}$. Note that in general, the term 'code', unless mentioned otherwise, refers to a deterministic code.

Definition 1 (A code, an achievable rate and capacity). A $(2^{nR}, n)$ code for the AVRC $\mathcal{L}$ consists of the following; a message set $[1 : 2^{nR}]$, where it is assumed throughout that 2^{nR} is an integer, an encoder $f : [1 : 2^{nR}] \to \mathcal{X}^n$, a sequence of n relaying functions $f_{1,i} : \mathcal{Y}_1^{i-1} \to \mathcal{X}_{1,i}$, $i \in [1 : n]$, and a decoding function $g : \mathcal{Y}^n \to [1 : 2^{nR}]$.

Given a message $m \in [1 : 2^{nR}]$, the encoder transmits $x^n = f(m)$. At time $i \in [1 : n]$, the relay transmits $x_{1,i} = f_{1,i}(y_1^{i-1})$ and then receives $y_{1,i}$. The relay codeword is given by $x_1^n = f_1^n(y_1^n) \triangleq (f_{1,i}(y_1^{i-1}))_{i=1}^{n}$. The decoder receives the output

sequence y^n and finds an estimate of the message $\hat{m} = g(y^n)$ (see Figure 1). We denote the code by $\mathscr{C} = (f(\cdot), f_1^n(\cdot), g(\cdot))$. Define the conditional probability of error of the code $\mathscr{C}$ given a state sequence $s^n \in \mathcal{S}^n$ by

$$P_{e|s^n}^{(n)}(\mathscr{C}) = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{(y^n, y_1^n) \,:\, g(y^n) \neq m} \left[\prod_{i=1}^n W_{Y,Y_1|X,X_1,S}(y_i, y_{1,i}|f_i(m), f_{1,i}(y_1^{i-1}), s_i) \right]. \tag{4}$$

Now, define the average probability of error of $\mathscr{C}$ for some distribution $q(s^n) \in \mathcal{P}(\mathcal{S}^n)$,

$$P_e^{(n)}(q, \mathscr{C}) = \sum_{s^n \in \mathcal{S}^n} q(s^n) \cdot P_{e|s^n}^{(n)}(\mathscr{C}). \tag{5}$$

Observe that $P_e^{(n)}(q, \mathscr{C})$ is linear in q, and thus continuous. We say that $\mathscr{C}$ is a $(2^{nR}, n, \varepsilon)$ code for the AVRC $\mathcal{L}$ if it further satisfies

$$P_e^{(n)}(q, \mathscr{C}) \leq \varepsilon, \quad \text{for all } q(s^n) \in \mathcal{P}(\mathcal{S}^n). \tag{6}$$

A rate R is called achievable if for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR}, n, \varepsilon)$ code. The operational capacity is defined as the supremum of the achievable rates and it is denoted by $\mathbb{C}(\mathcal{L})$. We use the term 'capacity' referring to this operational meaning, and in some places we call it the deterministic code capacity in order to emphasize that achievability is measured with respect to deterministic codes.

We proceed now to define the parallel quantities when using stochastic-encoders stochastic-decoder triplets with common randomness. The codes formed by these triplets are referred to as random codes.

Definition 2 (Random code). A $(2^{nR}, n)$ random code for the AVRC $\mathcal{L}$ consists of a collection of $(2^{nR}, n)$ codes $\{\mathscr{C}_\gamma = (f_\gamma, f_{1,\gamma}^n, g_\gamma)\}_{\gamma \in \Gamma}$, along with a probability distribution $\mu(\gamma)$ over the code collection Γ. We denote such a code by $\mathscr{C}^\Gamma = (\mu, \Gamma, \{\mathscr{C}_\gamma\}_{\gamma \in \Gamma})$. Analogously to the deterministic case, a $(2^{nR}, n, \varepsilon)$ random code has the additional requirement

$$P_e^{(n)}(q, \mathscr{C}^\Gamma) = \sum_{\gamma \in \Gamma} \mu(\gamma) P_e^{(n)}(q, \mathscr{C}_\gamma) \leq \varepsilon, \text{ for all } q(s^n) \in \mathcal{P}(\mathcal{S}^n). \tag{7}$$

The capacity achieved by random codes is denoted by $\mathbb{C}^\star(\mathcal{L})$, and it is referred to as the random code capacity.

III. Main Results – General AVRC

We present our results on the compound relay channel and the AVRC.

A. The Compound Relay Channel

We establish the cutset upper bound and the partial decode-forward lower bound for the compound relay channel. Consider a given compound relay channel $\mathcal{L}^\mathcal{Q}$. Let

$$\mathsf{R}_{CS}(\mathcal{L}^\mathcal{Q}) \triangleq \inf_{q \in \mathcal{Q}} \max_{p(x,x_1)} \min\left\{ I_q(X, X_1; Y), \ I_q(X; Y, Y_1|X_1) \right\}, \tag{8}$$

and

$$\mathsf{R}_{DF}(\mathcal{L}^\mathcal{Q}) \triangleq \max_{p(u,x,x_1)} \min\left\{ \inf_{q \in \mathcal{Q}} I_q(U, X_1; Y) + \inf_{q \in \mathcal{Q}} I_q(X; Y|X_1, U), \right.$$
$$\left. \inf_{q \in \mathcal{Q}} I_q(U; Y_1|X_1) + \inf_{q \in \mathcal{Q}} I_q(X; Y|X_1, U) \right\}, \tag{9}$$

where the subscripts 'CS' and 'DF' stand for 'cutset' and 'decode-forward', respectively.

Lemma 1. The capacity of the compound relay channel $\mathcal{L}^\mathcal{Q}$ is bounded by

$$\mathbb{C}(\mathcal{L}^\mathcal{Q}) \geq \mathsf{R}_{DF}(\mathcal{L}^\mathcal{Q}), \tag{10}$$
$$\mathbb{C}^\star(\mathcal{L}^\mathcal{Q}) \leq \mathsf{R}_{CS}(\mathcal{L}^\mathcal{Q}). \tag{11}$$

Specifically, if $R < \mathsf{R}_{DF}(\mathcal{L}^\mathcal{Q})$, then there exists a $(2^{nR}, n, e^{-an})$ block Markov code over $\mathcal{L}^\mathcal{Q}$ for sufficiently large n and some $a > 0$.

The proof of Lemma 1 is given in Appendix A. Observe that taking $U = \emptyset$ in (9) gives the direct transmission lower bound,

$$\mathbb{C}(\mathcal{L}^\mathcal{Q}) \geq \mathsf{R}_{DF}(\mathcal{L}^\mathcal{Q}) \geq \max_{p(x,x_1)} \inf_{q \in \mathcal{Q}} I_q(X; Y|X_1). \tag{12}$$

Taking $U = X$ in (9) results in a full decode-forward lower bound,

$$\mathbb{C}(\mathcal{L}^\mathcal{Q}) \geq \mathsf{R}_{DF}(\mathcal{L}^\mathcal{Q}) \geq \max_{p(x,x_1)} \inf_{q \in \mathcal{Q}} \min\left\{ I_q(X, X_1; Y), \ I_q(X; Y_1|X_1) \right\}. \tag{13}$$

This yields the following corollary.

Corollary 2. Let $\mathcal{L}^{\mathcal{Q}}$ be a compound relay channel, where $\mathcal{Q}$ is a compact convex set.

1) If $W_{Y,Y_1|X,X_1,S}$ is reversely degraded, such that

$$W_{Y,Y_1|X,X_1,S}(y,y_1|x,x_1,s) = W_{Y|X,X_1}(y|x,x_1)W_{Y_1|Y,X_1,S}(y_1|y,x_1,s), \tag{14}$$

then

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) = \min_{q \in \mathcal{Q}} \max_{p(x,x_1)} I_q(X;Y|X_1). \tag{15}$$

2) If $W_{Y,Y_1|X,X_1,S}$ is degraded, such that

$$W_{Y,Y_1|X,X_1,S}(y,y_1|x,x_1,s) = W_{Y_1|X,X_1}(y_1|x,x_1)W_{Y|Y_1,X_1,S}(y|y_1,x_1,s), \tag{16}$$

then

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) = \max_{p(x,x_1)} \min \left\{ \min_{q \in \mathcal{Q}} I_q(X,X_1;Y), \ I(X;Y_1|X_1) \right\}. \tag{17}$$

The proof of Corollary 2 is given in Appendix B. Part 1 follows from the direct transmission and cutset bounds, (12) and (8), respectively, while part 2 is based on the full decode-forward and cutset bounds, (13) and (8), respectively. The following corollary is a direct consequence of Lemma 1 and it is significant for the random code analysis of the AVRC.

Corollary 3. The capacity of the block-compound relay channel $\mathcal{L}^{\mathcal{Q} \times B}$ is bounded by

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q} \times B}) \geq \mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}}), \tag{18}$$

$$\mathbb{C}^{\star}(\mathcal{L}^{\mathcal{Q} \times B}) \leq \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}). \tag{19}$$

Specifically, if $R < \mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}})$, then there exists a $(2^{nR}, n, e^{-an})$ block Markov code over $\mathcal{L}^{\mathcal{Q} \times B}$ for sufficiently large n and some $a > 0$.

The proof of Corollary 3 is given in Appendix C.

B. The AVRC

We give lower and upper bounds, on the random code capacity and the deterministic code capacity, for the AVRC $\mathcal{L}$.

1) Random Code Lower and Upper Bounds: Define

$$\mathsf{R}_{DF}^{\star}(\mathcal{L}) \triangleq \mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}}) \Big|_{\mathcal{Q}=\mathcal{P}(\mathcal{S})}, \ \ \mathsf{R}_{CS}^{\star}(\mathcal{L}) \triangleq \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) \Big|_{\mathcal{Q}=\mathcal{P}(\mathcal{S})}. \tag{20}$$

Theorem 4. The random code capacity of an AVRC $\mathcal{L}$ is bounded by

$$\mathsf{R}_{DF}^{\star}(\mathcal{L}) \leq \mathbb{C}^{\star}(\mathcal{L}) \leq \mathsf{R}_{CS}^{\star}(\mathcal{L}). \tag{21}$$

The proof of Theorem 4 is given in Appendix D. Together with Corollary 2, this yields another corollary.

Corollary 5. Let $\mathcal{L}$ be an AVRC.

1) If $W_{Y,Y_1|X,X_1,S}$ is reversely degraded, such that $W_{Y,Y_1|X,X_1,S} = W_{Y|X,X_1}W_{Y_1|Y,X_1,S}$, then

$$\mathbb{C}^{\star}(\mathcal{L}) = \mathsf{R}_{DF}^{\star}(\mathcal{L}) = \mathsf{R}_{CS}^{\star}(\mathcal{L}) = \min_{q(s)} \max_{p(x,x_1)} I_q(X;Y|X_1). \tag{22}$$

2) If $W_{Y,Y_1|X,X_1,S}$ is degraded, such that $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X,X_1}W_{Y|Y_1,X_1,S}$, then

$$\mathbb{C}^{\star}(\mathcal{L}) = \mathsf{R}_{DF}^{\star}(\mathcal{L}) = \mathsf{R}_{CS}^{\star}(\mathcal{L}) = \max_{p(x,x_1)} \min \left\{ \min_{q(s)} I_q(X,X_1;Y), \ I(X;Y_1|X_1) \right\}. \tag{23}$$

Before we proceed to the deterministic code capacity, we note that Ahlswede's Elimination Technique [1] can be applied to the AVRC as well. Hence, the size of the code collection of any reliable random code can be reduced to polynomial size.

2) Deterministic Code Lower and Upper Bounds: In the next statements, we characterize the deterministic code capacity of the AVRC $\mathcal{L}$. We consider conditions under which the deterministic code capacity coincides with the random code capacity, and conditions under which it is lower. For every $x_1 \in \mathcal{X}_1$, let $\mathcal{W}_1(x_1)$ and $\mathcal{W}(x_1)$ denote the marginal AVCs from the sender to the relay and from the sender to the destination receiver, respectively,

$$\mathcal{W}_1(x_1) = \{W_{Y_1|X,X_1,S}(\cdot|\cdot, x_1, \cdot)\}, \ \mathcal{W}(x_1) = \{W_{Y|X,X_1,S}(\cdot|\cdot, x_1, \cdot)\}. \tag{24}$$

Lemma 6. If the marginal sender-relay and sender-reciever AVCs have positive capacities, *i.e.* $\mathbb{C}(\mathcal{W}_1(x_{1,1})) > 0$ and $\mathbb{C}(\mathcal{W}(x_{1,2})) > 0$, for some $x_{1,1}, x_{1,2} \in \mathcal{X}_1$, then the capacity of the AVRC $\mathcal{L}$ coincides with the random code capacity, *i.e.* $\mathbb{C}(\mathcal{L}) = \mathbb{C}^\star(\mathcal{L})$.

The proof of Lemma 6 is given in Appendix E. Next, we give a computable sufficient condition, under which the deterministic code capacity coincides with the random code capacity. For the point to point AVC, this occurs if and only if the channel is non-symmetrizable [10][8, Definition 2]. Our condition here is given in terms of an extended definition of symmetrizability, akin to [12, Definition 9].

Definition 3. A state-dependent relay channel $W_{Y,Y_1|X,X_1,S}$ is said to be *symmetrizable-$\mathcal{X}|\mathcal{X}_1$* if for some conditional distribution $J(s|x)$,

$$\sum_{s \in \mathcal{S}} W_{Y,Y_1|X,X_1,S}(y, y_1|x, x_1, s) J(s|\tilde{x}) = \sum_{s \in \mathcal{S}} W_{Y,Y_1|X,X_1,S}(y, y_1|\tilde{x}, x_1, s) J(s|x),$$

$$\forall x, \tilde{x} \in \mathcal{X}, \ x_1 \in \mathcal{X}_1, \ y \in \mathcal{Y}, \ y_1 \in \mathcal{Y}_1. \tag{25}$$

Equivalently, for every given $x_1 \in \mathcal{X}_1$, the DMC $W_{\bar{Y}|X,X_1,S}(\cdot|\cdot, x_1, \cdot)$ is symmetrizable, where $\bar{Y} = (Y, Y_1)$.

A similar definition applies to the marginals $W_{Y|X,X_1,S}$ and $W_{Y_1|X,X_1,S}$.

Corollary 7. Let $\mathcal{L}$ be an AVRC.

1) If $W_{Y|X,X_1,S}$ and $W_{Y_1|X,X_1,S}$ are non-symmetrizable-$\mathcal{X}|\mathcal{X}_1$, then $\mathbb{C}(\mathcal{L}) = \mathbb{C}^\star(\mathcal{L})$. In this case,

$$\mathsf{R}^\star_{DF}(\mathcal{L}) \leq \mathbb{C}(\mathcal{L}) \leq \mathsf{R}^\star_{CS}(\mathcal{L}). \tag{26}$$

2) If $W_{Y,Y_1|X,X_1,S}$ is reversely degraded, such that $W_{Y,Y_1|X,X_1,S} = W_{Y|X,X_1} W_{Y_1|Y,X_1,S}$, where $W_{Y_1|X,X_1,S}$ is non-symmetrizable-$\mathcal{X}|\mathcal{X}_1$ and $W_{Y|X,X_1}(y|x, x_1) \neq W_{Y|X,X_1}(y|\tilde{x}, x_1)$ for some $x, \tilde{x} \in \mathcal{X}$, $x_1 \in \mathcal{X}_1$ and $y \in \mathcal{Y}$, then

$$\mathbb{C}(\mathcal{L}) = \mathbb{C}^\star(\mathcal{L}) = \mathsf{R}^\star_{DF}(\mathcal{L}) = \mathsf{R}^\star_{CS}(\mathcal{L}) = \min_{q(s)} \max_{p(x,x_1)} I_q(X; Y|X_1). \tag{27}$$

3) If $W_{Y,Y_1|X,X_1,S}$ is degraded, such that $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X,X_1} W_{Y|Y_1,X_1,S}$, where $W_{Y|X,X_1,S}$ is non-symmetrizable-$\mathcal{X}|\mathcal{X}_1$ and $W_{Y_1|X,X_1}(y_1|x, x_1) \neq W_{Y_1|X,X_1}(y_1|\tilde{x}, x_1)$ for some $x, \tilde{x} \in \mathcal{X}$, $x_1 \in \mathcal{X}_1$ and $y_1 \in \mathcal{Y}_1$, then

$$\mathbb{C}(\mathcal{L}) = \mathbb{C}^\star(\mathcal{L}) = \mathsf{R}^\star_{DF}(\mathcal{L}) = \mathsf{R}^\star_{CS}(\mathcal{L}) = \max_{p(x,x_1)} \min \left\{ \min_{q(s)} I_q(X, X_1; Y), \ I(X; Y_1|X_1) \right\}. \tag{28}$$

The proof of Corollary 7 is given in Appendix F. Note that there are 4 symmetrizability cases in terms of the sender-relay channel $W_{Y_1|X,X_1,S}$ and the sender-receiver channel $W_{Y|X,X_1,S}$. For the case where $W_{Y_1|X,X_1,S}$ and $W_{Y|X,X_1,S}$ are both non-symmetrizable-$\mathcal{X}|\mathcal{X}_1$, the lemma above asserts that the capacity coincides with the random code capacity. In other cases, one may expect the capacity to be lower than the random code capacity. For instance, if $W_{Y|X,X_1,S}$ is non-symmetrizable-$\mathcal{X}|\mathcal{X}_1$, while $W_{Y_1|X,X_1,S}$ *is* symmetrizable-$\mathcal{X}|\mathcal{X}_1$, then the capacity is positive by direct transmission. Furthermore, in this case, if the channel is reversely degraded, then the capacity coincides with the random code capacity. However, it remains in question whether this is true in general, when the channel is not reversely degraded.

Next, we consider conditions under which the capacity is zero. Observe that if $W_{Y,Y_1|X,X_1,S}$ is symmetrizable-$\mathcal{X}|\mathcal{X}_1$ then so are $W_{Y|X,X_1,S}$ and $W_{Y_1|X,X_1,S}$. Intuitively, this means that the AVRC is a poor channel as well. For example, say $Y_1 = X + X_1 + S$ and $Y = X \cdot X_1 \cdot S$, then the jammer can confuse the decoder by taking the state sequence to be some codeword. The following lemma validates this intuition.

Lemma 8. If the AVRC $\mathcal{L}$ is symmetrizable-$\mathcal{X}|\mathcal{X}_1$, then it has zero capacity, *i.e.* $\mathbb{C}(\mathcal{L}) = 0$.

Lemma 8 is proved in Appendix G. If the AVRC is degraded then, we have a simpler symmetrizability condition under which the capacity is zero.

Definition 4. Let $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X,X_1} W_{Y|Y_1,X_1,S}$ be a degraded relay channel. We say that $W_{Y,Y_1|X,X_1,S}$ is *symmetrizable-$\mathcal{X}_1 \times \mathcal{Y}_1$* if for some conditional distribution $J(s|x_1, y_1)$,

$$\sum_{s \in \mathcal{S}} W_{Y|Y_1,X_1,S}(y|y_1, x_1, s) J(s|\tilde{x}_1, \tilde{y}_1) = \sum_{s \in \mathcal{S}} W_{Y|Y_1,X_1,S}(y|\tilde{y}_1, \tilde{x}_1, s) J(s|x_1, y_1),$$

$$\forall \tilde{x}_1, x_1 \in \mathcal{X}_1, \ y \in \mathcal{Y}, \ y_1, \tilde{y}_1 \in \mathcal{Y}_1. \tag{29}$$

Equivalently, the DMC $W_{Y|\bar{Y}_1,S}$ is symmetrizable, where $\bar{Y}_1 = (Y_1, X_1)$.

Lemma 9. If the AVRC $\mathcal{L}$ is degraded and symmetrizable-$\mathcal{X}_1 \times \mathcal{Y}_1$, then it has zero capacity, *i.e.* $\mathbb{C}(\mathcal{L}) = 0$.

Lemma 9 is proved in Appendix H. An example is given below.

Example 1. Consider a state-dependent relay channel $W_{Y,Y_1|X,X_1,S}$, specified by

$$Y_1 = X + Z \quad \text{mod } 2,$$
$$Y = X_1 + S,$$

where $\mathcal{X} = \mathcal{X}_1 = \mathcal{Z} = \mathcal{S} = \mathcal{Y}_1 = \{0,1\}$ and $\mathcal{Y} = \{0,1,2\}$, and the additive noise is distributed according to $Z \sim \text{Bernoulli}(\theta)$, $0 \le \theta \le 1$. It is readily seen that $W_{Y,Y_1|X,X_1,S}$ is degraded and symmetrizable-$\mathcal{X}_1 \times \mathcal{Y}_1$, by (2) and (29). In particular, (29) is satisfied with $J(s|x_1,y_1) = 1$ for $s = x_1$, and $J(s|x_1,y_1) = 0$ otherwise. Hence, by Lemma 9, the capacity is $\mathbb{C}(\mathcal{L}) = 0$. On the other hand, we show that the random code capacity is given by $\mathbb{C}^{\star}(\mathcal{L}) = \min\left\{\frac{1}{2}, 1 - h(\theta)\right\}$, using Corollary 5. The derivation is given in Appendix I.

C. AVRC with Orthogonal Sender Components

Consider the special case of a relay channel $W_{Y,Y_1|X,X_1,S}$ with orthogonal sender components [11] [9, Section 16.6.2], where $X = (X', X'')$ and

$$W_{Y,Y_1|X',X'',X_1,S}(y,y_1|x',x'',x_1,s) = W_{Y|X',X_1,S}(y|x',x_1,s) \cdot W_{Y_1|X'',X_1,S}(y_1|x'',x_1,s). \tag{30}$$

Here, we address the case where the channel output depends on the state only through the relay, *i.e.* $W_{Y|X',X_1,S}(y|x',x_1,s) = W_{Y|X',X_1}(y|x',x_1)$.

Lemma 10. Let $\mathcal{L} = \{W_{Y|X',X_1} W_{Y_1|X'',X_1,S}\}$ be an AVRC with orthogonal sender components. The random code capacity of $\mathcal{L}$ is given by

$$\mathbb{C}^{\star}(\mathcal{L}) = \mathsf{R}^{\star}_{DF}(\mathcal{L}) = \mathsf{R}^{\star}_{CS}(\mathcal{L}) = \max_{p(x_1)p(x'|x_1)p(x''|x_1)} \min\left\{I(X',X_1;Y),\ \min_{q(s)} I_q(X'';Y_1|X_1) + I(X';Y|X_1)\right\}. \tag{31}$$

If $W_{Y_1|X'',X_1,S}$ is non-symmetrizable-$\mathcal{X}''|\mathcal{X}_1$, and $W_{Y|X',X_1}(y|x',x_1) \ne W_{Y|X',X_1}(y|\tilde{x}',x_1)$ for some $x_1 \in \mathcal{X}_1$, $x', \tilde{x}' \in \mathcal{X}'$, $y \in \mathcal{Y}$, then the deterministic code capacity is given by $\mathbb{C}(\mathcal{L}) = \mathsf{R}^{\star}_{DF}(\mathcal{L}) = \mathsf{R}^{\star}_{CS}(\mathcal{L})$.

The proof of Lemma 10 is given in Appendix J. To prove Lemma 10, we apply the methods of [11] to our results. Specifically, we use the partial decode-forward lower bound in Theorem 4, taking $U = X''$ (see (9) and (20)).

APPENDIX A
PROOF OF LEMMA 1

A. Partial Decode-Forward Lower Bound

We construct a block Markov code, where the backward decoder uses joint typicality with respect to a state type, which is "close" to some $q \in \mathcal{Q}$. Let $\delta > 0$ be arbitrarily small. Define a set of state types $\hat{\mathcal{Q}}_n$ by

$$\hat{\mathcal{Q}}_n = \{\hat{P}_{s^n} : s^n \in \mathcal{A}^{\delta_1}(q) \text{ for some } q \in \mathcal{Q}\}, \tag{32}$$

where

$$\delta_1 \triangleq \frac{\delta}{2 \cdot |\mathcal{S}|}. \tag{33}$$

Namely, $\hat{\mathcal{Q}}_n$ is the set of types that are δ_1-close to some state distribution $q(s)$ in $\mathcal{Q}$. A code $\mathscr{C}$ for the compound relay channel is constructed as follows.

The encoders use B blocks, each consists of n channel uses to convey $(B-1)$ independent messages to the receiver. Furthermore, each message M_b, for $b \in [1 : B-1]$, is divided into two independent messages. That is, $M_b = (M'_b, M''_b)$, where M'_b and M''_b are uniformly distributed, *i.e.*

$$M'_b \sim \text{Unif}[1 : 2^{nR'}],\ M''_b \sim \text{Unif}[1 : 2^{nR''}],\ \text{with } R' + R'' = R, \tag{34}$$

for $b \in [1 : B-1]$. For convenience of notation, set $M'_0 = M'_B \equiv 1$ and $M''_0 = M''_B \equiv 1$. The average rate $\frac{B-1}{B} \cdot R$ is arbitrarily close to R.

Codebook Generation: Fix the distribution $P_{U,X,X_1}(u,x,x_1)$, and let

$$P^q_{X,Y,Y_1|U,X_1}(x,y,y_1|u,x_1) = P_{X|U,X_1}(x|u,x_1) \sum_{s \in \mathcal{S}} q(s) W_{Y,Y_1|X,X_1,S}(y,y_1|x,x_1,s). \tag{35}$$

Block	1	2	$\cdots$	$B-1$	B
Encoder	$x_1^n(m_1', m_1''\vert 1)$	$x_2^n(m_2', m_2''\vert m_1')$	$\cdots$	$x_{B-1}^n(m_{B-1}', m_{B-1}''\vert m_{B-2}')$	$x_B^n(1,1\vert m_{B-1}')$
Relay Decoder	$\widetilde{m}_1' \to$	$\widetilde{m}_2' \to$	$\cdots$	$\widetilde{m}_{B-1}'$	$\emptyset$
Relay Encoder	$x_{1,1}^n(1)$	$x_{1,2}^n(\widetilde{m}_1')$	$\cdots$	$x_{1,B-1}^n(\widetilde{m}_{B-2}')$	$x_{1,B}^n(m_{B-1}')$
Output	$\emptyset$	$\hat{m}_1'$	$\cdots$	$\leftarrow \hat{m}_{B-2}'$	$\leftarrow \hat{m}_{B-1}'$
	$\hat{m}_1''$	$\hat{m}_2''$	$\cdots$	$\hat{m}_{B-1}''$	$\emptyset$

Fig. 2. Partial decode-forward coding scheme. The block index $b \in [1:B]$ is indicated at the top. In the following rows, we have the corresponding elements: (1) sequences transmitted by the encoder; (2) estimated messages at the relay; (3) sequences transmitted by the relay; (4) estimated messages at the destination decoder. The arrows in the second row indicate that the relay encodes forwards with respect to the block index, while the arrows in the fourth row indicate that the receiver decodes backwards.

We construct B independent codebooks. For $b \in [2:B-1]$, generate $2^{nR'}$ independent sequences $x_{1,b}^n(m_{b-1}')$, $m_{b-1}' \in [1:2^{nR'}]$, at random, each according to $\prod_{i=1}^n P_{X_1}(x_{1,i})$. Then, generate $2^{nR'}$ sequences,

$$u_b^n(m_b'\vert m_{b-1}') \sim \prod_{i=1}^n P_{U\vert X_1}(u_i\vert x_{1,b,i}(m_{b-1}')), \ m_b' \in [1:2^{nR'}], \tag{36}$$

conditionally independent given $x_{1,b}^n(m_{b-1}')$. Then, for every $m_b' \in [1:2^{nR'}]$, generate $2^{nR''}$ sequences,

$$x_b^n(m_b', m_b''\vert m_{b-1}') \sim \prod_{i=1}^n P_{X\vert U,X_1}(x_i\vert u_{b,i}(m_b'\vert m_{b-1}'), x_{1,b,i}(m_{b-1}')), \ m_b'' \in [1:2^{nR''}], \tag{37}$$

conditionally independent given $(u_b^n(m_b'\vert m_{b-1}'), x_{1,b}^n(m_{b-1}'))$. We have thus generated $B-2$ independent codebooks,

$$\mathcal{F}_b = \left\{ \left(x_{1,b}^n(m_{b-1}'), u_b^n(m_b'\vert m_{b-1}'), x_b^n(m_b', m_b''\vert m_{b-1}')\right) : m_{b-1}', m_b' \in [1:2^{nR'}], \ m_b'' \in [1:2^{nR''}]\right\}, \tag{38}$$

for $b \in [2:B-1]$. The codebooks $\mathcal{F}_1$ and $\mathcal{F}_B$ are generated in the same manner, with fixed $m_0' = m_B' \equiv 1$ and $m_0'' = m_B'' \equiv 1$. Encoding and decoding is illustrated in Figure 2.

Encoding: To send the message sequence $(m_1', m_1'', \ldots, m_{B-1}', m_{B-1}'')$, transmit $x_b^n(m_b', m_b''\vert m_{b-1}')$ at block b, for $b \in [1:B]$.

Relay Encoding: In block 1, the relay transmits $x_{1,1}^n(1)$. Set $\widetilde{m}_0' \equiv 1$. At the end of block $b \in [1:B-1]$, the relay receives $y_{1,b}^n$, and finds some $\widetilde{m}_b' \in [1:2^{nR'}]$ such that

$$(u_b^n(\widetilde{m}_b'\vert \widetilde{m}_{b-1}'), x_{1,b}^n(\widetilde{m}_{b-1}'), y_{1,b}^n) \in \mathcal{A}^\delta(P_{U,X_1}P_{Y_1\vert U,X_1}^q), \text{ for some } q \in \hat{\mathcal{Q}}_n. \tag{39}$$

If there is none or there is more than one such, set $\widetilde{m}_b' = 1$. In block $b+1$, the relay transmits $x_{1,b+1}^n(\widetilde{m}_b')$.

Backward Decoding: Once all blocks $(y_b^n)_{b=1}^B$ are received, decoding is performed backwards. Set $\hat{m}_B' = \hat{m}_B'' \equiv 1$. For $b = B-1, B-2, \ldots, 1$, find a unique $\hat{m}_b' \in [1:2^{nR'}]$ such that

$$(u_{b+1}^n(\hat{m}_{b+1}'\vert \hat{m}_b'), x_{1,b+1}^n(\hat{m}_b'), y_{b+1}^n) \in \mathcal{A}^\delta(P_{U,X_1}P_{Y\vert U,X_1}^q), \text{ for some } q \in \hat{\mathcal{Q}}_n. \tag{40}$$

If there is none, or more than one such $\hat{m}_b' \in [1:2^{nR'}]$, declare an error.

Then, the decoder uses $\hat{m}_1', \ldots, \hat{m}_{B-1}'$ as follows. For $b = B-1, B-2, \ldots, 1$, find a unique $\hat{m}_b'' \in [1:2^{nR''}]$ such that

$$(u_b^n(\hat{m}_b'\vert \hat{m}_{b-1}'), x_b^n(\hat{m}_b', \hat{m}_b''\vert \hat{m}_{b-1}'), x_{1,b}(\hat{m}_{b-1}'), y_b^n) \in \mathcal{A}^\delta(P_{U,X,X_1}P_{Y\vert X,X_1}^q), \text{ for some } q \in \hat{\mathcal{Q}}_n. \tag{41}$$

If there is none, or more than one such $\hat{m}_b'' \in [1:2^{nR''}]$, declare an error. We note that using the set of types $\hat{\mathcal{Q}}_n$ instead of the original set of state distributions $\mathcal{Q}$ alleviates the analysis, since $\mathcal{Q}$ is not necessarily finite nor countable.

Analysis of Probability of Error: Assume without loss of generality that the user sent $(M_b', M_b'') = (1,1)$, and let $q^*(s) \in \mathcal{Q}$ denote the *actual* state distribution chosen by the jammer. The error event is bounded by the union of the events

$$\mathcal{E}_1(b) = \{\widetilde{M}_b' \neq 1\}, \ \mathcal{E}_2(b) = \{\hat{M}_b' \neq 1\}, \ \mathcal{E}_3(b) = \{\hat{M}_b'' \neq 1\}, \text{ for } b \in [1:B-1]. \tag{42}$$

Then, the probability of error is bounded by

$$P_e^{(n)}(q, \mathscr{C}) \le \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_1(b)\right) + \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_2(b) \mid \mathcal{E}_1^c(b)\right) + \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_3(b) \mid \mathcal{E}_1^c(b) \cap \mathcal{E}_2^c(b) \cap \mathcal{E}_2^c(b-1)\right), \tag{43}$$

with $\mathcal{E}_2(0) = \emptyset$, where the conditioning on $(M'_b, M''_b) = (1,1)$ is omitted for convenience of notation.

We begin with the probability of erroneous relaying, $\Pr\left(\mathcal{E}_1(b)\right)$. Define

$$\mathcal{E}_{1,1}(b) = \{(U_b^n(1|\widetilde{M}'_{b-1}), X_{1,b}^n(\widetilde{M}'_{b-1}), Y_{1,b}^n) \notin \mathcal{A}^\delta(P_{U,X_1} P_{Y_1|U,X_1}^{q'}) \text{ for all } q' \in \hat{\mathcal{Q}}_n\}$$

$$\mathcal{E}_{1,2}(b) = \{(U_b^n(m'_b|\widetilde{M}'_{b-1}), X_{1,b}^n(\widetilde{M}'_{b-1}), Y_{1,b}^n) \in \mathcal{A}^\delta(P_{U,X_1} P_{Y_1|U,X_1}^{q'}), \text{ for some } m'_b \neq 1, q' \in \hat{\mathcal{Q}}_n\}. \tag{44}$$

For $b \in [1 : B-1]$, the relay error event is bounded as

$$\mathcal{E}_1(b) \subseteq \mathcal{E}_1(b-1) \cup \mathcal{E}_{1,1}(b) \cup \mathcal{E}_{1,2}(b)$$
$$= \mathcal{E}_1(b-1) \cup (\mathcal{E}_1(b-1)^c \cap \mathcal{E}_{1,1}(b)) \cup (\mathcal{E}_1(b-1)^c \cap \mathcal{E}_{1,2}(b)), \tag{45}$$

with $\mathcal{E}_1(0) = \emptyset$. Thus, by the union of events bound,

$$\Pr\left(\mathcal{E}_1(b)\right) \le \Pr\left(\mathcal{E}_1(b-1)\right) + \Pr\left(\mathcal{E}_{1,1}(b) \mid \mathcal{E}_1(b-1)^c\right) + \Pr\left(\mathcal{E}_{1,2}(b) \mid \mathcal{E}_1(b-1)^c\right). \tag{46}$$

Consider the second term on the RHS of (46). We now claim that given that $\mathcal{E}_1(b-1)^c$ occurred, *i.e.* $\widetilde{M}'_{b-1} = 1$, the event $\mathcal{E}_{1,1}(b)$ implies that $(U_b^n(1|1), X_{1,b}^n(1), Y_{1,b}^n) \notin \mathcal{A}^{\delta/2}(P_{U,X_1} P_{Y_1|U,X_1}^{q''})$ for all $q'' \in \mathcal{Q}$. This claim is due to the following. Assume to the contrary that $\mathcal{E}_{1,1}(b)$ holds, but $(U_b^n(1|1), X_{1,b}^n(1), Y_{1,b}^n) \in \mathcal{A}^{\delta/2}(P_{U,X_1} P_{Y_1|U,X_1}^{q''})$ for some $q'' \in \mathcal{Q}$. Then, for a sufficiently large n, there exists a type $q'(s)$ such that

$$|q'(s) - q''(s)| \le \delta_1, \tag{47}$$

for all $s \in \mathcal{S}$, and by the definition in (32), $q' \in \hat{\mathcal{Q}}_n$. Then, (47) implies that

$$|P_{Y_1|U,X_1}^{q'}(y_1|u, x_1) - P_{Y_1|U,X_1}^{q''}(y_1|u, x_1)| \le |\mathcal{S}| \cdot \delta_1 = \frac{\delta}{2}, \tag{48}$$

for all $u \in \mathcal{U}$, $x_1 \in \mathcal{X}_1$ and $y_1 \in \mathcal{Y}_1$ (see (35) and (33)). Hence, $(U_b^n(1|1), X_{1,b}^n(1), Y_{1,b}^n) \in \mathcal{A}^\delta(P_{U,X_1} P_{Y_1|U,X_1}^{q'})$, which contradicts the first assumption. It follows that

$$\Pr\left(\mathcal{E}_{1,1}(b) \mid \mathcal{E}_1(b-1)^c\right) \le \Pr\left((U_b^n(1|1), X_{1,b}^n(1), Y_{1,b}^n) \notin \mathcal{A}^{\delta/2}(P_{U,X_1} P_{Y_1|U,X_1}^{q''}) \text{ for all } q'' \in \mathcal{Q} \mid \mathcal{E}_1(b-1)^c\right)$$
$$\le \Pr\left((U_b^n(1|1), X_{1,b}^n(1), Y_{1,b}^n) \notin \mathcal{A}^{\delta/2}(P_{U,X_1} P_{Y_1|U,X_1}^{q^*}) \mid \mathcal{E}_1(b-1)^c\right). \tag{49}$$

Since the codebooks $\mathcal{F}_1, \ldots, \mathcal{F}_B$ are independent, the sequence $(U_b^n(1|1), X_{1,b}^n(1))$ from the codebook $\mathcal{F}_b$ is independent of the relay estimate $\widetilde{M}_{b-1}$, which is a function of $Y_{1,b-1}^n$ and the codebook $\mathcal{F}_{b-1}$. Thus, the RHS of (49) tends to zero exponentially as $n \to \infty$ by the law of large numbers and Chernoff's bound.

We move to the third term in the RHS of (46). By the union of events bound, the fact that the number of type classes in $\mathcal{S}^n$ is bounded by $(n+1)^{|\mathcal{S}|}$, and the independence of the codebooks, we have that

$$\Pr\left(\mathcal{E}_{1,2}(b) \mid \mathcal{E}_1(b-1)^c\right) \le (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \hat{\mathcal{Q}}_n} \Pr\left((U_b^n(m'_b|1), X_{1,b}^n(1), Y_{1,b}^n) \in \mathcal{A}^\delta(P_{U,X_1} P_{Y_1|U,X_1}^{q'}) \text{ for some } m'_b \neq 1\right)$$

$$\le (n+1)^{|\mathcal{S}|} \cdot 2^{nR'} \cdot \sup_{q' \in \hat{\mathcal{Q}}_n} \left[\sum_{u^n, x_1^n} P_{U^n, X_1^n}(u^n, x_1^n) \cdot \sum_{y_1^n : (u^n, x_1^n, y_1^n) \in \mathcal{A}^\delta(P_{U,X_1} P_{Y_1|U,X_1}^{q'})} P_{Y_1^n|X_1^n}^{q^*}(y_1^n|x_1^n) \right], \tag{50}$$

where the last line follows since $U_b^n(m'_b|1)$ is conditionally independent of $Y_{1,b}^n$ given $X_{1,b}^n(1)$, for every $m'_b \neq 1$. Let y_1^n satisfy $(u^n, x_1^n, y_1^n) \in \mathcal{A}^\delta(P_{U,X_1} P_{Y_1|U,X_1}^{q'})$. Then, $(x_1^n, y_1^n) \in \mathcal{A}^{\delta_2}(P_{X_1,Y_1}^{q'})$ with $\delta_2 \triangleq |\mathcal{U}| \cdot \delta$. By Lemmas 2.6 and 2.7 in [7],

$$P_{X_1^n, Y_1^n}^{q^*}(x_1^n, y_1^n) = 2^{-n\left(H(\hat{P}_{x_1^n, y_1^n}) + D(\hat{P}_{x_1^n, y_1^n} \| P_{X_1, Y_1}^{q^*})\right)} \le 2^{-nH(\hat{P}_{x_1^n, y_1^n})} \le 2^{-n\left(H_{q'}(X_1, Y_1) - \varepsilon_1(\delta)\right)},$$

hence,

$$P_{Y_1^n|X_1^n}^{q^*}(y_1^n|x_1^n) \le 2^{-n\left(H_{q'}(Y_1|X_1) - \varepsilon_2(\delta)\right)}, \tag{51}$$

where $\varepsilon_1(\delta), \varepsilon_2(\delta) \to 0$ as $\delta \to 0$. Therefore, by (50)−(51), along with [7, Lemma 2.13],

$$\Pr\left(\mathcal{E}_{1,2}(b) \mid \mathcal{E}_1(b-1)^c\right) \le (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \mathcal{Q}} 2^{-n[I_{q'}(U;Y_1|X_1) - R' - \varepsilon_3(\delta)]}, \tag{52}$$

with $\varepsilon_3(\delta) \to 0$ as $\delta \to 0$. Using induction, we have by (46) that $\Pr(\mathcal{E}_1(b))$ tends to zero exponentially as $n \to \infty$, for $b \in [1 : B - 1]$, provided that $R' < \inf_{q' \in \mathcal{Q}} I_{q'}(U; Y_1 | X_1) - \varepsilon_3(\delta)$.

As for the erroneous decoding of M_b' at the receiver, observe that given $\mathcal{E}_1(b)^c$, the relay sends $X_{1,b}^n(1)$ in block $b+1$, hence

$$(U_{b+1}^n(1|1), X_{b+1}^n(1, 1|1), X_{1,b+1}^n(1)) \sim P_{U,X,X_1}(u, x, x_1). \tag{53}$$

At the destination receiver, decoding is performed backwards, hence the error events have a different form compared to those of the relay (*cf.* (44) and the events below). Define the events,

$$\mathcal{E}_{2,1}(b) = \{(U_{b+1}^n(\hat{M}_{b+1}'|1), X_{1,b+1}^n(1), Y_{b+1}^n) \notin \mathcal{A}^\delta(P_{U,X_1} P_{Y|U,X_1}^{q'}) \text{ for all } q' \in \hat{\mathcal{Q}}_n\}$$

$$\mathcal{E}_{2,2}(b) = \{(U_{b+1}^n(\hat{M}_{b+1}'|m_b'), X_{1,b+1}^n(m_b'), Y_{b+1}^n) \in \mathcal{A}^\delta(P_{U,X_1} P_{Y_1|U,X_1}^{q'}), \text{ for some } m_b' \neq 1, q' \in \hat{\mathcal{Q}}_n\} \tag{54}$$

For $b \in [1 : B - 1]$, the error event $\mathcal{E}_2(b)$ is bounded by

$$\begin{aligned}\mathcal{E}_2(b) \subseteq &\mathcal{E}_2(b+1) \cup \mathcal{E}_{2,1}(b) \cup \mathcal{E}_{2,2}(b)\\ = &\mathcal{E}_2(b+1) \cup (\mathcal{E}_2(b+1)^c \cap \mathcal{E}_{2,1}(b)) \cup (\mathcal{E}_2(b+1)^c \cap \mathcal{E}_{2,2}(b)),\end{aligned} \tag{55}$$

with $\mathcal{E}_2(B) = \emptyset$. Thus,

$$\Pr(\mathcal{E}_2(b) \mid \mathcal{E}_1(b)^c) \leq \Pr(\mathcal{E}_2(b+1) \mid \mathcal{E}_1(b)^c) + \Pr(\mathcal{E}_{2,1}(b) \mid \mathcal{E}_1(b)^c, \mathcal{E}_2(b+1)^c) + \Pr(\mathcal{E}_{2,2}(b) \mid \mathcal{E}_1(b)^c, \mathcal{E}_2(b+1)^c). \tag{56}$$

By similar arguments to those used above, we have that

$$\Pr(\mathcal{E}_{2,1}(b) \mid \mathcal{E}_1(b)^c, \mathcal{E}_2(b+1)^c) \leq \Pr\left((U_{b+1}^n(1|1), X_{1,b+1}^n(1), Y_{b+1}^n) \notin \mathcal{A}^{\delta/2}(P_{U,X_1} P_{Y|U,X_1}^{q^*}) \mid \mathcal{E}_1(b)^c\right), \tag{57}$$

which tends to zero exponentially as $n \to \infty$, due to (53), and by the law of large numbers and Chernoff's bound. Then, by similar arguments to those used for the bound on $\Pr(\mathcal{E}_{1,2}(b) \mid \mathcal{E}_1(b-1)^c)$, the third term on the RHS of (56) tends to zero as $n \to \infty$, provided that $R' < \inf_{q' \in \mathcal{Q}} I_{q'}(U, X_1; Y) - \varepsilon_4(\delta)$, where $\varepsilon_4(\delta) \to 0$ as $\delta \to 0$. Using induction, we have by (56) that the second term on the RHS of (43) tends to zero exponentially as $n \to \infty$, for $b \in [1 : B - 1]$.

Moving to the error event for M_b'', define

$$\mathcal{E}_{3,1}(b) = \{(U_b^n(\hat{M}_b'|\hat{M}_{b-1}'), X_b^n(\hat{M}_b', 1|\hat{M}_{b-1}'), X_{1,b}(\hat{M}_{b-1}'), Y_b^n) \notin \mathcal{A}^\delta(P_{U,X,X_1} P_{Y|X,X_1}^{q'}), \text{ for all } q' \in \hat{\mathcal{Q}}_n\}$$

$$\mathcal{E}_{3,2}(b) = \{(U_b^n(\hat{M}_b'|\hat{M}_{b-1}'), X_b^n(\hat{M}_b', m_b''|\hat{M}_{b-1}'), X_{1,b}(\hat{M}_{b-1}'), Y_b^n) \in \mathcal{A}^\delta(P_{U,X,X_1} P_{Y|X,X_1}^{q'}), \text{ for some } m_b'' \neq 1, q' \in \hat{\mathcal{Q}}_n\}. \tag{58}$$

Given $\mathcal{E}_2(b)^c \cap \mathcal{E}_2(b-1)^c$, we have that $\hat{M}_b' = 1$ and $\hat{M}_{b-1}' = 1$. Then, by similar arguments to those used above,

$$\begin{aligned}&\Pr(\mathcal{E}_3(b) \mid \mathcal{E}_1(b)^c \cap \mathcal{E}_2(b)^c \cap \mathcal{E}_2(b-1)^c)\\ \leq &\Pr(\mathcal{E}_{3,1}(b) \mid \mathcal{E}_1(b)^c \cap \mathcal{E}_2(b)^c \cap \mathcal{E}_2(b-1)^c) + \Pr(\mathcal{E}_{3,2}(b) \mid \mathcal{E}_1(b)^c \cap \mathcal{E}_2(b)^c \cap \mathcal{E}_2(b-1)^c)\\ \leq &e^{-a_0 n} + (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \mathcal{Q}} \sum_{m_b'' \neq 1} \Pr\left((U_b^n(1|1), X_b^n(1, m_b''|1), X_{1,b}(1), Y_b^n) \in \mathcal{A}^\delta(P_{U,X,X_1} P_{Y|X,X_1}^{q'}) \mid \mathcal{E}_1(b)^c\right)\\ \leq &e^{-a_0 n} + (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \mathcal{Q}} 2^{-n[I_{q'}(X;Y|U,X_1) - R'' - \varepsilon_5(\delta)]}\end{aligned} \tag{59}$$

where $a_0 > 0$ and $\varepsilon_5(\delta) \to 0$ as $\delta \to 0$. The second inequality holds by (53) along with the law of large numbers and Chernoff's bound, and the last inequality holds as $X_b^n(1, m_b''|1)$ is conditionally independent of Y_b^n given $(U_b^n(1|1), X_{1,b}^n(1))$ for every $m_b'' \neq 1$. Thus, the third term on the RHS of (43) tends to zero exponentially as $n \to \infty$, provided that $R'' < \inf_{q' \in \mathcal{Q}} I_{q'}(X; Y|U, X_1) - \varepsilon_5(\delta)$. Eliminating R' and R'', we conclude that the probability of error, averaged over the class of the codebooks, exponentially decays to zero as $n \to \infty$, provided that $R < \mathsf{R}_{DF}(\mathcal{L}^\mathcal{Q})$. Therefore, there must exist a $(2^{nR}, n, \varepsilon)$ deterministic code, for a sufficiently large n. $\qquad\square$

B. Cutset Upper Bound

This is a straightforward consequence of the cutset bound in [6]. Assume to the contrary that there exists an achievable rate $R > \mathsf{R}_{CS}(\mathcal{L}^\mathcal{Q})$. Then, for some $q^*(s)$ in the closure of $\mathcal{Q}$,

$$R > \max_{p(x,x_1)} \min\{I_{q^*}(X, X_1; Y), I_{q^*}(X; Y, Y_1|X_1)\}. \tag{60}$$

By the achievability assumption, we have that for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR}, n)$ random code $\mathscr{C}^\Gamma$ such that $P_e^{(n)}(q, \mathscr{C}) \leq \varepsilon$ for every i.i.d. state distribution $q \in \mathcal{Q}$, and in particular for q^*. This holds even if q^* is in the closure of $\mathcal{Q}$ but not in $\mathcal{Q}$ itself, since $P_e^{(n)}(q, \mathscr{C})$ is continuous in q. Consider using this code over a standard relay channel $W_{Y,Y_1|X,X_1}$ without a state, where $W_{Y,Y_1|X,X_1}(y, y_1|x, x_1) = \sum_{s \in \mathcal{S}} q^*(s) W_{Y,Y_1|X,X_1,S}(y, y_1|x, x_1, s)$. It follows that the rate R as in (60) can be achieved over the relay channel $W_{Y,Y_1|X,X_1}$, in contradiction to [6]. We deduce that the assumption is false, and $R > \mathsf{R}_{CS}(\mathcal{L}^\mathcal{Q})$ cannot be achieved. $\qquad\square$

APPENDIX B
PROOF OF COROLLARY 2

This is a straightforward consequence of Lemma 1, which states that the capacity of the compound relay channel is bounded by $\mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}}) \leq \mathbb{C}(\mathcal{L}^{\mathcal{Q}}) \leq \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}})$. Thus, if $W_{Y,Y_1|X,X_1,S}$ is reversely degraded such that $W_{Y,Y_1|X,X_1,S} = W_{Y|X,X_1}W_{Y_1|Y,X_1,S}$, then $I_q(X;Y,Y_1|X_1) = I_q(X;Y|X_1)$, and the bounds coincide by the minimax theorem [16], $cf.$ (8) and (12). Similarly, if $W_{Y,Y_1|X,X_1,S}$ is degraded such that $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X,X_1}W_{Y|Y_1,X_1,S}$, then $I_q(X;Y,Y_1|X_1) = I(X;Y_1|X_1)$, and by (8) and (13),

$$\mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) = \min_{q(s)\in\mathcal{Q}} \max_{p(x,x_1)} \min\left\{I_q(X,X_1;Y),\, I(X;Y_1|X_1)\right\}\,, \tag{61}$$

$$\mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}}) = \max_{p(x,x_1)} \min_{q(s)\in\mathcal{Q}} \min\left\{I_q(X,X_1;Y),\, I(X;Y_1|X_1)\right\}\,. \tag{62}$$

Observe that $\min\left\{I_q(X,X_1;Y),\, I(X;Y_1|X_1)\right\}$ is concave in $p(x,x_1)$ and quasi-convex in $q(s)$ (see *e.g.* [5, Section 3.4]), hence the bounds (61) and (62) coincide by the minimax theorem [16]. $\qquad\square$

APPENDIX C
PROOF OF COROLLARY 3

Consider the block-compound relay channel $\mathcal{L}^{\mathcal{Q}\times B}$, where the state distribution $q_b \in \mathcal{Q}$ varies from block to block. Since the encoder, relay and receiver are aware of this jamming scheme, the capacity is the same as that of the ordinary compound channel, *i.e.* $\mathbb{C}(\mathcal{L}^{\mathcal{Q}\times B}) = \mathbb{C}(\mathcal{L}^{\mathcal{Q}})$ and $\mathbb{C}^\star(\mathcal{L}^{\mathcal{Q}\times B}) = \mathbb{C}^\star(\mathcal{L}^{\mathcal{Q}})$. Hence, (18) and (19) follow from Lemma 1. As for the second part of the corollary, observe that the block Markov coding scheme used in the proof of the decode forward lower bound can be applied as is to the block-compound relay channel, since the relay and the destination receiver do not estimate the state distribution while decoding the messages (see Appendix A). Furthermore, the analysis also holds, where the actual state distribution q^*, in (49)–(51) and (57), is now replaced by the state distribution q_b^* which corresponds to block $b \in [1:B]$. $\qquad\square$

APPENDIX D
PROOF OF THEOREM 4

First, we explain the general idea. We modify Ahlswede's Robustification Technique (RT) [2] to the relay channel. Namely, we use codes for the compound relay channel to construct a random code for the AVRC using randomized permutations. However, in our case, the strictly causal nature of the relay imposes a difficulty, and the application of the RT is not straightforward.

In [2], there is noncausal state information and a random code is defined via permutations of the codeword symbols and the received sequence. Here, however, the relay cannot apply permutations to its transmission x_1^n, because it depends on the received sequence y_1^n in a strictly causal manner. We resolve this difficulty using block Markov codes for the block-compound relay channel to construct a random code for the AVRC, applying B in-block permutations to the relay transmission, which depends only on the sequence received in the *previous block*. The details are given below.

A. Partial Decode Forward Lower Bound

We show that every rate $R < \mathsf{R}_{DF}^\star(\mathcal{L})$ (see (20)) can be achieved by random codes over the AVRC $\mathcal{L}$, *i.e.* $\mathbb{C}(\mathcal{L}) \geq \mathsf{R}_{DF}^\star(\mathcal{L})$. We start with Ahlswede's RT [2], stated below. Let $h : \mathcal{S}^n \to [0,1]$ be a given function. If, for some fixed $\alpha_n \in (0,1)$, and for all $q(s^n) = \prod_{i=1}^n q(s_i)$, with $q \in \mathcal{P}(\mathcal{S})$,

$$\sum_{s^n \in \mathcal{S}^n} q(s^n)h(s^n) \leq \alpha_n\,, \tag{63}$$

then,

$$\frac{1}{n!}\sum_{\pi\in\Pi_n} h(\pi s^n) \leq \beta_n\,, \quad \text{for all } s^n \in \mathcal{S}^n\,, \tag{64}$$

where Π_n is the set of all n-tuple permutations $\pi : \mathcal{S}^n \to \mathcal{S}^n$, and $\beta_n = (n+1)^{|\mathcal{S}|} \cdot \alpha_n$.

According to Corollary 3, for every $R < \mathsf{R}_{DF}^\star(\mathcal{L})$, there exists a $(2^{nR(B-1)}, nB, e^{-2\theta n})$ block Markov code for the block-compound relay channel $\mathcal{L}^{\mathcal{P}(\mathcal{S})\times B}$ for some $\theta > 0$ and sufficiently large n, where $B > 0$ is arbitrarily large. Recall that the code constructed in the proof in Appendix A has the following form. The encoders use $B > 0$ blocks to convey $B-1$ messages m_b, $b \in [1 : B-1]$. Each message consists of two parts, *i.e.* $m_b = (m_b', m_b'')$, where $m_b' \in [1 : 2^{nR'}]$ and $m_b'' \in [1 : 2^{nR''}]$. In block $b \in [1 : B]$, the encoder sends $x_b^n = f_b(m_b', m_b''|m_{b-1}')$, with fixed m_0 and m_B, and the relay transmits $x_{1,b}^n = f_{1,b}(y_{1,b-1}^n)$, using the sequence received in the previous block. After receiving the entire output sequence $(y_b^n)_{b=1}^B$, the decoder finds an estimate for the messages. Set $\hat{m}_B' = 1$. The first part of each message is decoded backwards as $\hat{m}_b' = g_b'(y_{b+1}^n, \hat{m}_{b+1}')$, for $b = B-1, B-2, \ldots, 1$. Then, the second part of each message is decoded as $\hat{m}_b'' = g_b''(y_b^n, \hat{m}_1', \ldots, \hat{m}_{B-1}')$, for $b \in [1 : B-1]$. The overall blocklength is then $n \cdot B$ and the average rate is $\frac{B-1}{B}(R' + R'')$.

Given such a block Markov code $\mathscr{C}_{BM}$ for the block-compound relay channel $\mathcal{L}^{\mathcal{P}(\mathcal{S})\times B}$, we have that

$$\mathrm{Pr}_{\mathscr{C}_{BM}}\big(\mathcal{E}_b' \mid (\mathcal{E}_{b+1}')^c\big) \le e^{-2\theta n}\,,\quad \mathrm{Pr}_{\mathscr{C}_{BM}}\big(\mathcal{E}_b'' \mid \mathcal{E}_1'^c,\dots,\mathcal{E}_{b-1}'^c\big) \le e^{-2\theta n} \tag{65}$$

for $b = B-1,\dots,1$, where $\mathcal{E}_0' = \mathcal{E}_B' = \emptyset$, and $\mathcal{E}_b' = \{\hat{M}_b' \ne M_b'\}$, $\mathcal{E}_b'' = \{\hat{M}_b'' \ne M_b''\}$, $b \in [1:B-1]$. That is, for every sequence of state distributions $q_1,\dots,q_{b+1}$, where $q_t(s_t^n) = \prod_{i=1}^n q_t(s_{t,i})$ for $t \in [1:b+1]$,

$$\sum_{s_1^n \in \mathcal{S}^n} q_1(s_1^n) \sum_{s_2^n \in \mathcal{S}^n} q_2(s_2^n)\,\cdots\,\sum_{s_{b+1}^n \in \mathcal{S}^n} q_{b+1}(s_{b+1}^n)\cdot h_b'(s_1^n,s_2^n,\dots,s_{b+1}^n) \le e^{-2\theta n}\,, \tag{66}$$

and

$$\sum_{s_1^n \in \mathcal{S}^n} q_1(s_1^n) \sum_{s_2^n \in \mathcal{S}^n} q_2(s_2^n)\,\cdots\,\sum_{s_b^n \in \mathcal{S}^n} q_b(s_b^n)\cdot h_b''(s_1^n,s_2^n,\dots,s_b^n) \le e^{-2\theta n}\,, \tag{67}$$

where

$$h_b'(s_1^n,s_2^n,\dots,s_{b+1}^n) = \frac{1}{2^{n(b+1)(R'+R'')}}\sum_{(m_1',m_1''),\dots,(m_{b+1}',m_{b+1}'')}$$

$$\sum_{y_{1,b}^n \in \mathcal{Y}_1^n} \mathrm{Pr}\left(Y_{1,b}^n = y_{1,b}^n \mid (M_1',M_1'') = (m_1',m_1''),\dots,(M_b',M_b'') = (m_b',m_b''), S_1^n = s_1^n,\dots,S_b^n = s_b^n\right)$$

$$\times \sum_{y_{b+1}^n:g_b'(y_{b+1}^n,m_{b+1}')\ne m_b'} W_{Y^n|X^n,X_1^n,S^n}(y_{b+1}^n|f_{b+1}(m_{b+1}',m_{b+1}''|m_b'),f_{1,b+1}(y_{1,b}^n),s_{b+1}^n) \tag{68}$$

and

$$h_b''(s_1^n,s_2^n,\dots,s_b^n) = \frac{1}{2^{nR''}}\sum_{m_b''=1}^{2^{nR''}}\frac{1}{2^{nR'(B-1)}}\sum_{m_1',\dots,m_{B-1}'}$$

$$\sum_{y_{1,b-1}^n \in \mathcal{Y}_1^n} \mathrm{Pr}\left(Y_{1,b-1}^n = y_{1,b-1}^n \mid (M_1',M_1'') = (m_1',m_1''),\dots,(M_{b-1}',M_{b-1}'') = (m_{b-1}',m_{b-1}''), S_1^n = s_1^n,\dots,S_{b-1}^n = s_{b-1}^n\right)$$

$$\times \sum_{y_b^n,y_{1,b}^n:g_b''(y_b^n,m_1',\dots,m_{B-1}')\ne m_b''} W_{Y^n|X^n,X_1^n,S^n}(y_b^n|f_b(m_b',m_b''|m_{b-1}'),f_{1,b}(y_{1,b-1}^n),s_b^n)\,. \tag{69}$$

The conditioning in the equations above can be explained as follows. In (68), due to the code construction, the sequence $Y_{1,b}^n$ received at the relay in block $b \in [1:B]$ depends only on the messages (M_t',M_t'') with $t \le b$. The decoded message $\hat{M}_b'$, at the destination receiver, depends on messages M_t' with $t > b$, since the receiver decodes this part of the message backwards. In (69), since the second part of the message M_b'' is decoded after backward decoding is complete, the estimation of M_b'' at the decoder depends on the entire sequence $\hat{M}_1',\dots,\hat{M}_{B-1}'$. By (66)–(67), for every $t \in [1:b]$, h_b' and h_b'' as functions of s_{t+1}^n and s_t^n, respectively, satisfy (63) with $\alpha_n = e^{-2\theta n}$, given that the state sequences in the other blocks are fixed. Hence, applying Ahlswede's RT recursively, we obtain

$$\frac{1}{(n!)^{b+1}}\sum_{\pi_1,\pi_2,\dots,\pi_{b+1}\in\Pi_n} h_b'(\pi_1 s_1^n,\pi_2 s_2^n,\dots,\pi_{b+1}s_{b+1}^n) \le (n+1)^{B|\mathcal{S}|}e^{-2\theta n} \le e^{-\theta n}\,,\,,$$

$$\frac{1}{(n!)^{b}}\sum_{\pi_1,\pi_2,\dots,\pi_{b}\in\Pi_n} h_b''(\pi_1 s_1^n,\pi_2 s_2^n,\dots,\pi_{b}s_{b}^n) \le (n+1)^{B|\mathcal{S}|}e^{-2\theta n} \le e^{-\theta n}\,, \tag{70}$$

for all $(s_1^n,s_2^n,\dots,s_{b+1}^n) \in \mathcal{S}^{(b+1)n}$ and sufficiently large n, such that $(n+1)^{B|\mathcal{S}|} \le e^{\theta n}$.

On the other hand, for every $\pi_1,\pi_2,\dots,\pi_{b+1} \in \Pi_n$, we have that

$$h_b'(\pi_1 s_1^n,\pi_2 s_2^n,\dots,\pi_{b+1}s_{b+1}^n) = \mathbb{E}\, h_b'(\pi_1 s_1^n,\pi_2 s_2^n,\dots,\pi_{b+1}s_{b+1}^n|M_t',M_t'',t=1,\dots,b+1)\,, \tag{71}$$

with

$$h'_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_{b+1} s_{b+1}^n | m'_t, m''_t, t = 1, \ldots, b+1)$$

$$= \sum_{y_{1,1},\ldots,y_{1,b}} \prod_{t=0}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(y_{1,t+1}^n | f_{t+1}(m'_{t+1}, m''_{t+1}|m'_t), f_{1,t+1}(y_{1,t}^n), \pi_{t+1} s_{t+1}^n)$$

$$\times \sum_{y_{b+1}^n : g'_b(y_{b+1}^n, m'_{b+1}) \neq m'_b} W_{Y^n|X^n,X_1^n,S^n}(y_{b+1}^n | f_{b+1}(m'_{b+1}, m''_{b+1}|m'_b), f_{1,b+1}(y_{1,b}^n), \pi_{b+1} s_{b+1}^n)$$

$$\overset{(a)}{=} \sum_{y_{1,1},\ldots,y_{1,b}} \prod_{t=0}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(\pi_{t+1} y_{1,t+1}^n | f_{t+1}(m'_{t+1}, m''_{t+1}|m'_t), f_{1,b+1}(\pi_t y_{1,t}^n), \pi_{t+1} s_{t+1}^n)$$

$$\times \sum_{y_{b+1}^n : g'_b(\pi_{b+1} y_{b+1}^n, m'_{b+1}) \neq m'_b} W_{Y^n|X^n,X_1^n,S^n}(\pi_{b+1} y_{b+1}^n | f_{b+1}(m'_{b+1}, m''_{b+1}|m'_b), f_{1,b+1}(\pi_b y_{1,b}^n), \pi_{b+1} s_{b+1}^n)$$

$$\overset{(b)}{=} \sum_{y_{1,1},\ldots,y_{1,b}} \prod_{t=0}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(y_{1,t+1}^n | \pi_{t+1}^{-1} f_{t+1}(m'_{t+1}, m''_{t+1}|m'_t), \pi_{t+1}^{-1} f_{1,b+1}(\pi_t y_{1,t}^n), s_{t+1}^n)$$

$$\times \sum_{y_{b+1}^n : g'_b(\pi_{b+1} y_{b+1}^n, m'_{b+1}) \neq m'_b} W_{Y^n|X^n,X_1^n,S^n}(y_{b+1}^n | \pi_{b+1}^{-1} f_{b+1}(m'_{b+1}, m''_{b+1}|m'_b), \pi_{b+1}^{-1} f_{1,b+1}(\pi_b y_{1,b}^n), s_{b+1}^n), \tag{72}$$

where (a) is obtained by changing the order of summation over $y_{1,1}^n, \ldots, y_{1,b}^n$ and y_{b+1}^n; and (b) holds because the relay channel is memoryless. Similarly,

$$h''_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n) = \mathbb{E} h''_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n | M'_1, \ldots, M'_{B-1}, M''_t, t = 1, \ldots, b), \tag{73}$$

with

$$h''_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n | m'_1, \ldots, m'_{B-1}, m''_t, t = 1, \ldots, b)$$

$$= \sum_{y_{1,1},\ldots,y_{1,b-1}} \prod_{t=1}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(y_{1,t}^n | f_t(m'_t, m''_t|m'_{t-1}), f_{1,t}(y_{1,t}^n), \pi_t s_t^n)$$

$$\times \sum_{y_b^n : g''_b(y_b^n, m'_1, \ldots, m'_{B-1}) \neq m''_b} W_{Y^n|X^n,X_1^n,S^n}(y_b^n | f_b(m'_b, m''_b|m'_{b-1}), f_{1,b}(y_{1,b-1}^n), \pi_b s_b^n)$$

$$\overset{(a)}{=} \sum_{y_{1,1},\ldots,y_{1,b-1}} \prod_{t=1}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(\pi_t y_{1,t}^n | f_t(m'_t, m''_t|m'_{t-1}), f_{1,t}(\pi_{t-1} y_{1,t-1}^n), \pi_t s_t^n)$$

$$\times \sum_{y_b^n : g''_b(\pi_b y_b^n, m'_1, \ldots, m'_{B-1}) \neq m''_b} W_{Y^n|X^n,X_1^n,S^n}(\pi_b y_b^n | f_b(m'_b, m''_b|m'_{b-1}), f_{1,b}(\pi_{b-1} y_{1,b-1}^n), \pi_b s_b^n)$$

$$\overset{(b)}{=} \sum_{y_{1,1},\ldots,y_{1,b-1}} \prod_{t=1}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(y_{1,t}^n | \pi_t^{-1} f_t(m'_t, m''_t|m'_{t-1}), \pi_t^{-1} f_{1,t}(\pi_{t-1} y_{1,t-1}^n), s_t^n)$$

$$\times \sum_{y_b^n : g''_b(\pi_b y_b^n, m'_1, \ldots, m'_{B-1}) \neq m''_b} W_{Y^n|X^n,X_1^n,S^n}(y_b^n | \pi_b^{-1} f_b(m'_b, m''_b|m'_{b-1}), \pi_b^{-1} f_{1,b}(\pi_{b-1} y_{1,b-1}^n), s_b^n). \tag{74}$$

Then, consider the $(2^{nR(B-1)}, nB)$ random Markov block code $\mathscr{C}_{BM}^\Pi$, specified by

$$f_{b,\pi}(m'_b, m''_b|m'_{b-1}) = \pi_b^{-1} f_b(m'_b, m''_b|m'_{b-1}), \quad f_{1,b,\pi}(y_{1,b-1}^n) = \pi_b^{-1} f_{1,b}(\pi_{b-1} y_{1,b-1}^n), \tag{75a}$$

and

$$g'_{b,\pi}(y_{b+1}^n, \hat{m}'_{b+1}) = g'_b(\pi_{b+1} y_{b+1}^n, \hat{m}'_{b+1}), \quad g''_{b,\pi}(y_b^n, \hat{m}'_1, \ldots, \hat{m}'_{B-1}) = g''_b(\pi y_b^n, \hat{m}'_1, \ldots, \hat{m}'_{B-1}), \tag{75b}$$

for $\pi_1, \ldots, \pi_B \in \Pi_n$, with a uniform distribution $\mu(\pi_1, \ldots, \pi_B) = \frac{1}{|\Pi_n|^B} = \frac{1}{(n!)^B}$. That is, a set of B independent permutations is chosen at random and applied to all blocks simultaneously, while the order of the blocks remains intact. As we restricted ourselves to a block Markov code, the relaying function in a given block depends only on symbols received in the previous block, hence, the relay can implement those in-block permutations, and the coding scheme does not violate the causality requirement.

From (72) and (74), we see that using the random code $\mathscr{C}_{BM}^{\Pi}$, the error probabilities for the messages M_b' and M_b'' are given by

$$\Pr_{\mathscr{C}_{BM}^{\Pi}}\big(\mathcal{E}_b' \,|\, (\mathcal{E}_{b+1}')^c, S_1^n = s_1^n, \ldots, S_{b+1}^n = s_{b+1}^n\big) = \sum_{\pi_1, \ldots, \pi_B \in \Pi_n} \mu(\pi_1, \ldots, \pi_B) h_b'(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_{b+1} s_{b+1}^n),$$

$$\Pr_{\mathscr{C}_{BM}^{\Pi}}\big(\mathcal{E}_b'' \,|\, \mathcal{E}_1'^c, \ldots, \mathcal{E}_{B-1}'^c, S_1^n = s_1^n, \ldots, S_b^n = s_b^n\big) = \sum_{\pi_1, \ldots, \pi_B \in \Pi_n} \mu(\pi_1, \ldots, \pi_B) h_b''(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n), \tag{76}$$

for all $s_1^n, \ldots, s_{b+1}^n \in \mathcal{S}^n$, $b \in [1 : B-1]$, and therefore, together with (70), we have that the probability of error of the random code $\mathscr{C}_{BM}^{\Pi}$ is bounded by $P_e^{(n)}(q, \mathscr{C}_{BM}^{\Pi}) \le e^{-\theta n}$, for every $q(s^{nB}) \in \mathcal{P}(\mathcal{S}^{nB})$. That is, $\mathscr{C}_{BM}^{\Pi}$ is a $(2^{nR(B-1)}, nB, e^{-\theta n})$ random code for the AVRC $\mathcal{L}$, where the overall blocklength is nB, and the average rate $\frac{B-1}{B} \cdot R$ tends to R as $B \to \infty$. This completes the proof of the partial decode-forward lower bound. $\qquad\square$

B. Cutset Upper Bound

The proof immediately follows from Lemma 1, since the random code capacity of the AVRC is bounded by the random code capacity of the compound relay channel, *i.e.* $\mathbb{C}^{\star}(\mathcal{L}) \le \mathbb{C}^{\star}(\mathcal{L}^{\mathcal{P}(\mathcal{S})})$. $\qquad\square$

<div align="center">APPENDIX E
PROOF OF LEMMA 6</div>

We use the approach of [1], with the required adjustments. We use the random code constructed in the proof of Theorem 4. Let $R < \mathbb{C}^{\star}(\mathcal{L})$, and consider the case where the marginal sender-relay and sender-receiver AVCs have positive capacity, *i.e.*

$$\mathbb{C}(\mathcal{W}_1(x_{1,1})) > 0, \text{ and } \mathbb{C}(\mathcal{W}(x_{1,2})) > 0, \tag{77}$$

for some $x_{1,1}, x_{1,2} \in \mathcal{X}_1$ (see (24)). By Theorem 4, for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR}, n, \varepsilon)$ random code $\mathscr{C}^{\Gamma} = \big(\mu(\gamma) = \frac{1}{k}, \Gamma = [1 : k], \{\mathscr{C}_{\gamma}\}_{\gamma \in \Gamma}\big)$, where $\mathscr{C}_{\gamma} = (f_{\gamma}^n, f_{1,\gamma}, g_{\gamma})$, for $\gamma \in \Gamma$. Following Ahlswede's Elimination Technique [1], it can be assumed that the size of the code collection is bounded by $k = |\Gamma| \le n^2$. By (77), we have that for every $\varepsilon' > 0$ and sufficiently large ν', the code index $\gamma \in [1 : k]$ can be sent through the relay channel $W_{Y_1|X, X_1, S}$ using a $(2^{\nu'\widetilde{R}'}, \nu', \varepsilon')$ deterministic code $\mathscr{C}_i' = (\widetilde{f}^{\nu'}, \widetilde{g}')$, where $\widetilde{R}' > 0$, while the relay repeatedly transmits the symbol $x_{1,1}$. Since k is at most polynomial, the encoder can reliably convey γ to the relay with a negligible blocklength, *i.e.* $\nu' = o(n)$. Similarly, there exists $(2^{\nu''\widetilde{R}''}, \nu'', \varepsilon'')$ code $\mathscr{C}_i'' = (\widetilde{f}^{\nu''}, \widetilde{g}'')$ for the transmission of $\gamma \in [1 : k]$ through the channel $W_{Y|X, X_1, S}$ to the receiver, where $\nu'' = o(n)$ and $\widetilde{R}'' > 0$, while the relay repeatedly transmits the symbol $x_{1,2}$.

Now, consider a code formed by the concatenation of $\mathscr{C}_i'$ and $\mathscr{C}_i''$ as consecutive prefixes to a corresponding code in the code collection $\{\mathscr{C}_{\gamma}\}_{\gamma \in \Gamma}$. That is, the encoder first sends the index γ to the relay and the receiver, and then it sends the message $m \in [1 : 2^{nR}]$ to the receiver. Specifically, the encoder first transmits the $(\nu' + \nu'')$-sequence $(\widetilde{f}^{\nu'}(\gamma), \widetilde{f}^{\nu''}(\gamma))$ to convey the index γ, while the relay transmits the $(\nu' + \nu'')$-sequence $(\tilde{x}_1^{\nu'}, \tilde{x}_1^{\nu''})$, where $\tilde{x}_1^{\nu'} = (x_{1,1}, x_{1,1}, \ldots, x_{1,1})$ and $\tilde{x}_1^{\nu''} = (x_{1,2}, x_{1,2}, \ldots, x_{1,2})$. At the end of this transmission, the relay uses the first ν' symbols it received to estimate the code index as $\widehat{\gamma}' = \widetilde{g}'(\tilde{y}_1^{\nu'})$.

Then, the message m is transmitted by the codeword $x^n = f_{\gamma}(m)$, while the relay transmits $x_1^n = f_{1,\widehat{\gamma}'}^n(y_1^n)$. Subsequently, decoding is performed in two stages as well; the decoder estimates the index at first, with $\widehat{\gamma}'' = \widetilde{g}''(\tilde{y}^{\nu''})$, and the message is then estimated by $\widehat{m} = g_{\widehat{\gamma}''}(y^n)$. By the union of events bound, the probability of error is then bounded by $\varepsilon_c = \varepsilon + \varepsilon' + \varepsilon''$, for every joint distribution in $\mathcal{P}(\mathcal{S}^{\nu'+\nu''+n})$. That is, the concatenated code is a $(2^{(\nu'+\nu''+n)\widetilde{R}_n}, \nu' + \nu'' + n, \varepsilon_c)$ code over the AVRC $\mathcal{L}$, where the blocklength is $n + o(n)$, and the rate $\widetilde{R}_n = \frac{n}{\nu'+\nu''+n} \cdot R$ approaches R as $n \to \infty$. $\qquad\square$

<div align="center">APPENDIX F
PROOF OF COROLLARY 7</div>

Consider part 1. By Definition 3, if $W_{Y_1|X, X_1, S}$ and $W_{Y|X, X_1, S}$ are not symmetrizable -$\mathcal{X}|\mathcal{X}_1$ then there exist $x_{1,1}, x_{1,2} \in \mathcal{X}_1$ such that the DMCs $W_{Y_1|X, X_1, S}(\cdot|\cdot, x_{1,1}, \cdot)$ and $W_{Y|X, X_1, S}(\cdot|\cdot, x_{1,2}, \cdot)$ are non-symmetrizable in the sense of [8, Definition 2]. This, in turn, implies that $\mathbb{C}(\mathcal{W}_1(x_{1,1})) > 0$ and $\mathbb{C}(\mathcal{W}(x_{1,2})) > 0$, due to [8, Theorem 1]. Hence, by Lemma 6, $\mathbb{C}(\mathcal{L}) = \mathbb{C}^{\star}(\mathcal{L})$, and by Theorem 4, $\mathsf{R}_{DF}^{\star}(\mathcal{L}) \le \mathbb{C}(\mathcal{L}) \le \mathsf{R}_{CS}^{\star}(\mathcal{L})$. Parts 2 and 3 immediately follow from part 1 and Corollary 5. $\qquad\square$

<div align="center">APPENDIX G
PROOF OF LEMMA 8</div>

The proof is based on [10]. Let $\mathcal{L}$ be a symmetrizable-$\mathcal{X}|\mathcal{X}_1$. Assume to the contrary that a positive rate $R > 0$ can be achieved. That is, for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR}, n, \varepsilon)$ code $\mathscr{C} = (f, f_1, g)$. Hence, the size of the message set is at least 2, *i.e.*

$$\mathsf{M} \triangleq 2^{nR} \ge 2. \tag{78}$$

We now show that there exists a distribution $q(s^n)$ such that the probability of error $P_e^{(n)}(q, \mathscr{C})$ is bounded from below by a positive constant, in contradiction to the assumption above.

By Definition 3, there exists a conditional distribution $J(s|x)$ that satisfies (25). Then, consider the state sequence distribution $q(s^n) = \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} J^n(s^n|x^n(m))$, where $J^n(s^n|x^n) = \prod_{i=1}^{n} J(s_i|x_i)$ and $x^n(m) = f(m)$. For this distribution, the probability of error is given by

$$
\begin{aligned}
P_e^{(n)}(q, \mathscr{C}) =& \sum_{s^n \in \mathcal{S}^n} \left[\frac{1}{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} J^n(s^n|x^n(\widetilde{m})) \right] \cdot \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \sum_{(y^n, y_1^n): g(y^n) \neq m} W^n(y^n, y_1^n | x^n(m), f_1^n(y_1^n), s^n) \\
=& \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} \sum_{(y^n, y_1^n): g(y^n) \neq m} \sum_{s^n \in \mathcal{S}^n} W^n(y^n, y_1^n | x^n(m), f_1^n(y_1^n), s^n) J^n(s^n|x^n(\widetilde{m})) \\
&+ \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} \sum_{(y^n, y_1^n): g(y^n) \neq \widetilde{m}} \sum_{s^n \in \mathcal{S}^n} W^n(y^n, y_1^n | x^n(\widetilde{m}), f_1^n(y_1^n), s^n) J^n(s^n|x^n(m)) \quad (79)
\end{aligned}
$$

with $W^n \equiv W_{Y^n, Y_1^n | X^n, X_1^n, S^n}$ for short notation, where in the last sum we interchanged the summation indices m and $\widetilde{m}$. Then, consider the last sum, and observe that by (25), we have that

$$
\begin{aligned}
\sum_{s^n \in \mathcal{S}^n} W^n(y^n, y_1^n | x^n(\widetilde{m}), f_1^n(y_1^n), s^n) J^n(s^n|x^n(m)) &= \prod_{i=1}^{n} \left[\sum_{s_i \in \mathcal{S}} W(y_i, y_{1,i} | x_i(\widetilde{m}), f_{1,i}(y_1^{i-1}), s_i) J(s_i|x_i(m)) \right] \\
&= \prod_{i=1}^{n} \left[\sum_{s_i \in \mathcal{S}} W(y_i, y_{1,i} | x_i(m), f_{1,i}(y_1^{i-1}), s_i) J(s_i|x_i(\widetilde{m})) \right] \\
&= \sum_{s^n \in \mathcal{S}^n} W^n(y^n, y_1^n | x^n(m), f_1^n(y_1^n), s^n) J^n(s^n|x^n(\widetilde{m})) . \quad (80)
\end{aligned}
$$

Substituting (80) in (79), we have

$$
\begin{aligned}
P_e^{(n)}(q, \mathscr{C}) =& \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} \sum_{s^n \in \mathcal{S}^n} \left[\sum_{(y^n, y_1^n): g(y^n) \neq m} W^n(y^n, y_1^n | x^n(m), f_1^n(y_1^n), s^n) J^n(s^n|x^n(\widetilde{m})) \right. \\
&+ \left. \sum_{(y^n, y_1^n): g(y^n) \neq \widetilde{m}} W^n(y^n, y_1^n | x^n(m), f_1^n(y_1^n), s^n) J^n(s^n|x^n(\widetilde{m})) \right] \\
\geq& \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m} \neq m} \sum_{s^n \in \mathcal{S}^n} \sum_{y^n, y_1^n} W^n(y^n, y_1^n | x^n(m), f_1^n(y_1^n), s^n) J^n(s^n|x^n(\widetilde{m})) \\
=& \frac{\mathsf{M}(\mathsf{M} - 1)}{2\mathsf{M}^2} \geq \frac{1}{4} , \quad (81)
\end{aligned}
$$

where the last inequality follows from (78), hence a positive rate cannot be achieved. $\qquad\square$

APPENDIX H
PROOF OF LEMMA 9

Let $\mathcal{L} = \{W_{Y_1|X, X_1} W_{Y|Y_1, X_1, S}\}$ be a symmetrizable-$\mathcal{X}_1 \times \mathcal{Y}_1$ degraded AVRC. Assume to the contrary that a positive rate $R > 0$ can be achieved. That is, for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR}, n, \varepsilon)$ code $\mathscr{C} = (f, f_1, g)$. Hence, the size of the message set is at least 2, *i.e.*

$$
\mathsf{M} \triangleq 2^{nR} \geq 2 . \quad (82)
$$

We now show that there exists a distribution $q(s^n)$ such that the probability of error $P_e^{(n)}(q, \mathscr{C})$ is bounded from below by a positive constant, in contradiction to the assumption above. By Definition 4, there exists a conditional distribution $J(s|x_1, y_1)$ that satisfies (29). Then, consider the following state sequence distribution,

$$
q(s^n) = \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \sum_{y_1^n \in \mathcal{Y}_1} W_{Y_1^n | X^n, X_1^n}(y_1^n | f(m), f_1^n(y_1^n)) J^n(s^n | f_1^n(y_1^n), y_1^n) , \quad (83)
$$

where $J^n(s^n|x_1^n, y_1^n) = \prod_{i=1}^n J(s_i|x_{1,i}, y_{1,i})$. For this distribution, the probability of error is given by

$$
P_e^{(n)}(q, \mathscr{C}) = \sum_{s^n \in \mathcal{S}^n} \left[\frac{1}{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} \sum_{\tilde{y}_1^n} W_{Y_1^n|X^n, X_1^n}(\tilde{y}_1^n|f(\widetilde{m}), f_1^n(\tilde{y}_1^n)) J^n(s^n|f_1^n(\tilde{y}_1^n), \tilde{y}_1^n) \right]
$$

$$
\times \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \sum_{(y^n, y_1^n): g(y^n) \neq m} W_{Y_1^n|X^n, X_1^n}(y_1^n|f(m), f_1^n(y_1^n)) W^n(y^n|y_1^n, f_1^n(y_1^n), s^n)
$$

$$
= \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} \sum_{y_1^n, \tilde{y}_1^n} W_{Y_1^n|X^n, X_1^n}(\tilde{y}_1^n|f(\widetilde{m}), f_1^n(\tilde{y}_1^n)) \cdot W_{Y_1^n|X^n, X_1^n}(y_1^n|f(m), f_1^n(y_1^n))
$$

$$
\times \sum_{y^n: g(y^n) \neq m} \sum_{s^n \in \mathcal{S}^n} W^n(y^n|y_1^n, f_1^n(y_1^n), s^n) J^n(s^n|f_1^n(\tilde{y}_1^n), \tilde{y}_1^n)
$$

$$
+ \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} \sum_{y_1^n, \tilde{y}_1^n} W_{Y_1^n|X^n, X_1^n}(y_1^n|f(m), f_1^n(y_1^n)) \cdot W_{Y_1^n|X^n, X_1^n}(\tilde{y}_1^n|f(\widetilde{m}), f_1^n(\tilde{y}_1^n))
$$

$$
\times \sum_{y^n: g(y^n) \neq \widetilde{m}} \sum_{s^n \in \mathcal{S}^n} W^n(y^n|\tilde{y}_1^n, f_1^n(\tilde{y}_1^n), s^n) J^n(s^n|f_1^n(y_1^n), y_1^n) \tag{84}
$$

with $W^n \equiv W_{Y^n|Y_1^n, X_1^n, S^n}$ for short notation, where in the last sum we interchanged the summation variables (m, y_1^n) and $(\widetilde{m}, \tilde{y}_1^n)$. Then, consider the last sum, and observe that by (29), we have that

$$
\sum_{s^n \in \mathcal{S}^n} W^n(y^n|\tilde{y}_1^n, f_1^n(\tilde{y}_1^n), s^n) J^n(s^n|f_1^n(y_1^n), y_1^n) = \prod_{i=1}^n \left[\sum_{s_i \in \mathcal{S}} W(y_i|\tilde{y}_{1,i}, f_{1,i}(\tilde{y}_1^{i-1}), s_i) J(s_i|f_{1,i}(y_1^{i-1}), y_{1,i}) \right]
$$

$$
= \prod_{i=1}^n \left[\sum_{s_i \in \mathcal{S}} W(y_i|y_{1,i}, f_{1,i}(y_1^{i-1}), s_i) J(s_i|f_{1,i}(\tilde{y}_1^{i-1}), \tilde{y}_{1,i}) \right]
$$

$$
= \sum_{s^n \in \mathcal{S}^n} W^n(y^n|y_1^n, f_1^n(y_1^n), s^n) J^n(s^n|f_1^n(\tilde{y}_1^n), \tilde{y}_1^n). \tag{85}
$$

Substituting (85) in (84), we have

$$
P_e^{(n)}(q, \mathscr{C}) = \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m}=1}^{\mathsf{M}} \sum_{y_1^n, \tilde{y}_1^n} W_{Y_1^n|X^n, X_1^n}(\tilde{y}_1^n|f(\widetilde{m}), f_1^n(\tilde{y}_1^n)) W_{Y_1^n|X^n, X_1^n}(y_1^n|f(m), f_1^n(y_1^n))
$$

$$
\times \sum_{s^n \in \mathcal{S}^n} \left[\sum_{y^n: g(y^n) \neq m} W^n(y^n|y_1^n, f_1^n(y_1^n), s^n) J^n(s^n|f_1^n(\tilde{y}_1^n), \tilde{y}_1^n) + \sum_{y^n: g(y^n) \neq \widetilde{m}} W^n(y^n|y_1^n, f_1^n(y_1^n), s^n) J^n(s^n|f_1^n(\tilde{y}_1^n), \tilde{y}_1^n) \right]
$$

$$
\geq \frac{1}{2\mathsf{M}^2} \sum_{m=1}^{\mathsf{M}} \sum_{\widetilde{m} \neq m} \sum_{y_1^n, \tilde{y}_1^n} W^n(\tilde{y}_1^n|f(\widetilde{m}), f_1^n(\tilde{y}_1^n)) W^n(y_1^n|f(m), f_1^n(y_1^n)) \sum_{s^n \in \mathcal{S}^n} \sum_{y^n \in \mathcal{Y}^n} W^n(y^n|y_1^n, f_1^n(y_1^n), s^n) J^n(s^n|f_1^n(\tilde{y}_1^n), \tilde{y}_1^n)
$$

$$
= \frac{\mathsf{M}(\mathsf{M}-1)}{2\mathsf{M}^2} \geq \frac{1}{4}, \tag{86}
$$

where the last inequality follows from (82), hence a positive rate cannot be achieved. $\qquad \square$

APPENDIX I
ANALYSIS OF EXAMPLE 1

We show that the random code capacity of the AVRC in Example 1 is given by $\mathbb{C}^\star(\mathcal{L}) = \min\left\{\frac{1}{2}, 1 - h(\theta)\right\}$. As the AVRC is degraded, the random code capacity is given by

$$
\mathbb{C}^\star(\mathcal{L}) = \mathsf{R}_{DF}^\star(\mathcal{L}) = \mathsf{R}_{CS}^\star(\mathcal{L}) = \max_{p(x, x_1)} \min\left\{ \min_{0 \leq q \leq 1} I_q(X, X_1; Y), \, I(X; Y_1|X_1) \right\}, \tag{87}
$$

due to part 2 of Corollary 5, where $q \equiv q(1) = 1 - q(0)$. Now, consider the direct part. Set $p(x, x_1) = p(x)p(x_1)$, where $X \sim \text{Bernoulli}(1/2)$ and $X_1 \sim \text{Bernoulli}(1/2)$. Then,

$$
I(X; Y_1|X_1) = 1 - h(\theta),
$$

$$
H_q(Y) = \frac{1}{2}\left[-q \log\left(\frac{1}{2}q\right) - (1-q)\log\left(\frac{1}{2}(1-q)\right) \right] - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1 + \frac{1}{2}h(q),
$$

$$
H_q(Y|X, X_1) = h(q). \tag{88}
$$

Hence,

$$\mathbb{C}^{\star}(\mathcal{L}) \geq \min\left\{\min_{0 \leq q \leq 1}\left[1 - \frac{1}{2}h(q)\right], 1 - h(\theta)\right\} = \min\left\{\frac{1}{2}, 1 - h(\theta)\right\}. \tag{89}$$

As for the converse part, we have the following bounds,

$$\mathbb{C}^{\star}(\mathcal{L}) \leq \max_{p(x,x_1)} I(X; Y_1|X_1) = 1 - h(\theta), \tag{90}$$

and

$$\mathbb{C}^{\star}(\mathcal{L}) \leq \max_{p(x,x_1)} \min_{0 \leq q \leq 1} I_q(X, X_1; Y) \leq \max_{p(x,x_1)} \left[H_q(Y) - H_q(Y|X, X_1)\right]\Big|_{q=\frac{1}{2}}$$
$$= \max_{0 \leq p \leq 1}\left[1 + \frac{1}{2}h(p)\right] - 1 = \frac{1}{2}, \tag{91}$$

where $p \triangleq \Pr(X_1 = 1)$. $\qquad\square$

APPENDIX J
PROOF OF LEMMA 10

The proof follows the lines of [11]. Consider an AVRC $\mathcal{L} = \{W_{Y|X',X_1} \, W_{Y_1|X'',X_1,S}\}$ with orthogonal sender components. We apply Theorem 4, which states that $\mathsf{R}^{\star}_{DF}(\mathcal{L}) \leq \mathbb{C}^{\star}(\mathcal{L}) \leq \mathsf{R}^{\star}_{CS}(\mathcal{L})$.

A. Achievability Proof

To show achievability, we set $U = X''$ and $p(x', x'', x_1) = p(x_1)p(x'|x_1)p(x''|x_1)$ in the partial decode-forward lower bound $\mathsf{R}^{\star}_{DF}(\mathcal{L}) \triangleq \mathsf{R}_{DF}(\mathcal{L}^{\mathcal{Q}})\Big|_{\mathcal{Q}=\mathcal{P}(\mathcal{S})}$. Hence, by (9),

$$\mathsf{R}^{\star}_{DF}(\mathcal{L}_2) \geq \max_{p(x_1)p(x'|x_1)p(x''|x_1)} \min\left\{I(X', X'', X_1; Y), \, \min_{q(s)} I_q(X''; Y_1|X_1) + I(X'; Y|X_1, X'')\right\}. \tag{92}$$

Now, by (30), we have that $(X'', Y_1) \multimap (X', X_1) \multimap Y$ form a Markov chain. As $(X_1, X', X'') \sim p(x_1)p(x'|x_1)p(x''|x_1)$, it further follows that $(X'', Y_1) \multimap X_1 \multimap Y$ form a Markov chain, hence $I(X', X'', X_1; Y) = I(X', X_1; Y)$ and $I(X'; Y|X_1, X'') = I(X'; Y|X_1)$. Thus, (92) reduces to the expression in the RHS of (31). If $W_{Y_1|X'',X_1,S}$ is non-symmetrizable-$\mathcal{X}''|\mathcal{X}_1$, then (92) is achievable by deterministic codes as well, due to Corollary 7. $\qquad\square$

B. Converse Proof

By (8) and (20), the cutset upper bound takes the following form,

$$\mathsf{R}^{\star}_{CS}(\mathcal{L}) = \min_{q(s)} \max_{p(x',x'',x_1)} \min\left\{I(X', X'', X_1; Y), \, I_q(X', X''; Y, Y_1|X_1)\right\}$$
$$= \max_{p(x',x'',x_1)} \min\left\{I(X', X'', X_1; Y), \, \min_{q(s)} I_q(X', X''; Y, Y_1|X_1)\right\}, \tag{93}$$

where the last line is due to the minimax theorem [16]. For the AVRC with orthogonal sender components, as specified by (30), we have the following Markov relations,

$$Y_1 \multimap (X'', X_1) \multimap (X', Y), \tag{94}$$
$$(X'', Y_1) \multimap (X', X_1) \multimap Y. \tag{95}$$

Hence, by (95), $I(X', X'', X_1; Y) = I(X', X_1; Y)$. As for the second mutual information in the RHS of (93), by the mutual information chain rule,

$$I_q(X', X''; Y, Y_1|X_1) = I_q(X''; Y_1|X_1) + I_q(X'; Y_1|X'', X_1) + I_q(X', X''; Y|X_1, Y_1)$$
$$\overset{(a)}{=} I_q(X''; Y_1|X_1) + I_q(X', X''; Y|X_1, Y_1)$$
$$\overset{(b)}{=} I_q(X''; Y_1|X_1) + H_q(Y|X_1, Y_1) - H(Y|X', X_1)$$
$$\overset{(c)}{\leq} I_q(X''; Y_1|X_1) + I(X'; Y|X_1) \tag{96}$$

where (a) is due to (94), (b) is due to (95), and (c) holds since conditioning reduces entropy. Therefore,

$$\mathsf{R}^{\star}_{CS}(\mathcal{L}) \leq \max_{p(x',x'',x_1)} \min\left\{I(X', X_1; Y), \, \min_{q(s)} I_q(X''; Y_1|X_1) + I(X'; Y|X_1)\right\}. \tag{97}$$

Without loss of generality, the maximization in (97) can be restricted to distributions of the form $p(x', x'', x_1) = p(x_1)\cdot p(x'|x_1)\cdot p(x''|x_1)$. $\qquad\square$

References

[1] R. Ahlswede. "Elimination of correlation in random codes for arbitrarily varying channels". *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* 44.2 (June 1978), pp. 159–175.

[2] R. Ahlswede. "Arbitrarily varying channels with states sequence known to the sender". *IEEE Trans. Inform. Theory* 32.5 (Sept. 1986), pp. 621–629.

[3] A. Behboodi and P. Piantanida. "On the simultaneous relay channel with informed receivers". *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2009)*. Seoul, South Korea, June 2009, pp. 1179–1183.

[4] D. Blackwell, L. Breiman, and A. J. Thomasian. "The capacities of certain channel classes under random coding". *Ann. Math. Statist.* 31.3 (Sept. 1960), pp. 558–567.

[5] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[6] T. Cover and A. E. Gamal. "Capacity theorems for the relay channel". *IEEE Trans. Inform. Theory* 25.5 (Sept. 1979), pp. 572–584.

[7] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. 2nd ed. Cambridge University Press, 2011.

[8] I. Csiszár and P. Narayan. "The capacity of the arbitrarily varying channel revisited: positivity, constraints". *IEEE Trans. Inform. Theory* 34.2 (Mar. 1988), pp. 181–193.

[9] A. El Gamal and Y.H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[10] T. Ericson. "Exponential error bounds for random codes in the arbitrarily varying channel". *IEEE Trans. Inform. Theory* 31.1 (Jan. 1985), pp. 42–48.

[11] A. El Gamal and S. Zahedi. "Capacity of a class of relay channels with orthogonal components". *IEEE Trans. Inform. Theory* 51.5 (May 2005), pp. 1815–1817.

[12] E. Hof and S. I. Bross. "On the deterministic-code capacity of the two-user discrete memoryless Arbitrarily Varying General Broadcast channel with degraded message sets". *IEEE Trans. Inform. Theory* 52.11 (Nov. 2006), pp. 5023–5044.

[13] E. C. van der Meulen. "Three-terminal communication channels". *Adv. Appl. Prob.* 3.1 (1971), pp. 120–154.

[14] S. Sigurjónsson and Y. H. Kim. "On multiple user channels with state information at the transmitters". *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2005)*. Adelaide, Australia, Sept. 2005, pp. 72–76.

[15] O. Simeone, D. Gündüz, and S. Shamai. "Compound relay channel with informed relay and destination". *Proc. Allerton Conf. Commun., Control, Computing*. Monticello, Illinois, Sept. 2009, pp. 692–699.

[16] M. Sion. "On General Minimax Theorems". *Pacific J. Math* 8.1 (Mar. 1958), pp. 171–176.