entropy

MDPI

Article

The Arbitrarily Varying Relay Channel †

Uzi Pereg * and Yossef Steinberg

Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel; ysteinbe@ee.technion.ac.il

* Correspondence: uzipereg@campus.technion.ac.il

† Parts of this work have been presented at the 2018 International Symposium on Information Theory, Vail, Colorado, 17–22 June 2018, and at the 56th Annual Allerton Conference on Communication, Control, and Computing, Monticelo, Illinoi, 3–5 October 2018. This work was supported by the Israel Science Foundation (grant No. 1285/16).

Received: 26 March 2019; Accepted: 20 May 2019; Published: 22 May 2019

Abstract: We study the arbitrarily varying relay channel, which models communication with relaying in the presence of an active adversary. We 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. In addition, we consider the arbitrarily varying Gaussian relay channel with sender frequency division under input and state constraints. We determine the random code capacity, and establish lower and upper bounds on the deterministic code capacity. Furthermore, we show that as opposed to previous relay models, the primitive relay channel has a different behavior compared to the non-primitive relay channel in the arbitrarily varying scenario.

Keywords: arbitrarily varying channel; relay channel; decode-forward; Markov block code; minimax theorem; deterministic code; random code; symmetrizability.

1. Introduction

The relay channel was first introduced by van der Meulen [1] 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. The relay channel is generally perceived as a fundamental building block for multihop networks (see e.g., [2,3], Chapter 16), where some nodes receive and transmit in order to assist the information flow between other nodes. 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 [4]. It was also shown in [4] that for the reversely degraded relay channel, direct transmission is capacity achieving. For the degraded relay channel, the decode-forward lower bound and the cutset upper bound coincide, thus characterizing the capacity for this model [4].

In general, the partial decode-forward lower bound is tighter than both direct transmission and decode-forward lower bounds. El Gamal and Zahedi [5] determined the capacity of the relay channel with orthogonal sender components, by showing that the partial decode-forward lower bound and cutset upper bound coincide. A variation of the relay channel, referred to as the primitive relay channel, was introduced by Kim [2], and attracted a lot of attention (see e.g., [6–12] and references therein). Recently, there has also been a growing interest in the Gaussian relay channel, as e.g., in [5,7,9,13–16] and references therein. In particular, El Gamal and Zahedi [5] introduced the Gaussian relay channel with sender frequency division (SFD), as a special case of a relay channel with orthogonal sender components. There are many other relaying scenarios, including secrecy [17,18], networking [15,19–22], parallel relaying [23–25], diamond channels [26–28], side information [29–33], etc.

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 [34]. In real systems, such variations are caused by fading in wireless communication [35–42], memory faults in storage [43–47], malicious attacks on identification and authorization systems [48,49], etc.It is especially relevant to communication in the presence of an adversary, or a *jammer*, attempting to disrupt communication. Jamming attacks are not limited to point-to-point communication, and cause a major security concern for cognitive radio networks [50] and wireless sensor networks [42,51–54], for instance.

Considering the AVC without a relay, Blackwell et al. determined the random code capacity [34], i.e., the capacity achieved by stochastic-encoder stochastic-decoder coding schemes with common randomness. It was also demonstrated in [34] that the random code capacity is not necessarily achievable using deterministic codes. A well-known result by Ahlswede [55] is the dichotomy property of the AVC. Specifically, the deterministic code capacity either equals the random code capacity or else, it is zero. Subsequently, Ericson [56] and Csiszár and Narayan [57] established a simple single-letter condition, namely non-symmetrizability, which is both necessary and sufficient for the capacity to be positive. Ahlswede's Robustification Technique (RT) is a useful technique for the AVC analysis, developed and applied to classical AVC settings [58,59]. Essentially, the RT uses a reliable code for the compound channel to construct a random code for the AVC applying random permutations to the codeword symbols. A continuing line of works on arbitrarily varying networks includes among others the arbitrarily varying broadcast channel [60–65], multiple-access channel [60,66–75], and wiretap channel [76–84]. The reference lists here are far from being exhaustive.

In this work, we introduce a new model, namely, the arbitrarily varying relay channel (AVRC). The AVRC combines the previous models, i.e., the relay channel and the AVC, and we believe that it is a natural problem to consider, in light of the jamming attacks on current and future networks, as mentioned above. In the analysis, we incorporate the block Markov coding schemes of [4] in Ahlswede's Robustification and Elimination Techniques [55,59]. A straightforward application of Ahlswede's RT fails to comply with the strictly causal relay transmission. In a recent work [85,86], by the authors of this paper, a modified RT technique was presented and applied to the point-to-point AVC with causal side information under input and state constraints, without a relay. This was the first time where the application of the RT exploited the structure of the original compound channel code to construct a random code for the AVC, as opposed to earlier work where the original code is treated as a "black box". Here, we present another modification of the RT, which also exploits the structure of the original compound channel code, but in a different manner. The analysis also requires to redefine the compound channel, and we refer to the newly defined channel as the *block-compound relay channel*.

We establish the cutset upper bound and the full/partial decode-forward lower bound on the random code capacity of the AVRC. The random code capacity is determined in special cases of the degraded AVRC, the reversely degraded AVRC, and the AVRC with orthogonal sender components. Then, we give extended non-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. Then, we consider the Gaussian AVRC with SFD, under input and state constraints. The random code capacity is determined using the previous results, whereas the deterministic code capacity is lower and upper bounded using an independent approach. Specifically, we extend the techniques from [87], where Csiszár and Narayan determine the capacity of the Gaussian AVC under input and state constraint. It is shown that for low values on the input constraint, the deterministic code capacity can be strictly lower than the random code capacity, but yet non-zero.

Furthermore, we give similar bounds for the primitive AVRC, where there is a noiseless link between the relay and the receiver of limited capacity [2]. We find the capacity of the primitive counterpart of the Gaussian AVRC with SFD, in which case the deterministic and random code capacities coincide, regardless of the value of the input constraint. We deduce that Kim's assertion—that

"the primitive relay channel captures most essential features and challenges of relaying, and thus serves as a good testbed for new relay coding techniques" [2]—is not true in the arbitrarily varying scenario.

This work is organized as follows. In Section 2, the basic definitions and notation are provided. In Section 3, we give the main results on the general AVRC. The Gaussian AVRC with SFD is introduced in Section 4, and the main results are given in Section 5. The definition and results on the primitive AVRC are in Section 6.

2. Definitions

2.1. Notation

We use the following notation conventions throughout. Calligraphic letters $\mathcal{X}, \mathcal{S}, \mathcal{Y}, \ldots$ 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 $\|\mathbf{x}\|$.

2.2. 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 [29], a relay channel $W_{Y,Y_1|X,X_1,S}$ is called degraded if the channel can be expressed as

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

and it is called reversely degraded if

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

We say that the relay channel is *strongly* degraded or reversely degraded, if the respective definition holds such that the sender-relay marginal is independent of the state. That is, $W_{Y,Y_1|X,X_1,S}$ is strongly degraded if $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X,X_1} W_{Y|Y_1,X_1,S}$, and similarly, $W_{Y,Y_1|X,X_1,S}$ is strongly reversely degraded if $W_{Y,Y_1|X,X_1,S} = W_{Y|X,X_1,S} W_{Y_1|Y,X_1}$. For example, if $Y_1 = X + Z$ and $Y = Y_1 + X_1 + S$, where Z is an independent additive noise, then $W_{Y,Y_1|X,X_1,S}$ is strongly degraded. Whereas, if $Y = X + X_1 + S$ and $Y_1 = Y + Z$, then $W_{Y,Y_1|X,X_1,S}$ is strongly reversely degraded.

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 [30,88]. 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.

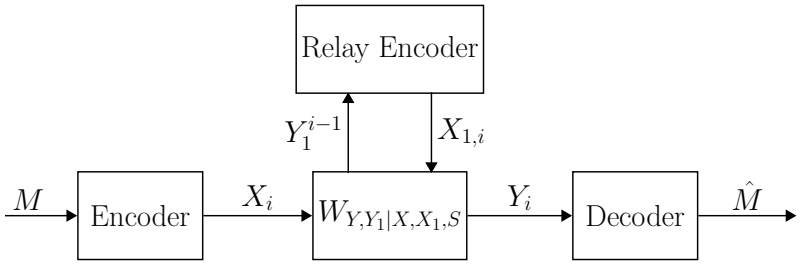

Figure 1. Communication over the arbitrarily varying relay channel $\mathcal{L} = \{W_{Y,Y_1|X,X_1,S}\}$. Given a message M, the encoder transmits $X^n = f(M)$. At time $i \in [1:n]$, the relay transmits $X_{1,i}$ based on all the symbols of the past Y_1^{i-1} and then receives a new symbol $Y_{1,i}$.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

2.3. 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 $x_1^n = f_1^n(y_1^n) \triangleq \left(f_{1,i}(y_1^{i-1})\right)_{i=1}^n$. 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 $\mathscr{C} = \left(f(\cdot), f_1^n(\cdot), g(\cdot)\right)$. 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.

3. Main Results—General AVRC

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

3.1. 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

$$\begin{aligned}\mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}}) \triangleq \max_{p(u,x,x_1)}\min\Big\{&\inf_{q\in\mathcal{Q}} I_q(U,X_1;Y) + \inf_{q\in\mathcal{Q}} I_q(X;Y|X_1,U)\,,\\ &\inf_{q\in\mathcal{Q}} I_q(U;Y_1|X_1) + \inf_{q\in\mathcal{Q}} I_q(X;Y|X_1,U)\Big\}\,,\end{aligned} \tag{9}$$

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

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

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q}}) \geq \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}})\,, \tag{10}$$

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

Specifically, if $R < \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}})$, 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. The achievability proof is based on block Markov coding interlaced with the partial decode-forward scheme. That is, the encoder sends a

sequence of messages over multiple blocks. The message in each block consists of two components, a decode-forward component, and a direct transmission component, where only the former is decoded by the relay. The name 'decode-forward component' stands for the fact that the relay decodes this message component and sends its estimation forwards, to the destination receiver. Once the decoder has received all blocks, the decode-forward components are decoded backwards, i.e., starting with the message in the last block going backwards. Using the estimation of the decode-forward components, the direct transmission components are decoded forwards, i.e., starting with the message in the first block going forwards. The ambiguity of the state distribution needs to be dealt with throughout all of those estimations. In both decoding stages, the receiver performs joint typicality decoding using a set of types that "quantizes" the set $\mathcal{Q}$ of state distributions.

Remark 1. *If the set of state distributions $\mathcal{Q}$ is convex, then the upper bound expression in the RHS of Equation (8) has a* min max *form. On the other hand, in the lower bound expression in the RHS of Equation (9), the maximum comes first, and then we have multiple* min *terms, which makes this expression a lot more complicated than the classical partial decode-forward bound [4] (see also [3], Theorem 16.3), where Markov properties lead to a simpler expression. We note that this phenomenon (or one might say, disturbance) where the lower bound has multiple* min *terms is not exclusive to the AVRC. A noteworthy example is the arbitrarily varying wiretap channel [76,89], where the lower bound has the form of* $\max[\min I_q(U;Y) - \max I_q(U;Z)]$. *While the capacity of the classical wiretap channel is known, the arbitrarily varying counterpart has remained an open problem for several years.*

Observe that taking $U = \emptyset$ in (9) gives the direct transmission lower bound,

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

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

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

This yields the following corollary. The corollary uses the terms of a strongly degraded relay channel, for which $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X,X_1}W_{Y|Y_1,X_1,S}$, and a strongly reversely degraded relay channel, for which $W_{Y,Y_1|X,X_1,S} = W_{Y|X,X_1,S}W_{Y_1|Y,X_1}$, as defined in Section 2.2.

Corollary 1. *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 strongly reversely degraded, then*

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) = \min_{q\in\mathcal{Q}} \max_{p(x,x_1)} I_q(X;Y|X_1)\,. \tag{14}$$

2. *If $W_{Y,Y_1|X,X_1,S}$ is strongly degraded, then*

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}}) = \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) = \max_{p(x,x_1)} \min\left\{\min_{q\in\mathcal{Q}} I_q(X,X_1;Y)\,,\ I(X;Y_1|X_1)\right\}\,. \tag{15}$$

The proof of Corollary 1 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, along with the convexity considerations in the remark below.

Remark 2. *On a technical level, there are two purposes for considering the strongly degraded relay channel, for which the marginal channel to the relay is independent of the state, i.e., $W_{Y_1|X,X_1,S} = W_{Y_1|X,X_1}$ (see Section 2.2). First, this ensures that $X - (X_1,Y_1) - Y$ form a Markov chain, without conditioning on S. Secondly,*

as pointed out in Remark 1, there is a difference between the order of the min *and* max *in the lower and upper bounds (cf. (8) and (9)). Thereby, proving the capacity results of Corollary 1 above, we apply the minimax theorem. In general, a pointwise minimum of two convex functions may not necessarily yield a convex function. Nevertheless, having assumed that the relay channel is strongly degraded, the functional* $G(p,q) = \min\{I_q(X, X_1; Y)\,,\ I(X; Y_1|X_1)\}$ *is* quasi*-convex in the state distribution, i.e.,*

$$G(p, (1-\alpha)q_1 + \alpha q_2)) \leq \max\left(G(p,q_1), G(p,q_2)\right)\,, \tag{16}$$

for every $p \in \mathcal{P}(\mathcal{X} \times \mathcal{X}_1)$, $q_1, q_2 \in \mathcal{Q}$, *and* $0 \leq \alpha \leq 1$. *The quasi-convex shape is illustrated in Figure 2, which depicts* $G(p,q)$ *for an example given in the sequel. By [90] (Theorem 3.4), the minimax theorem applies to quasi-convex functions as well, which alleviates the proof of Corollary 1.*

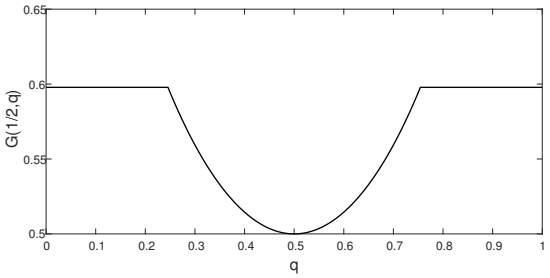

Figure 2. The functional $G(p,q) = \min\{I_q(X, X_1; Y)\,,\ I(X; Y_1|X_1)\}$, for $S \sim \text{Bernoulli}(q)$, $0 \leq q \leq 1$, as a function of q. The figure corresponds to Example 1, where $G(p,q) = \min\{1 - \frac{1}{2}h(q)\,,\ 1 - h(\theta)\}$, for $p(x, x_1) = p(x)p(x_1)$, with $X \sim \text{Bernoulli}(\frac{1}{2})$ and $X_1 \sim \text{Bernoulli}(\frac{1}{2})$, and $\theta = 0.08$. Clearly, $G(p,q)$ is not convex in q, but rather quasi-convex in q.

The following corollary is a direct consequence of Lemma 1 and it is significant for the random code analysis of the AVRC.

Corollary 2. *The capacity of the block-compound relay channel* $\mathcal{L}^{\mathcal{Q}\times B}$ *is bounded by*

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q}\times B}) \geq \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}})\,, \tag{17}$$

$$\mathbb{C}^{\star}(\mathcal{L}^{\mathcal{Q}\times B}) \leq \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}})\,. \tag{18}$$

Specifically, if $R < \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}})$, *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 2 is given in Appendix C.

3.2. The AVRC

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

3.2.1. Random Code Lower and Upper Bounds

The random code bounds below are obtained through a modified version of Ahlswede's RT, using our results on the block-compound relay channel in Corollary 2. Define

$$\mathsf{R}^{\star}_{PDF}(\mathcal{L}) \triangleq \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}})\Big|_{\mathcal{Q}=\mathcal{P}(\mathcal{S})}\,,\ \mathsf{R}^{\star}_{CS}(\mathcal{L}) \triangleq \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}})\Big|_{\mathcal{Q}=\mathcal{P}(\mathcal{S})}\,. \tag{19}$$

Theorem 1. *The random code capacity of an AVRC $\mathcal{L}$ is bounded by*

$$\mathsf{R}^{\star}_{PDF}(\mathcal{L}) \leq \mathbb{C}^{\star}(\mathcal{L}) \leq \mathsf{R}^{\star}_{CS}(\mathcal{L}) \,. \tag{20}$$

The proof of Theorem 1 is given in Appendix D. To prove Theorem 1 we modify Ahlswede's RT. A straightforward application of Ahlswede's RT fails to comply with the strictly causal relay transmission. Essentially, the RT uses a reliable code for the compound channel code to construct a random code for the AVC, applying random permutations to the transmitted codeword. However, the relay cannot apply permutations to its transmission, since at time $i \in [1:n]$, the relay cannot compute $f_{1,j}(y_1^{j-1})$, for $j > i$, as the relay encoder only knows the past received symbols $y_{1,1}, \ldots, y_{1,i-1}$, and does not have access to the symbols $y_{1,i}, \ldots, y_{1,j-1}$ which will be received in the future. To resolve this difficulty, we use a block Markov code for the block compound channel. In a block Markov coding scheme, the relay sends $x_{1,b}^n$ in block b, using the sequence of symbols $y_{1,b-1}^n$ received in the previous block. Since the entire sequence $y_{1,b-1}^n$ is known to the relay encoder, permutations can be applied to the transmission in each block separately. Hence, our proof exploits the structure of the original block-compound channel code to construct a random code for the AVRC, as opposed to classical works where the RT is used such that the original code is treated as a "black box" [59].

Remark 3. *Block Markov coding with partial decode-forward is not a simple scheme by itself, and thus, using the RT requires careful attention. In particular, by close inspection of the proof of Theorem 1, one may recognize that the necessity of using the block-compound relay channel, rather than the standard compound channel, stems from the fact that for the AVRC, the state sequences may have completely different types in each block. For each block, we use the RT twice. First, the RT is applied to the probability of the backward decoding error, for the message component which is decoded by the relay. Then, it is applied to the probability of forward decoding error, for the message component which is transmitted directly.*

Together with Corollary 1, the theorem above yields another corollary.

Corollary 3. *Let $\mathcal{L}$ be an AVRC.*

1. *If $W_{Y,Y_1|X,X_1,S}$ is strongly reversely degraded,*

$$\mathbb{C}^{\star}(\mathcal{L}) = \mathsf{R}^{\star}_{PDF}(\mathcal{L}) = \mathsf{R}^{\star}_{CS}(\mathcal{L}) = \min_{q(s)} \max_{p(x,x_1)} I_q(X;Y|X_1) \,. \tag{21}$$

2. *If $W_{Y,Y_1|X,X_1,S}$ is strongly degraded,*

$$\mathbb{C}^{\star}(\mathcal{L}) = \mathsf{R}^{\star}_{PDF}(\mathcal{L}) = \mathsf{R}^{\star}_{CS}(\mathcal{L}) = \max_{p(x,x_1)} \min \left\{ \min_{q(s)} I_q(X, X_1; Y) \,,\; I(X; Y_1|X_1) \right\} . \tag{22}$$

Before we proceed to the deterministic code capacity, we note that Ahlswede's Elimination Technique [55] 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.

3.2.2. 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 is positive, and it 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,

$$\mathcal{W}_1(x_1) = \{W_{Y_1|X,X_1,S}(\cdot|\cdot, x_1, \cdot)\} \,,\; \mathcal{W}(x_1) = \{W_{Y|X,X_1,S}(\cdot|\cdot, x_1, \cdot)\} \,. \tag{23}$$

See Figure 3.

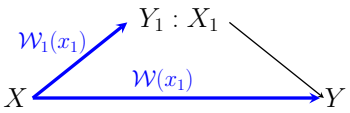

Figure 3. The marginals of the arbitrarily varying relay channel. For every relay transmission $x_1 \in \mathcal{X}_1$, the marginal sender-relay AVC is denoted by $\mathcal{W}_1(x_1) = \{W_{Y_1|X,X_1,S}(\cdot|\cdot,x_1,\cdot)\}$, and the marginal sender-receiver AVC is denoted by $\mathcal{W}(x_1) = \{W_{Y|X,X_1,S}(\cdot|\cdot,x_1,\cdot)$. A sufficient condition, under which the deterministic code capacity is the same as the random code capacity of the AVRC, is given in Lemma 2. This condition is also a sufficient condition for positive capacity, but as explained in Remark 4, it is not a necessary condition.

Lemma 2 gives a condition under which the deterministic code capacity is the same as the random code capacity. The condition is given in terms of the marginal AVCs $\mathcal{W}_1(x_1)$ and $\mathcal{W}(x_1)$.

Lemma 2. *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* $\mathcal{L}$ *is positive, and it coincides with the random code capacity, i.e.,* $\mathbb{C}(\mathcal{L}) = \mathbb{C}^\star(\mathcal{L}) > 0$.

The proof of Lemma 2 is given in Appendix E, extending Ahlswede's Elimination Technique [55].

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 [56,57] (Definition 2). Our condition here is given in terms of an extended definition of symmetrizability, akin to [67] (Definition 3.2).

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* $J(s|x)$*,*

$$\sum_{s\in\mathcal{S}} W_{Y,Y_1|X,X_1,S}(y,y_1|x,x_1,s)J(s|\tilde{x}) = \sum_{s\in\mathcal{S}} W_{Y,Y_1|X,X_1,S}(y,y_1|\tilde{x},x_1,s)J(s|x)\,,$$
$$\forall\, x,\tilde{x}\in\mathcal{X}\,,\ x_1\in\mathcal{X}_1\,,\ y\in\mathcal{Y}\,,\ y_1\in\mathcal{Y}_1\,. \quad (24)$$

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

A similar definition applies to the marginals $W_{Y|X,X_1,S}$ and $W_{Y_1|X,X_1,S}$. Note that symmetrizability of each of these marginals can be checked, without reference to whether the channel is degraded or strongly degraded.

Corollary 4. *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* $\mathbb{C}(\mathcal{L}) = \mathbb{C}^\star(\mathcal{L}) > 0$*. In this case,*

$$\mathsf{R}^\star_{PDF}(\mathcal{L}) \le \mathbb{C}(\mathcal{L}) \le \mathsf{R}^\star_{CS}(\mathcal{L})\,. \quad (25)$$

2. *If* $W_{Y,Y_1|X,X_1,S}$ *is strongly reversely degraded, where* $W_{Y_1|X,X_1,S}$ *is non-symmetrizable-*$\mathcal{X}|\mathcal{X}_1$*, then*

$$\mathbb{C}(\mathcal{L}) = \mathbb{C}^\star(\mathcal{L}) = \mathsf{R}^\star_{PDF}(\mathcal{L}) = \mathsf{R}^\star_{CS}(\mathcal{L}) = \min_{q(s)}\max_{p(x,x_1)} I_q(X;Y|X_1)\,. \quad (26)$$

3. *If* $W_{Y,Y_1|X,X_1,S}$ *is strongly degraded, 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*

$$\mathbb{C}(\mathcal{L}) = \mathbb{C}^{\star}(\mathcal{L}) = \mathsf{R}^{\star}_{PDF}(\mathcal{L}) = \mathsf{R}^{\star}_{CS}(\mathcal{L}) = \max_{p(x,x_1)} \min \left\{ \min_{q(s)} I_q(X, X_1; Y)\,,\ I(X; Y_1|X_1) \right\}. \quad (27)$$

The proof of Corollary 4 is given in Appendix F.

Remark 4. *By Corollary 4, we have that non-symmetrizability of the marginal AVCs,* $\mathcal{W}_1(x_{1,1})$ *and* $\mathcal{W}(x_{1,2})$, *for some* $x_{1,1}, x_{1,2} \in \mathcal{X}_1$, *is a sufficient condition for positive capacity (see Figure 3). This raises the question whether it is a necessary condition as well. In other words: If* $\mathcal{W}_1(x_1)$ *and* $\mathcal{W}(x_1)$ *are symmetrizable for all* $x_1 \in \mathcal{X}_1$, *does that necessarily imply that the capacity is zero? The answer is no. We show this using a very simple example. Suppose that* $Y_1 = S$ *and* $Y = (X_1, X + S)$, *where all variables are binary. It is readily seen that for both* Y_1 *and* Y, *the input and the state are symmetric, for every given* $X_1 = x_1$. *Hence,* $\mathcal{W}_1(x_1)$ *and* $\mathcal{W}(x_1)$ *are symmetrizable for all* $x_1 \in \mathcal{X}_1$. *Nevertheless, we note that since the relay can send* $X_1 = Y_1 = S$, *this is equivalent to an AVC with state information at the decoder. As the decoder can use* X_1 *to eliminate the state, the capacity of this AVRC is* $\mathbb{C}(\mathcal{L}) = 1$. *In Lemma 3 below, we give a stronger condition which is a necessary condition for positive capacity.*

Remark 5. *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, if the AVRC is symmetrizable-$\mathcal{X}|\mathcal{X}_1$, then it is a poor channel. For example, say $Y_1 = X + X_1 + S$ and $Y = X \cdot X_1 \cdot S$, with $\mathcal{S} = \mathcal{X}$. Then, the jammer can confuse the decoder by taking the state sequence S^n to be some codeword. The following lemma validates this intuition.

Lemma 3. *If the AVRC* $\mathcal{L}$ *is symmetrizable-*$\mathcal{X}|\mathcal{X}_1$, *then it has zero capacity, i.e.,* $\mathbb{C}(\mathcal{L}) = 0$. *Equivalently, non-symmetrizability-*$\mathcal{X}|\mathcal{X}_1$ *of the AVRC* $\mathcal{L}$ *is a necessary condition for positive capacity.*

Lemma 3 is proved in Appendix G, using an extended version of Ericson's technique [56]. For a strongly degraded AVRC, 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 strongly 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* $J(s|x_1,y_1)$,

$$\sum_{s\in\mathcal{S}} W_{Y|Y_1,X_1,S}(y|y_1,x_1,s) J(s|\tilde{x}_1,\tilde{y}_1) = \sum_{s\in\mathcal{S}} W_{Y|Y_1,X_1,S}(y|\tilde{y}_1,\tilde{x}_1,s) J(s|x_1,y_1)\,,$$
$$\forall\, \tilde{x}_1, x_1 \in \mathcal{X}_1\,,\ y \in \mathcal{Y}\,,\ y_1, \tilde{y}_1 \in \mathcal{Y}_1\,. \quad (28)$$

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

Lemma 4. *If the AVRC $\mathcal{L}$ is strongly degraded and symmetrizable-$\mathcal{X}_1 \times \mathcal{Y}_1$, then it has zero capacity, i.e., $\mathbb{C}(\mathcal{L}) = 0$.*

Lemma 4 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*

$$\begin{aligned} Y_1 &= X + Z \mod 2\,, \\ Y &= X_1 + S\,, \end{aligned}$$

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 $Z \sim$ Bernoulli(θ), $0 \le \theta \le 1$. It is readily seen that $W_{Y,Y_1|X,X_1,S}$ is strongly degraded and symmetrizable-$\mathcal{X}_1 \times \mathcal{Y}_1$, by (2) and (28). In particular, (28) 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 4, 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 3. The derivation of the random code capacity is given in Appendix I.

3.3. 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 [5]; [3] (Section 16.6.2), where $X = (X', X'')$ and

$$W_{Y,Y_1|X',X'',X_1,S}(y,y_1|x',x'',x_1,s) = W_{Y|X',X_1,S}(y|x',x_1,s) \cdot W_{Y_1|X'',X_1,S}(y_1|x'',x_1,s)\,. \tag{29}$$

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 5. *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*

$$\mathbb{C}^\star(\mathcal{L}) = \mathsf{R}^\star_{PDF}(\mathcal{L}) = \mathsf{R}^\star_{CS}(\mathcal{L}) = \max_{p(x_1)p(x'|x_1)p(x''|x_1)} \min\left\{I(X',X_1;Y)\,,\ \min_{q(s)} I_q(X'';Y_1|X_1) + I(X';Y|X_1)\right\}. \tag{30}$$

If $W_{Y_1|X'',X_1,S}$ is non-symmetrizable-$\mathcal{X}''|\mathcal{X}_1$, and $W_{Y|X',X_1}(y|x',x_1) \neq W_{Y|X',X_1}(y|\tilde{x}',x_1)$ 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 $\mathbb{C}(\mathcal{L}) = \mathsf{R}^\star_{PDF}(\mathcal{L}) = \mathsf{R}^\star_{CS}(\mathcal{L})$.

The proof of Lemma 5 is given in Appendix J. To prove Lemma 5, we apply the methods of [5] to our results. Specifically, we use the partial decode-forward lower bound in Theorem 1, taking $U = X''$ (see (9) and (19)).

4. Gaussian AVRC with Sender Frequency Division

We give extended results for the Gaussian AVRC with sender frequency division (SFD), which is a special case of the AVRC with orthogonal sender components [5]. We determine the random code capacity of the Gaussian AVRC with SFD, and give lower and upper bounds on the deterministic code capacity. The derivation of the deterministic code bounds is mostly independent of our previous results, and it is based on the technique by [87]. The Gaussian relay channel $W_{Y,Y_1|X,X_1,S}$ with SFD is a special case of a relay channel with orthogonal sender components [5], specified by

$$\begin{aligned} Y_1 &= X'' + Z\,, \\ Y &= X' + X_1 + S\,, \end{aligned} \tag{31}$$

where the Gaussian additive noise $Z \sim \mathcal{N}(0,\sigma^2)$ is independent of the channel state. As opposed to Lemma 5, the main channel here depends on the state, while the channel to the relay does not.

In the case of a Gaussian channel, power limitations need to be accounted for, and thus, we consider the Gaussian relay channel under input and state constraints. Specifically, the user and the relay's transmission are subject to input constraints $\Omega > 0$ and $\Omega_1 > 0$, respectively, and the jammer is under a state constraint Λ, i.e.,

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}(X_i'^2 + X_i''^2) \le \Omega\,, \quad \frac{1}{n}\sum_{i=1}^{n} X_{1,i}^2 \le \Omega_1 \quad \text{w.p. } 1\,, \\ &\frac{1}{n}\sum_{i=1}^{n} S_i^2 \le \Lambda \quad \text{w.p. } 1\,. \end{aligned} \tag{32}$$

We note that Ahlswede's Elimination Technique cannot be used under a state constraint (see [57]). Indeed, if the jammer concentrates a lot of power on the shared randomness transmission, then this transmission needs to be robust against a state constraint that is higher than Λ. Thereby, the results given in Section 3.2.2 do not apply to the Gaussian AVRC under input and state constraints.

For the compound relay channel, the state constraint is in the average sense. That is, we say that the Gaussian compound relay channel $\mathcal{L}^{\mathcal{Q}}$ with SFD is under input constraints Ω and Ω_1 and state constraint Λ if

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}(X_i'^2 + X_i''^2) \le \Omega\,, \quad \frac{1}{n}\sum_{i=1}^{n} X_{1,i}^2 \le \Omega_1\,, \quad \text{w.p. } 1\,, \\ &\mathcal{Q} = \{q(s) \,:\, \mathbb{E}S^2 \le \Lambda\}\,. \end{aligned} \tag{33}$$

Coding definitions and notation are as follows. The definition of a code is similar to that of Section 2.3. The encoding function is denoted by $\mathbf{f} = (\mathbf{f}', \mathbf{f}'')$, with $\mathbf{f}' : [1 : 2^{nR}] \to \mathbb{R}^n$ and $\mathbf{f}'' : [1 : 2^{nR}] \to \mathbb{R}^n$, and the relay encoding function is denoted by $\mathbf{f}_1 : \mathbb{R}^n \to \mathbb{R}^n$, where $\mathrm{f}_{1,i} : \mathbb{R}^{i-1} \to \mathbb{R}$, for $i \in [1 : n]$. The boldface notation indicates that the encoding functions produce sequences. Here, the encoder and the relay satisfy the input constraints $\|\mathbf{f}'(m)\|^2 + \|\mathbf{f}''(m)\|^2 \le n\Omega$ and $\|\mathbf{f}_1(\mathbf{y}_1)\|^2 \le n\Omega_1$ for all $m \in [1 : 2^{nR}]$ and $\mathbf{y}_1 \in \mathbb{R}^n$. At time $i \in [1 : n]$, given a message $m \in [1 : 2^{nR}]$, the encoder transmits $(x_i', x_i'') = (\mathrm{f}_i'(m), \mathrm{f}_i''(m))$, and the relay transmits $x_{1,i} = \mathrm{f}_{1,i}(y_{1,1}, \ldots, y_{1,i-1})$. The decoder receives the output sequence $\mathbf{y}$, and finds an estimate $\hat{m} = g(\mathbf{y})$. A $(2^{nR}, n, \varepsilon)$ code $\mathscr{C}$ for the Gaussian AVRC satisfies $P_{e|\mathbf{s}}^{(n)}(\mathscr{C}) \le \varepsilon$, for all $\mathbf{s} \in \mathbb{R}^n$ with $\|\mathbf{s}\|^2 \le n\Lambda$, where

$$P_{e|\mathbf{s}}^{(n)}(\mathscr{C}) = \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}} \int_{\mathcal{D}(m,\mathbf{s})^c} \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\|\mathbf{z}\|^2/2\sigma^2}\, d\mathbf{z}\,, \tag{34}$$

with

$$\mathcal{D}(m, \mathbf{s}) = \left\{\mathbf{z} \in \mathbb{R}^n : g\big(\,\mathbf{f}'(m) + \mathbf{f}_1\big(\mathbf{f}''(m) + \mathbf{z}\big) + \mathbf{s}\,\big) = m\right\}\,. \tag{35}$$

Achivable rates, deterministic code capacity and random code capacity are defined as before. Next, we give our results on the Gaussian compound relay channel and the Gaussian AVRC with SFD.

5. Main Results—Gaussian AVRC with SFD

We give our results on the Gaussian compound and AVRC with SFD. The results on this compound relay channel and on the random code capacity of this AVRC are obtained through a straightforward extension of our previous results and derivations. However, the derivation of the deterministic code bounds is mostly independent of our previous results, and it is based on modifying the technique by Csiszär and Narayan in their paper on the Gaussian AVC [87].

5.1. Gaussian Compound Relay Channel

We determine the capacity of the Gaussian compound relay channel with SFD under input and state constraints. Let

$$\mathsf{F}_G(\alpha,\rho) \triangleq \min\left\{\frac{1}{2}\log\left(1+\frac{\Omega_1+\alpha\Omega+2\rho\sqrt{\alpha\Omega}\sqrt{\Omega_1}}{\Lambda}\right),\right.$$
$$\left.\frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\sigma^2}\right)+\frac{1}{2}\log\left(1+\frac{(1-\rho^2)\alpha\Omega}{\Lambda}\right)\right\}. \quad (36)$$

Lemma 6. *The capacity of the Gaussian compound relay channel with SFD, under input constraints Ω and Ω_1 and state constraint Λ, is given by*

$$\mathbb{C}(\mathcal{L}^{\mathcal{Q}}) = \max_{0\le\alpha,\rho\le 1} \mathsf{F}_G(\alpha,\rho)\,, \quad (37)$$

and it is identical to the random code capacity, i.e., $\mathbb{C}(\mathcal{L}^{\mathcal{Q}}) = \mathbb{C}^{\star}(\mathcal{L}^{\mathcal{Q}})$.

The proof of Lemma 6 is given in Appendix K, based on our results in the previous sections. The parameter $0 \le \alpha \le 1$ represents the fraction of input power invested in the transmission of the message component which is decoded by the relay, in the partial decode-forward coding scheme. Specifically, in the achievability proof in [5], $\alpha\Omega$ and $(1-\alpha)\Omega$ are the variances of X' and X'', respectively. The parameter ρ stands for the correlation coefficient between the decode-forward transmission X' and the relay transmission X_1.

5.2. Gaussian AVRC

We determine the random code capacity of the Gaussian AVRC with SFD under constraints.

Theorem 2. *The random code capacity of the Gaussian AVRC with SFD, under input constraints Ω and Ω_1 and state constraint Λ, is given by*

$$\mathbb{C}^{\star}(\mathcal{L}) = \mathbb{C}(\mathcal{L}^{\mathcal{Q}}) = \max_{0\le\alpha,\rho\le 1} \mathsf{F}_G(\alpha,\rho)\,. \quad (38)$$

The proof of Theorem 2 is given in Appendix L. The proof follows the same considerations as in our previous results.

Next, we give lower and upper bounds on the deterministic code capacity of the Gaussian AVRC with SFD under constraints, obtained by generalizing the non-standard techniques by Csiszár and Narayan in their 1991 paper on the Gaussian AVC [87]. Define

$$\begin{array}{lll} \mathsf{R}_{G,low}(\mathcal{L}) \triangleq & \max & \mathsf{F}_G(\alpha,\rho) \\ & \text{subject to} & 0\le\alpha,\rho\le 1\,, \\ & & (1-\rho^2)\alpha\Omega > \Lambda\,, \\ & & \frac{\Omega_1}{\Omega}(\sqrt{\Omega_1}+\rho\sqrt{\alpha\Omega})^2 > \Lambda+(1-\rho^2)\alpha\Omega\,. \end{array} \quad (39)$$

and

$$\begin{array}{lll} \mathsf{R}_{G,up}(\mathcal{L}) \triangleq & \max & \mathsf{F}_G(\alpha,\rho) \\ & \text{subject to} & 0\le\alpha,\rho\le 1\,, \\ & & \Omega_1+\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1} \ge \Lambda\,. \end{array} \quad (40)$$

It can be seen that $\mathsf{R}_{G,low} \leq \mathsf{R}_{G,up}$, since

$$\Omega_1 + \alpha\Omega + 2\rho\sqrt{\alpha\Omega \cdot \Omega_1} = (1-\rho^2)\alpha\Omega + (\sqrt{\Omega_1} + \rho\sqrt{\alpha\Omega})^2 \geq (1-\rho^2)\alpha\Omega\,. \tag{41}$$

The analysis is based on the following lemma by [87].

Lemma 7 (see [87] (Lemma 1))**.** *For every $\varepsilon > 0$, $8\sqrt{\varepsilon} < \eta < 1$, $K > 2\varepsilon$, and $\mathsf{M} = 2^{nR}$, with $2\varepsilon \leq R \leq K$, and $n \geq n_0(\varepsilon, \eta, K)$, there exist M unit vectors $\mathbf{a}(m) \in \mathbb{R}^n$, $m \in [1:\mathsf{M}]$, such that for every unit vector $\mathbf{c} \in \mathbb{R}^n$ and $0 \leq \theta, \zeta \leq 1$,*

$$\left|\left\{\widetilde{m} \in [1:\mathsf{M}] \,:\, \langle \mathbf{a}(\widetilde{m}), \mathbf{c}\rangle \geq \theta\right\}\right| \leq 2^{n\left([R+\frac{1}{2}\log(1-\theta^2)]_+ + \varepsilon\right)}\,, \tag{42}$$

and if $\theta \geq \eta$ and $\theta^2 + \zeta^2 > 1 + \eta - 2^{-2R}$, then

$$\frac{1}{\mathsf{M}}\left|\left\{m \in [1:\mathsf{M}] \,:\, |\langle \mathbf{a}(\widetilde{m}), \mathbf{a}(m)\rangle| \geq \theta\,,\ |\langle \mathbf{a}(\widetilde{m}), \mathbf{c}\rangle| \geq \zeta\,,\ \text{for some } \widetilde{m} \neq m\right\}\right| \leq 2^{-n\varepsilon}\,, \tag{43}$$

where $[t]_+ = \max\{0, t\}$ and $\langle \cdot, \cdot \rangle$ denotes inner product.

Intuitively, the lemma states that under certain conditions, a codebook can be constructed with an exponentially small fraction of "bad" messages, for which the codewords are non-orthogonal to each other and the state sequence.

Theorem 3. *The deterministic code capacity of the Gaussian AVRC with SFD, under input constraints Ω and Ω_1 and state constraint Λ, is bounded by*

$$\mathsf{R}_{G,low}(\mathcal{L}) \leq \mathbb{C}(\mathcal{L}) \leq \mathsf{R}_{G,up}(\mathcal{L})\,. \tag{44}$$

The proof of Theorem 3 is given in Appendix M.

Remark 6. *Csiszár and Narayan [87] have shown that for the classical Gaussian AVC, reliable decoding is guaranteed when the input constraint Ω is larger than the state constraint Λ. Here, we use a partial decode-forward coding scheme, where the message has two components, one which is decoded by the relay, and the other is transmitted directly. The respective optimization constraints $\frac{\Omega_1}{\Omega}(\sqrt{\Omega_1} + \rho\sqrt{\alpha\Omega})^2 > \Lambda + (1-\rho^2)\alpha\Omega$ and $(1-\rho^2)\alpha\Omega > \Lambda$ in the RHS of (39), guarantee reliability for each decoding step.*

Remark 7. *Csiszár and Narayan [87] have further shown that for the classical Gaussian AVC, if $\Omega \leq \Lambda$, the capacity is zero. The converse proof in [87] follows by considering a jammer who chooses the state sequence to be a codeword. Due to the symmetry between $\mathbf{X}$ and $\mathbf{S}$, the decoder cannot distinguish between the transmitted codeword and the impostor sent by the jammer. Here, we consider a jammer who simulates $\mathbf{X}' + \mathbf{X}_1$. Specifically, The jammer draws a codeword $\mathbf{X}' = \mathbf{f}'(\widetilde{m})$ uniformly at random, and then, generates a sequence $\widetilde{\mathbf{Y}}_1$ distributed according to the conditional distribution $P_{\mathbf{Y}_1|M=\widetilde{m}}$. If the sequence $\widetilde{\mathbf{S}} = \mathbf{f}'(\widetilde{m}) + \mathbf{f}_1(\widetilde{\mathbf{Y}}_1)$ satisfies the state constraint Λ, then the jammer chooses $\widetilde{\mathbf{S}}$ as the state sequence. Defining $\alpha\Omega$, Ω_1, and ρ as the empirical decode-forward transmission power, relay transmission power, and their correlation coefficient, respectively, we have that the state constraint $\left\|\widetilde{\mathbf{S}}\right\|^2 \leq n\Lambda$ holds with high probability, if $\Omega_1 + \alpha\Omega + 2\rho\sqrt{\alpha\Omega \cdot \Omega_1} < \Lambda$. The details are in Appendix M.*

Figure 4 depicts the bounds on the capacity of the Gaussian AVRC with SFD under input and state constraints, as a function of the input constraint $\Omega = \Omega_1$, under state constraint $\Lambda = 1$ and $\sigma^2 = 0.5$. The top dashed line depicts the random code capacity of the Gaussian AVRC. The solid lines depict the deterministic code lower and upper bounds $\mathsf{R}_{G,low}(\mathcal{L})$ and $\mathsf{R}_{G,up}(\mathcal{L})$. For low values, $\Omega < \frac{\Lambda}{4} = 0.25$, we have that $\mathsf{R}_{G,up}(\mathcal{L}) = 0$, hence the deterministic code capacity is zero, and it is strictly lower than

the random code capacity. The dotted lower line depicts the direct transmission lower bound, which is $\mathsf{F}_G(1,0)$ for $\Omega > \Lambda$, and zero otherwise [57]. For intermediate values of Ω, direct transmission is better than the lower bound in Theorem 3. Whereas, for high values of Ω, the optimization constraints in (39) and (40) are inactive, hence, our bounds are tight, and the capacity coincides with the random code capacity, i.e., $\mathbb{C}(\mathcal{L}) = \mathbb{C}^{\star}(\mathcal{L}) = \mathsf{R}_{G,low}(\mathcal{L}) = \mathsf{R}_{G,up}(\mathcal{L})$.

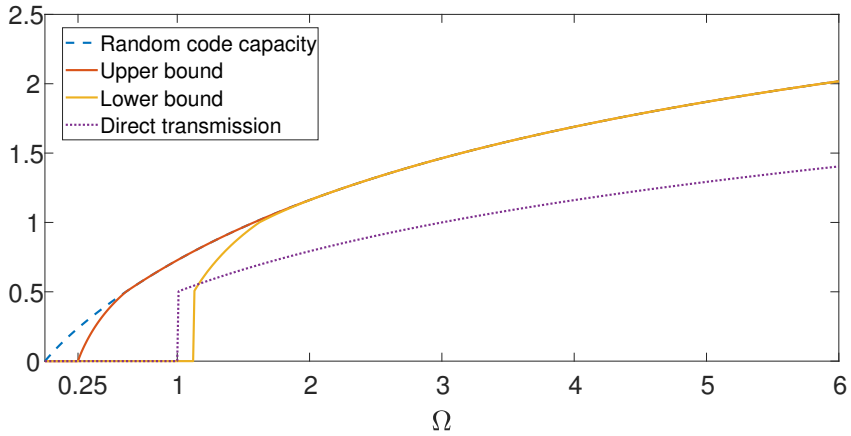

Figure 4. Bounds on the capacity of the Gaussian AVRC with sender frequency division. The dashed upper line depicts the random code capacity of the Gaussian AVRC as a function of the input constraint $\Omega = \Omega_1$, under state constraint $\Lambda = 1$ and $\sigma^2 = 0.5$. The solid lines depict the deterministic code lower and upper bounds $\mathsf{R}_{G,low}(\mathcal{L})$ and $\mathsf{R}_{G,up}(\mathcal{L})$. The dotted lower line depicts the direct transmission lower bound.

6. The Primitive AVRC

In this section, we give our results on the primitive AVRC [2], and then consider the Gaussian case. Part of the motivation given in [2] to consider the primitive relay channel was that the overall behavior and properties are the same as the non primitive ("regular") relay channel. We show that this is not true in the arbitrarily varying scenario. In particular, the behavior of the primitive Gaussian AVRC with SFD is different compared to the non-primitive counterpart considered above.

6.1. Definitions and Notation

Consider a setup where the sender transmits information over state-dependent memoryless relay channel $W_{Y,Y_1|X,S}$, while there is a noiseless link of capacity $C_1 > 0$ between the relay and the receiver. Communication over a primitive relay channel is depicted in Figure 5. Given a message $M \in [1 : 2^{nR}]$, the encoder transmits $X^n = f(M)$ over the channel $W_{Y,Y_1|X,S}$, which is referred to as the primitive relay channel. The relay receives Y_1^n and sends an index $L = f_1(Y_1^n)$ to the receiver, where $f_1 : \mathcal{Y}_1^n \to [1 : 2^{nC_1}]$. The decoder receives both the channel output sequence Y^n and the relay output L, and finds an estimate of the message $\hat{M} = g(Y^n, L)$. In accordance with the previous definitions, the primitive AVRC $\mathcal{L}_{\text{prim}} = \{W_{Y,Y_1|X,S}\}$ has a state sequence of unknown distribution, not necessarily independent nor stationary. The deterministic code capacity and the random code capacity are defined as before, and denoted by $\mathbb{C}(\mathcal{L}_{\text{prim}})$ and $\mathbb{C}^{\star}(\mathcal{L}_{\text{prim}})$, respectively.

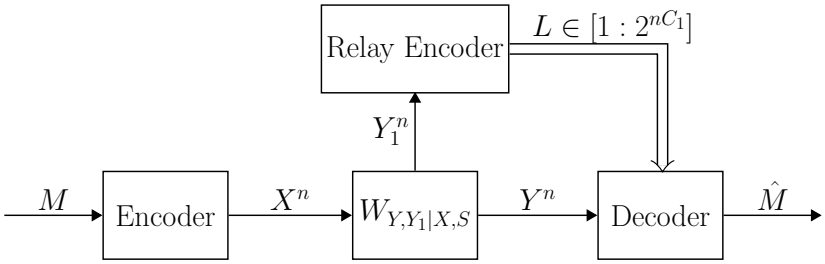

Figure 5. Communication over the primitive AVRC $\mathcal{L}$. Given a message M, the encoder transmits $X^n = f(M)$. The relay receives Y_1^n and sends $L = f_1(Y_1^n)$, where $f_1 : \mathcal{Y}_1^n \to [1 : 2^{nC_1}]$. The decoder receives both the channel output sequence Y^n and the relay output L, and finds an estimate of the message $\hat{M} = g(Y^n, L)$.

6.2. Main Results—Primitive AVRC

We give our results on the primitive AVRC below. However, since the proofs are based on the same arguments as given for the non primitive AVRC, we omit the proofs of the results in this section. The details are given in [91].

Using similar arguments to those given for the non primitive relay channel, we obtain the following bounds on the random code capacity,

$$\mathsf{R}_{CS}^{\star} \triangleq \min_{q(s)} \max_{p(x)} \min \left\{ I_q(X;Y) + C_1 \,,\; I_q(X;Y,Y_1) \right\} , \tag{45}$$

and

$$\mathsf{R}_{PDF}^{\star} \triangleq \max_{p(u,x)} \min \left\{ \min_{q(s)} I_q(U;Y) + \min_{q(s)} I_q(X;Y|U) + C_1 \,,\; \min_{q(s)} I_q(U;Y_1) + \min_{q(s)} I_q(X;Y|U) \right\} . \tag{46}$$

Theorem 4. *The random code capacity of a primitive AVRC $\mathcal{L}_{prim}$ is bounded by*

$$\mathsf{R}_{PDF}^{\star} \le \mathbb{C}^{\star}(\mathcal{L}_{prim}) \le \mathsf{R}_{CS}^{\star} . \tag{47}$$

Those bounds have the same form as the cutset upper bound and the partical decode-forward lower bound in Section 3 (cf. (8), (9) and (45), (46)). As in Section 3, we can use the bounds above to determine the capacity in the strongly degraded and reversely degraded cases, based on the direct transmission lower bound (for $U = \emptyset$), and the full decode-forward lower bound (for $U = X$).

Corollary 5. *Let $\mathcal{L}_{prim}$ be a primitive AVRC.*

1. *If $W_{Y,Y_1|X,S}$ is strongly reversely degraded, i.e., $W_{Y,Y_1|X,S} = W_{Y|X,S} W_{Y_1|Y}$, then*

$$\mathbb{C}^{\star}(\mathcal{L}_{prim}) = \min_{q(s)} \max_{p(x)} I_q(X;Y) . \tag{48}$$

2. *If $W_{Y,Y_1|X,X_1,S}$ is strongly degraded, i.e., $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X} W_{Y|Y_1,S}$, then*

$$\mathbb{C}^{\star}(\mathcal{L}_{prim}) = \max_{p(x)} \min \left\{ \min_{q(s)} I_q(X;Y) + C_1 \,,\; I(X;Y_1) \right\} . \tag{49}$$

As for the deterministic code capacity, we give the following theorem.

Theorem 5. *Let $\mathcal{L}_{prim}$ be a primitive AVRC.*

1. *If $W_{Y_1|X,S}$ is non-symmetrizable, then $\mathbb{C}(\mathcal{L}_{prim}) = \mathbb{C}^{\star}(\mathcal{L}_{prim})$. In this case,*

$$\mathsf{R}^{\star}_{PDF} \leq \mathbb{C}(\mathcal{L}_{prim}) \leq \mathsf{R}^{\star}_{CS} \,. \tag{50}$$

2. *If $W_{Y,Y_1|X,S}$ is strongly reversely degraded, where $W_{Y_1|X,S}$ is non-symmetrizable, then*

$$\mathbb{C}(\mathcal{L}_{prim}) = \min_{q(s)} \max_{p(x)} I_q(X;Y) \,. \tag{51}$$

3. *If $W_{Y,Y_1|X,S}$ is strongly degraded, such that $W_{Y_1|X}(y_1|x) \neq W_{Y_1|X}(y_1|\tilde{x})$ for some $x, \tilde{x} \in \mathcal{X}$, $y_1 \in \mathcal{Y}_1$, then*

$$\mathbb{C}(\mathcal{L}_{prim}) = \max_{p(x)} \min \left\{ \min_{q(s)} I_q(X;Y) + C_1 \,,\; I(X;Y_1) \right\} . \tag{52}$$

4. *If $W_{\widetilde{Y}|X,S}$ is symmetrizable, where $\widetilde{Y} = (Y, Y_1)$, then $\mathbb{C}(\mathcal{L}_{prim}) = 0$.*

The proof of Theorem 5 is available in [91]. To illustrate our results, we give the following example of a primitive AVRC.

Example 2. *Consider a state-dependent primitive relay channel $W_{Y,Y_1|X,S}$, specified by*

$$\begin{aligned} Y_1 &= X(1-S) \,, \\ Y &= X + S \,, \end{aligned}$$

where $\mathcal{X} = \mathcal{S} = \mathcal{Y}_1 = \{0,1\}$, $\mathcal{Y} = \{0,1,2\}$, and $C_1 = 1$, i.e., the link between the relay and the receiver is a noiseless bit pipe. It can be seen that both the sender-relay and the sender-receiver marginals are symmetrizable. Indeed, $W_{Y|X,S}$ satisfies

$$\sum_{s \in \mathcal{S}} W_{Y|X,S}(y_1|x,s) J(s|\tilde{x}) = \sum_{s \in \mathcal{S}} W_{Y|X,S}(y_1|\tilde{x},s) J(s|x) \,, \; x, \tilde{x} \in \mathcal{X} \,, \; y \in \mathcal{Y} \,, \tag{53}$$

with $J(s|x) = 1$ for $s = x$, and $J(s|x) = 0$ otherwise, while $W_{Y_1|X,S}$ satisfies (53) with $J(s|x) = 1$ for $s = 1 - x$, and $J(s|x) = 0$ otherwise. Nevertheless, the capacity of the primitive AVRC $\mathcal{L}_{prim} = \{W_{Y,Y_1|X,S}\}$ is $\mathbb{C}(\mathcal{L}_{prim}) = 1$, which can be achieved using a code of length $n = 1$, with $f(m) = m$, $f_1(y_1) = y_1$,

$$g(y, \ell) = g(y, y_1) = \begin{cases} 0 & y = 0 \\ 1 & y = 2 \\ y_1 & y = 1 \end{cases} \tag{54}$$

for $m, y_1 \in \{0,1\}$ and $y \in \{0,1,2\}$. This example shows that even if the sender-relay and sender-receiver marginals are symmetrizable, the capacity may still be positive. We further note that the condition in part 4 of Theorem 5 implies that $W_{Y|X,S}$ and $W_{Y_1|X,S}$ are both symmetrizable, but not vice versa, as shown by this example. That is, as the capacity is positive, we have that $W_{\widetilde{Y}|X,S}$ is non-symmetrizable, where $\widetilde{Y} = (Y, Y_1)$, despite the fact that the marginals $W_{Y|X,S}$ and $W_{Y_1|X,S}$ are both symmetrizable.

6.3. Primitive Gaussian AVRC

Consider the primitive Gaussian relay channel with SFD,

$$\begin{aligned} Y_1 &= X'' + Z \,, \\ Y &= X' + S \,, \end{aligned} \tag{55}$$

Suppose that input and state constraints are imposed as before, i.e., $\frac{1}{n}\sum_{i=1}^{n}(X_i'^2 + X_i''^2) \leq \Omega$ and $\frac{1}{n}\sum_{i=1}^{n} S_i^2 \leq \Lambda$ with probability 1. The capacity of the primitive Gaussian AVRC with SFD, under input constraint Ω and state constraint Λ is given by

$$\mathbb{C}(\mathcal{L}_{\text{prim}}) = \mathbb{C}^{\star}(\mathcal{L}_{\text{prim}}) = \max_{0\leq\alpha\leq 1}\left[\frac{1}{2}\log\left(1+\frac{\alpha\Omega}{\Lambda}\right) + \min\left\{C_1, \frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\Lambda}\right)\right\}\right]. \tag{56}$$

This result is due to the following. Observe that one could treat this primitive AVRC as two independent channels, one from X' to Y and the other from X'' to Y_1, dividing the input power to $\alpha\Omega$ and $(1-\alpha)\Omega$, respectively. Based on this observation, the random code direct part follows from [92]. Next, the deterministic code direct part follows from part 1 of Theorem 5, and the converse part follows straightforwardly from the cutset upper bound in Theorem 4.

7. Discussion

We have presented the model of the arbitrarily varying relay channel (AVRC), as a state dependent relay channel, where jamming attacks result in either a random or a deterministic state sequence, $S^n \sim q(s^n)$, where the joint distribution $q(s^n)$ is unknown and it is not necessarily of a product form. We have established the cutset upper bound and the partial decode-forward lower bound on the random code capacity of the AVRC. We have determined the random code capacity in special cases of the degraded AVRC, the reversely degraded AVRC, and the AVRC with orthogonal sender components. To do so, we used the direct transmission lower bound and the full decode-forward lower bound, along with quasi-convexity properties which are required in order to use the minimax theorem.

We have provided generalized symmetrizability conditions under which the deterministic code capacity coincides with the random code capacity. Specifically, we have shown that if the sender-relay and sender-receiver marginals are non-symmetrizable for a given relay transmission, then the capacity is positive. We further noted that this is a sufficient condition for positive capacity, which raises the question whether it is also a necessary condition. In other words, if those marginals are symmetrizable for every given relay transmission, does that necessarily imply that the capacity is zero? The answer is no, and we have refuted this assertion using a simple example, where the relay acts as a source of state information to the receiver. Then, we provided a stronger symmetrizability condition, which is necessary for the capacity to be positive. We have shown by example that the deterministic code capacity can be strictly lower than the random code capacity of the AVRC.

The Gaussian AVRC with sender frequency division (SFD) under input and state constraints is also addressed in this paper. The random code capacity is determined using the above results, whereas the deterministic code capacity is lower and upper bounded using an independent approach. Specifically, we extended the technique by Csiszár and Narayan in their 1991 paper on the Gaussian AVC [87]. We have shown that the deterministic code capacity can be strictly lower than the random code capacity, for low values on the input constraint.

Furthermore, we have considered the primitive AVRC, where there is a noiseless link between the relay and the receiver of limited capacity [2]. We tested Kim's assertion that "the primitive relay channel captures most essential features and challenges of relaying, and thus serves as a good testbed for new relay coding techniques" [2]. We have shown that this assertion is not true in the arbitrarily varying scenario. Specifically, for the primitive Gaussian AVRC with SFD, the deterministic code capacity and the random code capacity are always the same, regardless of the value of the input constraint (see (56)), in contrast to our findings for the non primitive case, as demonstrated in Figure 4.

Author Contributions: Formal analysis, U.P.; Investigation, U.P.; Methodology, U.P.; Supervision, Y.S.; Writing – original draft, U.P.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AVC	Arbitrarily varying channel
AVRC	Arbitrarily varying relay channel
DMC	Discrete memoryless channel
pmf	probability mass function
RT	Robustification technique
SFD	Sender frequency division
Eq.	Equation
RHS	Right hand side
LHS	Left hand side

Appendix A. Proof of Lemma 1

Appendix A.1. Partial Decode-Forward Lower Bound

We construct a block Markov code using the partial decode-forward scheme. That is, the encoder sends a sequence of messages over multiple blocks. The message in each block consists of two components, a decode-forward component, and a direct transmission component, where only the former is decoded by the relay. Once the decoder has received all blocks, the decode-forward components are decoded backwards, i.e., starting with the message in the last block going backwards. Using the estimation of the decode-forward components, the direct transmission components are decoded forwards, i.e., starting with the message in the first block going forwards. The ambiguity of the state distribution needs to be treated throughout all of those estimations. Hence, we use 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

$$\hat{\mathcal{Q}}_n = \{\hat{P}_{s^n} \,:\, s^n \in \mathcal{A}^{\delta_1}(q) \;\text{ for some } q \in \mathcal{Q}\}\,, \tag{A1}$$

where

$$\delta_1 \triangleq \frac{\delta}{2 \cdot |\mathcal{S}|}\,. \tag{A2}$$

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.,

$$M'_b \sim \text{Unif}[1 : 2^{nR'}]\,,\; M''_b \sim \text{Unif}[1 : 2^{nR''}]\,,\; \text{with } R' + R'' = R\,, \tag{A3}$$

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

$$P^q_{X,Y,Y_1|U,X_1}(x,y,y_1|u,x_1) = P_{X|U,X_1}(x|u,x_1) \sum_{s\in\mathcal{S}} q(s) W_{Y,Y_1|X,X_1,S}(y,y_1|x,x_1,s)\,. \tag{A4}$$

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,

$$u_b^n(m'_b|m'_{b-1}) \sim \prod_{i=1}^n P_{U|X_1}(u_i|x_{1,b,i}(m'_{b-1}))\,,\ m'_b \in [1 : 2^{nR'}]\,, \tag{A5}$$

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

$$x_b^n(m'_b, m''_b|m'_{b-1}) \sim \prod_{i=1}^n P_{X|U,X_1}(x_i|u_{b,i}(m'_b|m'_{b-1}), x_{1,b,i}(m'_{b-1}))\,,\ m''_b \in [1 : 2^{nR''}]\,, \tag{A6}$$

conditionally independent given $(u_b^n(m'_b|m'_{b-1}), x_{1,b}^n(m'_{b-1}))$. We have thus generated $B-2$ independent codebooks,

$$\mathcal{F}_b = \left\{ \left(x_{1,b}^n(m'_{b-1}), u_b^n(m'_b|m'_{b-1}), x_b^n(m'_b, m''_b|m'_{b-1})\right) \,:\, m'_{b-1}, m'_b \in [1 : 2^{nR'}]\,,\ m''_b \in [1 : 2^{nR''}] \right\}, \tag{A7}$$

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 A1.

Encoding: To send the message sequence $(m'_1, m''_1, \ldots, m'_{B-1}, m''_{B-1})$, transmit $x_b^n(m'_b, m''_b|m'_{b-1})$ 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

$$(u_b^n(\widetilde{m}'_b|\widetilde{m}'_{b-1}), x_{1,b}^n(\widetilde{m}'_{b-1}), y_{1,b}^n) \in \mathcal{A}^\delta(P_{U,X_1} P^q_{Y_1|U,X_1})\,,\ \text{for some } q \in \hat{\mathcal{Q}}_n\,. \tag{A8}$$

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

$$(u_{b+1}^n(\hat{m}'_{b+1}|\hat{m}'_b), x_{1,b+1}^n(\hat{m}'_b), y_{b+1}^n) \in \mathcal{A}^\delta(P_{U,X_1} P^q_{Y|U,X_1})\,,\ \text{for some } q \in \hat{\mathcal{Q}}_n\,. \tag{A9}$$

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

$$(u_b^n(\hat{m}'_b|\hat{m}'_{b-1}), x_b^n(\hat{m}'_b, \hat{m}''_b|\hat{m}'_{b-1}), x_{1,b}(\hat{m}'_{b-1}), y_b^n) \in \mathcal{A}^\delta(P_{U,X,X_1} P^q_{Y|X,X_1})\,,\ \text{for some } q \in \hat{\mathcal{Q}}_n\,. \tag{A10}$$

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.

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

Figure A1. The 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.

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

$$\mathcal{E}_1(b) = \{\widetilde{M}_b' \neq 1\}\,,\ \mathcal{E}_2(b) = \{\hat{M}_b' \neq 1\}\,,\ \mathcal{E}_3(b) = \{\hat{M}_b'' \neq 1\}\,,\ \text{for } b \in [1:B-1]\,. \tag{A11}$$

Then, the probability of error is bounded by

$$P_e^{(n)}(q,\mathscr{C}) \leq \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_1(b)\right) + \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_2(b) \mid \mathcal{E}_1^c(b)\right) + \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_3(b) \mid \mathcal{E}_1^c(b) \cap \mathcal{E}_2^c(b) \cap \mathcal{E}_2^c(b-1)\right)\,, \tag{A12}$$

with $\mathcal{E}_2(0) = \varnothing$, 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

$$\begin{aligned}
\mathcal{E}_{1,1}(b) =& \{(U_b^n(1|\widetilde{M}_{b-1}'), X_{1,b}^n(\widetilde{M}_{b-1}'), Y_{1,b}^n) \notin \mathcal{A}^\delta(P_{U,X_1}P_{Y_1|U,X_1}^{q'}) \ \text{ for all } q' \in \hat{\mathcal{Q}}_n\} \\
\mathcal{E}_{1,2}(b) =& \{(U_b^n(m_b'|\widetilde{M}_{b-1}'), X_{1,b}^n(\widetilde{M}_{b-1}'), Y_{1,b}^n) \in \mathcal{A}^\delta(P_{U,X_1}P_{Y_1|U,X_1}^{q'})\,, \text{ for some } m_b' \neq 1, q' \in \hat{\mathcal{Q}}_n\}\,.
\end{aligned} \tag{A13}$$

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

$$\begin{aligned}
\mathcal{E}_1(b) \subseteq& \mathcal{E}_1(b-1) \cup \mathcal{E}_{1,1}(b) \cup \mathcal{E}_{1,2}(b) \\
=& \mathcal{E}_1(b-1) \cup (\mathcal{E}_1(b-1)^c \cap \mathcal{E}_{1,1}(b)) \cup (\mathcal{E}_1(b-1)^c \cap \mathcal{E}_{1,2}(b))\,,
\end{aligned} \tag{A14}$$

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

$$\Pr\left(\mathcal{E}_1(b)\right) \leq \Pr\left(\mathcal{E}_1(b-1)\right) + \Pr\left(\mathcal{E}_{1,1}(b) \mid \mathcal{E}_1(b-1)^c\right) + \Pr\left(\mathcal{E}_{1,2}(b) \mid \mathcal{E}_1(b-1)^c\right)\,. \tag{A15}$$

Consider the second term on the RHS of (A15). 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

$$|q'(s) - q''(s)| \leq \delta_1\,, \tag{A16}$$

for all $s \in \mathcal{S}$, and by the definition in (A1), $q' \in \hat{\mathcal{Q}}_n$. Then, (A16) implies that

$$|P^{q'}_{Y_1|U,X_1}(y_1|u,x_1) - P^{q''}_{Y_1|U,X_1}(y_1|u,x_1)| \leq |\mathcal{S}| \cdot \delta_1 = \frac{\delta}{2}\,, \tag{A17}$$

for all $u \in \mathcal{U}$, $x_1 \in \mathcal{X}_1$ and $y_1 \in \mathcal{Y}_1$ (see (A4) and (A2)). Hence, $(U^n_b(1|1), X^n_{1,b}(1), Y^n_{1,b}) \in \mathcal{A}^\delta(P_{U,X_1}P^{q'}_{Y_1|U,X_1})$, which contradicts the first assumption. It follows that

$$\begin{aligned}
&\Pr\left(\mathcal{E}_{1,1}(b) \mid \mathcal{E}_1(b-1)^c\right)\\
&\leq \Pr\left((U^n_b(1|1), X^n_{1,b}(1), Y^n_{1,b}) \notin \mathcal{A}^{\delta/2}(P_{U,X_1}P^{q''}_{Y_1|U,X_1}) \text{ for all } q'' \in \mathcal{Q} \mid \mathcal{E}_1(b-1)^c\right)\\
&\leq \Pr\left((U^n_b(1|1), X^n_{1,b}(1), Y^n_{1,b}) \notin \mathcal{A}^{\delta/2}(P_{U,X_1}P^{q^*}_{Y_1|U,X_1}) \mid \mathcal{E}_1(b-1)^c\right)\,.
\end{aligned} \tag{A18}$$

Since the codebooks $\mathcal{F}_1, \ldots, \mathcal{F}_B$ are independent, the sequence $(U^n_b(1|1), X^n_{1,b}(1))$ from the codebook $\mathcal{F}_b$ is independent of the relay estimate $\widetilde{M}_{b-1}$, which is a function of $Y^n_{1,b-1}$ and the codebook $\mathcal{F}_{b-1}$. Thus, the RHS of (A18) 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 (A15). 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

$$\begin{aligned}
&\Pr\left(\mathcal{E}_{1,2}(b) \mid \mathcal{E}_1(b-1)^c\right)\\
&\leq (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \hat{\mathcal{Q}}_n} \Pr\left((U^n_b(m'_b|1), X^n_{1,b}(1), Y^n_{1,b}) \in \mathcal{A}^\delta(P_{U,X_1}P^{q'}_{Y_1|U,X_1}) \text{ for some } m'_b \neq 1\right)\\
&\leq (n+1)^{|\mathcal{S}|} \cdot 2^{nR'} \cdot \sup_{q' \in \hat{\mathcal{Q}}_n} \left[\sum_{u^n, x_1^n} P_{U^n,X_1^n}(u^n, x_1^n) \cdot \sum_{y_1^n\,:\,(u^n,x_1^n,y_1^n) \in \mathcal{A}^\delta(P_{U,X_1}P^{q'}_{Y_1|U,X_1})} P^{q^*}_{Y_1^n|X_1^n}(y_1^n|x_1^n)\right]\,,
\end{aligned} \tag{A19}$$

where the last line follows since $U^n_b(m'_b|1)$ is conditionally independent of $Y^n_{1,b}$ given $X^n_{1,b}(1)$, for every $m'_b \neq 1$. Let y_1^n satisfy $(u^n, x_1^n, y_1^n) \in \mathcal{A}^\delta(P_{U,X_1}P^{q'}_{Y_1|U,X_1})$. Then, $(x_1^n, y_1^n) \in \mathcal{A}^{\delta_2}(P^{q'}_{X_1,Y_1})$ with $\delta_2 \triangleq |\mathcal{U}| \cdot \delta$. By Lemmas 2.6 and 2.7 in [93],

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

hence,

$$P^{q^*}_{Y_1^n|X_1^n}(y_1^n|x_1^n) \leq 2^{-n\left(H_{q'}(Y_1|X_1) - \varepsilon_2(\delta)\right)}\,, \tag{A20}$$

where $\varepsilon_1(\delta), \varepsilon_2(\delta) \to 0$ as $\delta \to 0$. Therefore, by Equation (A19)−(A20), along with [93] (Lemma 2.13),

$$\Pr\left(\mathcal{E}_{1,2}(b) \mid \mathcal{E}_1(b-1)^c\right) \leq \ (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \mathcal{Q}} 2^{-n[I_{q'}(U;Y_1|X_1) - R' - \varepsilon_3(\delta)]}\,, \tag{A21}$$

with $\varepsilon_3(\delta) \to 0$ as $\delta \to 0$. Using induction, we have by (A15) that $\Pr\left(\mathcal{E}_1(b)\right)$ 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 $X^n_{1,b}(1)$ in block $b+1$, hence

$$(U^n_{b+1}(1|1), X^n_{b+1}(1,1|1), X^n_{1,b+1}(1)) \sim P_{U,X,X_1}(u,x,x_1)\,. \tag{A22}$$

At the destination receiver, decoding is performed backwards, hence the error events have a different form compared to those of the relay (cf. (A13) and the events below). Define the events,

$$\begin{aligned}\mathcal{E}_{2,1}(b) =&\{(U_{b+1}^n(\hat{M}'_{b+1}|1), X_{1,b+1}^n(1), Y_{b+1}^n) \notin \mathcal{A}^\delta(P_{U,X_1}P_{Y|U,X_1}^{q'}) \text{ for all } q' \in \hat{\mathcal{Q}}_n\}\\ \mathcal{E}_{2,2}(b) =&\{(U_{b+1}^n(\hat{M}'_{b+1}|m'_b), X_{1,b+1}^n(m'_b), Y_{b+1}^n) \in \mathcal{A}^\delta(P_{U,X_1}P_{Y_1|U,X_1}^{q'}), \text{ for some } m'_b \neq 1, q' \in \hat{\mathcal{Q}}_n\}\end{aligned} \tag{A23}$$

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

$$\begin{aligned}\mathcal{E}_2(b) \subseteq&\mathcal{E}_2(b+1) \cup \mathcal{E}_{2,1}(b) \cup \mathcal{E}_{2,2}(b)\\ =&\mathcal{E}_2(b+1) \cup (\mathcal{E}_2(b+1)^c \cap \mathcal{E}_{2,1}(b)) \cup (\mathcal{E}_2(b+1)^c \cap \mathcal{E}_{2,2}(b)),\end{aligned} \tag{A24}$$

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

$$\begin{aligned}\Pr(\mathcal{E}_2(b) \mid \mathcal{E}_1(b)^c) \leq& \Pr(\mathcal{E}_2(b+1) \mid \mathcal{E}_1(b)^c) + \Pr(\mathcal{E}_{2,1}(b) \mid \mathcal{E}_1(b)^c, \mathcal{E}_2(b+1)^c)\\ &+ \Pr(\mathcal{E}_{2,2}(b) \mid \mathcal{E}_1(b)^c, \mathcal{E}_2(b+1)^c).\end{aligned} \tag{A25}$$

By similar arguments to those used above, we have that

$$\Pr(\mathcal{E}_{2,1}(b) \mid \mathcal{E}_1(b)^c, \mathcal{E}_2(b+1)^c) \leq \Pr\left((U_{b+1}^n(1|1), X_{1,b+1}^n(1), Y_{b+1}^n) \notin \mathcal{A}^{\delta/2}(P_{U,X_1}P_{Y|U,X_1}^{q^*}) \mid \mathcal{E}_1(b)^c\right), \tag{A26}$$

which tends to zero exponentially as $n \to \infty$, due to (A22), 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 (A25) 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 (A25) that the second term on the RHS of (A12) tends to zero exponentially as $n \to \infty$, for $b \in [1 : B-1]$.

Moving to the error event for M''_b, define

$$\begin{aligned}\mathcal{E}_{3,1}(b) =&\{(U_b^n(\hat{M}'_b|\hat{M}'_{b-1}), X_b^n(\hat{M}'_b, 1|\hat{M}'_{b-1}), X_{1,b}(\hat{M}'_{b-1}), Y_b^n) \notin \mathcal{A}^\delta(P_{U,X,X_1}P_{Y|X,X_1}^{q'}), \text{ for all } q' \in \hat{\mathcal{Q}}_n\}\\ \mathcal{E}_{3,2}(b) =&\{(U_b^n(\hat{M}'_b|\hat{M}'_{b-1}), X_b^n(\hat{M}'_b, m''_b|\hat{M}'_{b-1}), X_{1,b}(\hat{M}'_{b-1}), Y_b^n) \in \mathcal{A}^\delta(P_{U,X,X_1}P_{Y|X,X_1}^{q'}),\\ &\text{for some } m''_b \neq 1, q' \in \hat{\mathcal{Q}}_n\}.\end{aligned} \tag{A27}$$

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,

$$\begin{aligned}&\Pr(\mathcal{E}_3(b) \mid \mathcal{E}_1(b)^c \cap \mathcal{E}_2(b)^c \cap \mathcal{E}_2(b-1)^c)\\ \leq& \Pr(\mathcal{E}_{3,1}(b) \mid \mathcal{E}_1(b)^c \cap \mathcal{E}_2(b)^c \cap \mathcal{E}_2(b-1)^c) + \Pr(\mathcal{E}_{3,2}(b) \mid \mathcal{E}_1(b)^c \cap \mathcal{E}_2(b)^c \cap \mathcal{E}_2(b-1)^c)\\ \leq& e^{-a_0 n} + (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \mathcal{Q}} \sum_{m''_b \neq 1} \Pr\left((U_b^n(1|1), X_b^n(1, m''_b|1), X_{1,b}(1), Y_b^n) \in \mathcal{A}^\delta(P_{U,X,X_1}P_{Y|X,X_1}^{q'}) \mid \mathcal{E}_1(b)^c\right)\\ \leq& e^{-a_0 n} + (n+1)^{|\mathcal{S}|} \cdot \sup_{q' \in \mathcal{Q}} 2^{-n[I_{q'}(X;Y|U,X_1) - R'' - \varepsilon_5(\delta)]}\end{aligned} \tag{A28}$$

where $a_0 > 0$ and $\varepsilon_5(\delta) \to 0$ as $\delta \to 0$. The second inequality holds by (A22) 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 (A12) 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 $R < \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}})$. Therefore, there must exist a $(2^{nR}, n, \varepsilon)$ deterministic code, for a sufficiently large n. □

Appendix A.2. Cutset Upper Bound

This is a straightforward consequence of the cutset bound in [4]. 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 $\mathcal{Q}$,

$$R > \max_{p(x,x_1)} \min \left\{ I_{q^*}(X, X_1; Y)\,,\ I_{q^*}(X; Y, Y_1|X_1) \right\}. \tag{A29}$$

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 (A29) can be achieved over the relay channel $W_{Y,Y_1|X,X_1}$, in contradiction to [4]. We deduce that the assumption is false, and $R > \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}})$ cannot be achieved. □

Appendix B. Proof of Corollary 1

This is a straightforward consequence of Lemma 1, which states that the capacity of the compound relay channel is bounded by $\mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}}) \leq \mathbb{C}(\mathcal{L}^{\mathcal{Q}}) \leq \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}})$. 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 [90], cf. (8) and (12). Similarly, if $W_{Y,Y_1|X,X_1,S}$ is strongly degraded, i.e., $W_{Y,Y_1|X,X_1,S} = W_{Y_1|X,X_1} W_{Y|Y_1,X_1,S}$, then $I_q(X; Y, Y_1|X_1) = I(X; Y_1|X_1)$, and by (8) and (13),

$$\mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) = \min_{q(s)\in\mathcal{Q}} \max_{p(x,x_1)} \min \left\{ I_q(X, X_1; Y)\,,\ I(X; Y_1|X_1) \right\}, \tag{A30}$$

$$\mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}}) = \max_{p(x,x_1)} \min_{q(s)\in\mathcal{Q}} \min \left\{ I_q(X, X_1; Y)\,,\ I(X; Y_1|X_1) \right\}. \tag{A31}$$

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., [94] (Section 3.4])), hence the bounds (A30) and (A31) coincide by the minimax theorem [90]. □

Appendix C. Proof of Corollary 2

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, they can use a block coding scheme that is synchronized with the jammer block strategy. Thus, 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, (17) and (18) follow from Lemma 1. As for the second part of Corollary 2, observe that the block Markov coding scheme used in the proof of the partial 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 (A18)–(A20) and (A26), is now replaced by the state distribution q_b^* which corresponds to block $b \in [1 : B]$. □

Appendix D. Proof of Theorem 1

First, we explain the general idea. We modify Ahlswede's Robustification Technique (RT) [59] 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 [59], 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.

Appendix D.1. Partial Decode-Forward Lower Bound

We show that every rate $R < \mathsf{R}^{\star}_{PDF}(\mathcal{L})$ (see (19)) can be achieved by random codes over the AVRC $\mathcal{L}$, i.e., $\mathbb{C}(\mathcal{L}) \geq \mathsf{R}^{\star}_{PDF}(\mathcal{L})$. We start with Ahlswede's RT [59], 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 $q \in \mathcal{P}(\mathcal{S})$,

$$\sum_{s^n \in \mathcal{S}^n} q(s^n) h(s^n) \leq \alpha_n \,, \tag{A32}$$

then,

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} h(\pi s^n) \leq \beta_n \,, \quad \text{for all } s^n \in \mathcal{S}^n \,, \tag{A33}$$

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 2, for every $R < \mathsf{R}^{\star}_{PDF}(\mathcal{L})$, 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 $x^n_b = f_b(m'_b, m''_b | m'_{b-1})$, with fixed m_0 and m_B, and the relay transmits $x^n_{1,b} = f_{1,b}(y^n_{1,b-1})$, using the sequence received in the previous block. After receiving the entire output sequence $(y^n_b)_{b=1}^B$, the decoder finds an estimate for the messages. Set $\hat{m}'_B = 1$. The first part of each message is decoded backwards as $\hat{m}'_b = g'_b(y^n_{b+1}, \hat{m}'_{b+1})$, for $b = B-1, B-2, \ldots, 1$. Then, the second part of each message is decoded as $\hat{m}''_b = g''_b(y^n_b, \hat{m}'_1, \ldots, \hat{m}'_{B-1})$, 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

$$\Pr_{\mathscr{C}_{BM}}(\mathcal{E}'_b \mid (\mathcal{E}'_{b+1})^c) \leq e^{-2\theta n} \,, \; \Pr_{\mathscr{C}_{BM}}(\mathcal{E}''_b \mid \mathcal{E}'^c_1, \ldots, \mathcal{E}'^c_{b-1}) \leq e^{-2\theta n} \tag{A34}$$

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

$$\sum_{s^n_1 \in \mathcal{S}^n} q_1(s^n_1) \sum_{s^n_2 \in \mathcal{S}^n} q_2(s^n_2) \; \cdots \sum_{s^n_{b+1} \in \mathcal{S}^n} q_{b+1}(s^n_{b+1}) \cdot h'_b(s^n_1, s^n_2, \ldots, s^n_{b+1}) \leq e^{-2\theta n} \,, \tag{A35}$$

and

$$\sum_{s^n_1 \in \mathcal{S}^n} q_1(s^n_1) \sum_{s^n_2 \in \mathcal{S}^n} q_2(s^n_2) \; \cdots \sum_{s^n_b \in \mathcal{S}^n} q_b(s^n_b) \cdot h''_b(s^n_1, s^n_2, \ldots, s^n_b) \leq e^{-2\theta n} \,, \tag{A36}$$

where

$$h'_b(s_1^n, s_2^n, \ldots, s_{b+1}^n) = \frac{1}{2^{n(b+1)(R'+R'')}} \sum_{(m'_1,m''_1),\ldots,(m'_{b+1},m''_{b+1})}$$
$$\sum_{y_{1,b}^n \in \mathcal{Y}_1^n} \Pr\left(Y_{1,b}^n = y_{1,b}^n \mid (M'_1, M''_1) = (m'_1, m''_1), \ldots, (M'_b, M''_b) = (m'_b, m''_b), S_1^n = s_1^n, \ldots, S_b^n = s_b^n\right)$$
$$\times \sum_{y_{b+1}^n : g'_b(y_{b+1}^n, m'_{b+1}) \neq m'_b} W_{Y^n|X^n,X_1^n,S^n}(y_{b+1}^n | f_{b+1}(m'_{b+1}, m''_{b+1}|m'_b), f_{1,b+1}(y_{1,b}^n), s_{b+1}^n) \tag{A37}$$

and

$$h''_b(s_1^n, s_2^n, \ldots, s_b^n) = \frac{1}{2^{nR''}} \sum_{m''_b=1}^{2^{nR''}} \frac{1}{2^{nR'(B-1)}} \sum_{m'_1,\ldots,m'_{B-1}}$$
$$\sum_{y_{1,b-1}^n \in \mathcal{Y}_1^n} \Pr\Big(Y_{1,b-1}^n = y_{1,b-1}^n \big| (M'_1, M''_1) = (m'_1, m''_1), \ldots, (M'_{b-1}, M''_{b-1}) = (m'_{b-1}, m''_{b-1}),$$
$$S_1^n = s_1^n, \ldots, S_{b-1}^n = s_{b-1}^n\Big)$$
$$\times \sum_{y_b^n, y_{1,b}^n : g''_b(y_b^n, m'_1, \ldots, m'_{B-1}) \neq m''_b} W_{Y^n|X^n,X_1^n,S^n}(y_b^n | f_b(m'_b, m''_b|m'_{b-1}), f_{1,b}(y_{1,b-1}^n), s_b^n)\,. \tag{A38}$$

The conditioning in the equations above can be explained as follows. In (A37), 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 $t \leq b$. 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 (A38), 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 $\hat{M}'_1, \ldots, \hat{M}'_{B-1}$. By (A35)–(A36), for every $t \in [1:b]$, h'_b and h''_b as functions of s_{t+1}^n and s_t^n, respectively, satisfy (A32) 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

$$\frac{1}{(n!)^{b+1}} \sum_{\pi_1,\pi_2,\ldots,\pi_{b+1} \in \Pi_n} h'_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_{b+1} s_{b+1}^n) \leq (n+1)^{B|\mathcal{S}|} e^{-2\theta n} \leq e^{-\theta n}\,,,$$
$$\frac{1}{(n!)^{b}} \sum_{\pi_1,\pi_2,\ldots,\pi_b \in \Pi_n} h''_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n) \leq (n+1)^{B|\mathcal{S}|} e^{-2\theta n} \leq e^{-\theta n}\,, \tag{A39}$$

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

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

$$h'_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_{b+1} s_{b+1}^n) = \mathbb{E}\, h'_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_{b+1} s_{b+1}^n | M'_t, M''_t, t = 1, \ldots, b+1)\,, \tag{A40}$$

with

$$\begin{aligned}
&h'_b(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_{b+1} s_{b+1}^n | m'_t, m''_t, t = 1, \ldots, b+1) \\
&= \sum_{y_{1,1},\ldots,y_{1,b}} \prod_{t=0}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(y_{1,t+1}^n | f_{t+1}(m'_{t+1}, m''_{t+1}|m'_t), f_{1,t+1}(y_{1,t}^n), \pi_{t+1} s_{t+1}^n) \\
&\quad \times \sum_{y_{b+1}^n : g'_b(y_{b+1}^n, m'_{b+1}) \neq m'_b} W_{Y^n|X^n,X_1^n,S^n}(y_{b+1}^n | f_{b+1}(m'_{b+1}, m''_{b+1}|m'_b), f_{1,b+1}(y_{1,b}^n), \pi_{b+1} s_{b+1}^n) \\
&\overset{(a)}{=} \sum_{y_{1,1},\ldots,y_{1,b}} \prod_{t=0}^{b-1} W_{Y_1^n|X^n,X_1^n,S^n}(\pi_{t+1} y_{1,t+1}^n | f_{t+1}(m'_{t+1}, m''_{t+1}|m'_t), f_{1,b+1}(\pi_t y_{1,t}^n), \pi_{t+1} s_{t+1}^n)
\end{aligned}$$

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

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

$$\times \sum_{y_{b+1}^n : g_b'(\pi_{b+1}y_{b+1}^n, m_{b+1}') \neq m_b'} W_{Y^n|X^n,X_1^n,S^n}(y_{b+1}^n | \pi_{b+1}^{-1} f_{b+1}(m_{b+1}', m_{b+1}''|m_b'), \pi_{b+1}^{-1} f_{1,b+1}(\pi_b y_{1,b}^n), s_{b+1}^n)\,, \quad \text{(A41)}$$

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,

$$h_b''(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n) = \mathbb{E} h_b''(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n | M_1', \ldots, M_{B-1}', M_t'', t = 1, \ldots, b)\,, \quad \text{(A42)}$$

with

$$h_b''(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n | m_1', \ldots, m_{B-1}', m_t'', t = 1, \ldots, b)$$

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

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

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

$$\times \sum_{y_b^n : g_b''(\pi_b y_b^n, m_1', \ldots, m_{B-1}') \neq m_b''} W_{Y^n|X^n,X_1^n,S^n}(\pi_b y_b^n | f_b(m_b', m_b''|m_{b-1}'), f_{1,b}(\pi_{b-1} y_{1,b-1}^n), \pi_b s_b^n)$$

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

$$\times \sum_{y_b^n : g_b''(\pi_b y_b^n, m_1', \ldots, m_{B-1}') \neq m_b''} W_{Y^n|X^n,X_1^n,S^n}(y_b^n | \pi_b^{-1} f_b(m_b', m_b''|m_{b-1}'), \pi_b^{-1} f_{1,b}(\pi_{b-1} y_{1,b-1}^n), s_b^n)\,. \quad \text{(A43)}$$

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

$$f_{b,\pi}(m_b', m_b''|m_{b-1}') = \pi_b^{-1} f_b(m_b', m_b''|m_{b-1}')\,, \quad f_{1,b,\pi}(y_{1,b-1}^n) = \pi_b^{-1} f_{1,b}(\pi_{b-1} y_{1,b-1}^n)\,, \quad \text{(A44a)}$$

and

$$g_{b,\pi}'(y_{b+1}^n, \hat{m}_{b+1}') = g_b'(\pi_{b+1} y_{b+1}^n, \hat{m}_{b+1}')\,, \quad g_{b,\pi}''(y_b^n, \hat{m}_1', \ldots, \hat{m}_{B-1}') = g_b''(\pi y_b^n, \hat{m}_1', \ldots, \hat{m}_{B-1}')\,, \quad \text{(A44b)}$$

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 (A41) and (A43), 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

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

$$\Pr_{\mathscr{C}_{BM}^{\Pi}}(\mathcal{E}_b'' \mid \mathcal{E}_1'^c, \ldots, \mathcal{E}_{B-1}'^c, S_1^n = s_1^n, \ldots, S_b^n = s_b^n) = \sum_{\pi_1,\ldots,\pi_B \in \Pi_n} \mu(\pi_1, \ldots, \pi_B) h_b''(\pi_1 s_1^n, \pi_2 s_2^n, \ldots, \pi_b s_b^n)\,, \quad \text{(A45)}$$

for all $s_1^n, \ldots, s_{b+1}^n \in \mathcal{S}^n$, $b \in [1 : B-1]$, and therefore, together with (A39), we have that the probability of error of the random code $\mathscr{C}_{BM}^{\Pi}$ is bounded by $P_e^{(n)}(q, \mathscr{C}_{BM}^{\Pi}) \leq e^{-\theta n}$, 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.

Appendix D.2. 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., $\mathbb{C}^{\star}(\mathcal{L}) \leq \mathbb{C}^{\star}(\mathcal{L}^{\mathcal{P}(\mathcal{S})})$. □

Appendix E. Proof of Lemma 2

We use the approach of [55], with the required adjustments. We use the random code constructed in the proof of Theorem 1. 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.,

$$\mathbb{C}(\mathcal{W}_1(x_{1,1})) > 0\,, \text{ and } \mathbb{C}(\mathcal{W}(x_{1,2})) > 0\,, \tag{A46}$$

for some $x_{1,1}, x_{1,2} \in \mathcal{X}_1$ (see (23)). By Theorem 1, for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR}, n, \varepsilon)$ random code $\mathscr{C}^{\Gamma} = \left(\mu(\gamma) = \frac{1}{k}, \Gamma = [1 : k], \{\mathscr{C}_{\gamma}\}_{\gamma \in \Gamma}\right)$, where $\mathscr{C}_{\gamma} = (f_{\gamma}^n, f_{1,\gamma}, g_{\gamma})$, for $\gamma \in \Gamma$. Following Ahlswede's Elimination Technique [55], it can be assumed that the size of the code collection is bounded by $k = |\Gamma| \leq n^2$. By (A46), 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 $\mathscr{C}_{\mathrm{i}}' = (\widetilde{f}^{\nu'}, \widetilde{g}')$, 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 $\mathscr{C}_{\mathrm{i}}'' = (\widetilde{f}^{\nu''}, \widetilde{g}'')$ 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 $\mathscr{C}_{\mathrm{i}}'$ and $\mathscr{C}_{\mathrm{i}}''$ 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 $n \to \infty$. □

Appendix F. Proof of Corollary 4

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 [57] (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 [57] (Theorem 1). Hence, by Lemma 2, $\mathbb{C}(\mathcal{L}) = \mathbb{C}^{\star}(\mathcal{L})$, and by Theorem 1, $\mathsf{R}_{PDF}^{\star}(\mathcal{L}) \leq \mathbb{C}(\mathcal{L}) \leq \mathsf{R}_{CS}^{\star}(\mathcal{L})$.

Part 3 immediately follows from part 1 and Corollary 3. As for part 2, consider a strongly reversely degraded relay channel. We claim that if $W_{Y|X,X_1,S}$ is symmetrizable-$\mathcal{X}|\mathcal{X}_1$, then $W_{Y_1|X,X_1,S}$

is also symmetrizable-$\mathcal{X}|\mathcal{X}_1$. Indeed, suppose that $W_{Y|X,X_1,S}$ is symmetrized by some $J(s|x,x_1)$ (see Definition 24). Then, for every $x,\tilde{x}\in\mathcal{X}$, $x_1\in\mathcal{X}_1$, and $y_1\in\mathcal{Y}_1$,

$$\begin{aligned}\sum_{s\in\mathcal{S}} J(s|\tilde{x},x_1)W_{Y_1|X,X_1,S}(y_1|x,x_1,s) &= \sum_{s\in\mathcal{S}} J(s|\tilde{x},x_1)\sum_{y\in\mathcal{Y}} W_{Y,Y_1|X,X_1,S}(y,y_1|x,x_1,s)\\ &\overset{(a)}{=} \sum_{y\in\mathcal{Y}} W_{Y_1|Y,X_1}(y_1|y,x_1)\sum_{s\in\mathcal{S}} J(s|\tilde{x},x_1)W_{Y|X,X_1,X}(y|x,x_1,s)\\ &\overset{(b)}{=} \sum_{y\in\mathcal{Y}} W_{Y_1|Y,X_1}(y_1|y,x_1)\sum_{s\in\mathcal{S}} J(s|x,x_1)W_{Y|X,X_1,X}(y|\tilde{x},x_1,s)\\ &\overset{(c)}{=} \sum_{s\in\mathcal{S}} J(s|x,x_1)\sum_{y\in\mathcal{Y}} W_{Y,Y_1|X,X_1,S}(y,y_1|\tilde{x},x_1,s)\\ &= \sum_{s\in\mathcal{S}} J(s|x,x_1)W_{Y_1|X,X_1,S}(y_1|\tilde{x},x_1,s)\,, \end{aligned} \tag{A47}$$

where (a) and (c) hold since $W_{Y,Y_1|X,X_1,S}$ is strongly reversely degraded, and (b) holds since $W_{Y|X,X_1,S}$ is symmetrized by $J(s|x,x_1)$. This means that $W_{Y_1|X,X_1,S}$ is also symmetrizable-$\mathcal{X}|\mathcal{X}_1$. It can be deduced that given the conditions of part 2, both $W_{Y|X,X_1,S}$ and $W_{Y_1|X,X_1,S}$ are non-symmetrizable-$\mathcal{X}|\mathcal{X}_1$. Hence, the proof follows from part 1 and Corollary 3. □

Appendix G. Proof of Lemma 3

The proof is based on generalizing the technique by [56]. 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.,

$$\mathsf{M}\triangleq 2^{nR}\geq 2\,. \tag{A48}$$

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 (24). 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

$$\begin{aligned}P_e^{(n)}(q,\mathscr{C}) &= \sum_{s^n\in\mathcal{S}^n}\left[\frac{1}{\mathsf{M}}\sum_{\tilde{m}=1}^{\mathsf{M}} J^n(s^n|x^n(\tilde{m}))\right]\cdot\frac{1}{\mathsf{M}}\sum_{m=1}^{\mathsf{M}}\sum_{(y^n,y_1^n):g(y^n)\neq m} W^n(y^n,y_1^n|x^n(m),f_1^n(y_1^n),s^n)\\ &= \frac{1}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\tilde{m}=1}^{\mathsf{M}}\sum_{(y^n,y_1^n):g(y^n)\neq m}\sum_{s^n\in\mathcal{S}^n} W^n(y^n,y_1^n|x^n(m),f_1^n(y_1^n),s^n)J^n(s^n|x^n(\tilde{m}))\\ &\quad+\frac{1}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\tilde{m}=1}^{\mathsf{M}}\sum_{(y^n,y_1^n):g(y^n)\neq \tilde{m}}\sum_{s^n\in\mathcal{S}^n} W^n(y^n,y_1^n|x^n(\tilde{m}),f_1^n(y_1^n),s^n)J^n(s^n|x^n(m)) \end{aligned} \tag{A49}$$

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 $\tilde{m}$. Then, consider the last sum, and observe that by (24), we have that

$$\begin{aligned}\sum_{s^n\in\mathcal{S}^n} W^n(y^n,y_1^n|x^n(\tilde{m}),f_1^n(y_1^n),s^n)J^n(s^n|x^n(m)) &= \prod_{i=1}^n\left[\sum_{s_i\in\mathcal{S}} W(y_i,y_{1,i}|x_i(\tilde{m}),f_{1,i}(y_1^{i-1}),s_i)J(s_i|x_i(m))\right]\\ &= \prod_{i=1}^n\left[\sum_{s_i\in\mathcal{S}} W(y_i,y_{1,i}|x_i(m),f_{1,i}(y_1^{i-1}),s_i)J(s_i|x_i(\tilde{m}))\right]\\ &= \sum_{s^n\in\mathcal{S}^n} W^n(y^n,y_1^n|x^n(m),f_1^n(y_1^n),s^n)J^n(s^n|x^n(\tilde{m}))\,. \end{aligned} \tag{A50}$$

Substituting (A50) in (A49), we have

$$\begin{aligned} P_e^{(n)}(q,\mathscr{C}) =& \frac{1}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}=1}^{\mathsf{M}}\sum_{s^n\in\mathcal{S}^n}\Bigg[\sum_{(y^n,y_1^n):g(y^n)\neq m} W^n(y^n,y_1^n|x^n(m),f_1^n(y_1^n),s^n)J^n(s^n|x^n(\widetilde{m})) \\ &+\sum_{(y^n,y_1^n):g(y^n)\neq \widetilde{m}} W^n(y^n,y_1^n|x^n(m),f_1^n(y_1^n),s^n)J^n(s^n|x^n(\widetilde{m}))\Bigg] \\ \geq& \frac{1}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}\neq m}\sum_{s^n\in\mathcal{S}^n}\sum_{y^n,y_1^n} W^n(y^n,y_1^n|x^n(m),f_1^n(y_1^n),s^n)J^n(s^n|x^n(\widetilde{m})) \\ =& \frac{\mathsf{M}(\mathsf{M}-1)}{2\mathsf{M}^2} \geq \frac{1}{4}, \end{aligned} \tag{A51}$$

where the last inequality follows from (A48), hence a positive rate cannot be achieved. □

Appendix H. Proof of Lemma 4

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. The proof follows similar lines as in Appendix G. First, assume to the contrary that there exists a $(2^{nR},n,\varepsilon)$ code $\mathscr{C} = (f,f_1,g)$, with $\mathsf{M}\triangleq 2^{nR}\geq 2$. By Definition 4, there exists $J(s|x_1,y_1)$ that satisfies (28). Hence, defining

$$q(s^n) = \frac{1}{\mathsf{M}}\sum_{m=1}^{\mathsf{M}}\sum_{y_1^n\in\mathcal{Y}_1} W_{Y_1^n|X^n,X_1^n}(y_1^n|f(m),f_1^n(y_1^n))J^n(s^n|f_1^n(y_1^n),y_1^n), \tag{A52}$$

where $J^n(s^n|x_1^n,y_1^n) = \prod_{i=1}^n J(s_i|x_{1,i},y_{1,i})$, we have that

$$\sum_{s^n\in\mathcal{S}^n} W^n(y^n|\tilde{y}_1^n,f_1^n(\tilde{y}_1^n),s^n)J^n(s^n|f_1^n(y_1^n),y_1^n) = \sum_{s^n\in\mathcal{S}^n} W^n(y^n|y_1^n,f_1^n(y_1^n),s^n)J^n(s^n|f_1^n(\tilde{y}_1^n),\tilde{y}_1^n). \tag{A53}$$

By similar manipulations as in Appendix G, we obtain

$$\begin{aligned} P_e^{(n)}(q,\mathscr{C}) =& \frac{1}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}=1}^{\mathsf{M}}\sum_{y_1^n,\tilde{y}_1^n} W_{Y_1^n|X^n,X_1^n}(\tilde{y}_1^n|f(\widetilde{m}),f_1^n(\tilde{y}_1^n))W_{Y_1^n|X^n,X_1^n}(y_1^n|f(m),f_1^n(y_1^n)) \\ &\times\sum_{s^n\in\mathcal{S}^n}\Bigg[\sum_{y^n:g(y^n)\neq m} W^n(y^n|y_1^n,f_1^n(y_1^n),s^n)J^n(s^n|f_1^n(\tilde{y}_1^n),\tilde{y}_1^n) \\ &+\sum_{y^n:g(y^n)\neq\widetilde{m}} W^n(y^n|y_1^n,f_1^n(y_1^n),s^n)J^n(s^n|f_1^n(\tilde{y}_1^n),\tilde{y}_1^n)\Bigg] \\ \geq& \frac{\mathsf{M}(\mathsf{M}-1)}{2\mathsf{M}^2} \geq \frac{1}{4}, \end{aligned} \tag{A54}$$

hence a positive rate cannot be achieved. □

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

$$\mathbb{C}^\star(\mathcal{L}) = \mathsf{R}_{PDF}^\star(\mathcal{L}) = \mathsf{R}_{CS}^\star(\mathcal{L}) = \max_{p(x,x_1)}\min\left\{\min_{0\leq q\leq 1} I_q(X,X_1;Y),\ I(X;Y_1|X_1)\right\}, \tag{A55}$$

due to part 2 of Corollary 3, 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,

$$\begin{aligned} &I(X; Y_1|X_1) = 1 - h(\theta)\,, \\ &H_q(Y) = \frac{1}{2}\left[-q\log\left(\frac{1}{2}q\right) - (1-q)\log\left(\frac{1}{2}(1-q)\right)\right] - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1 + \frac{1}{2}h(q)\,, \\ &H_q(Y|X, X_1) = h(q)\,. \end{aligned} \tag{A56}$$

Hence,

$$\mathbb{C}^\star(\mathcal{L}) \geq \min\left\{\min_{0\leq q\leq 1}\left[1 - \frac{1}{2}h(q)\right], 1 - h(\theta)\right\} = \min\left\{\frac{1}{2}, 1 - h(\theta)\right\}. \tag{A57}$$

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

$$\mathbb{C}^\star(\mathcal{L}) \leq \max_{p(x,x_1)} I(X; Y_1|X_1) = 1 - h(\theta)\,, \tag{A58}$$

and

$$\begin{aligned} \mathbb{C}^\star(\mathcal{L}) &\leq \max_{p(x,x_1)} \min_{0\leq q\leq 1} I_q(X, X_1; Y) \leq \max_{p(x,x_1)} \left[H_q(Y) - H_q(Y|X, X_1)\right]\Big|_{q=\frac{1}{2}} \\ &= \max_{0\leq p\leq 1}\left[1 + \frac{1}{2}h(p)\right] - 1 = \frac{1}{2}\,, \end{aligned} \tag{A59}$$

where $p \triangleq \Pr(X_1 = 1)$. □

Appendix J. Proof of Lemma 5

The proof follows the lines of [5]. Consider an AVRC $\mathcal{L} = \{W_{Y|X',X_1}\ W_{Y_1|X'',X_1,S}\}$ with orthogonal sender components. We apply Theorem 1, which states that $\mathsf{R}^\star_{PDF}(\mathcal{L}) \leq \mathbb{C}^\star(\mathcal{L}) \leq \mathsf{R}^\star_{CS}(\mathcal{L})$.

Appendix J.1. 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 $\mathsf{R}^\star_{PDF}(\mathcal{L}) \triangleq \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}})\Big|_{\mathcal{Q}=\mathcal{P}(\mathcal{S})}$. Hence, by (9),

$$\mathsf{R}^\star_{PDF}(\mathcal{L}_2) \geq \max_{p(x_1)p(x'|x_1)p(x''|x_1)} \min\left\{I(X', X'', X_1; Y)\,,\ \min_{q(s)} I_q(X''; Y_1|X_1) + I(X'; Y|X_1, X'')\right\}. \tag{A60}$$

Now, by (29), we have that $(X'', Y_1) - (X', X_1) - Y$ 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 $(X'', Y_1) - X_1 - Y$ 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, (A60) reduces to the expression in the RHS of (30). If $W_{Y_1|X'',X_1,S}$ is non-symmetrizable-$\mathcal{X}''|\mathcal{X}_1$, then (A60) is achievable by deterministic codes as well, due to Corollary 4.

Appendix J.2. Converse Proof

By (8) and (19), the cutset upper bound takes the following form,

$$\begin{aligned} \mathsf{R}^\star_{CS}(\mathcal{L}) &= \min_{q(s)} \max_{p(x',x'',x_1)} \min\left\{I(X', X'', X_1; Y)\,,\ I_q(X', X''; Y, Y_1|X_1)\right\} \\ &= \max_{p(x',x'',x_1)} \min\left\{I(X', X'', X_1; Y)\,,\ \min_{q(s)} I_q(X', X''; Y, Y_1|X_1)\right\}, \end{aligned} \tag{A61}$$

where the last line is due to the minimax theorem [90]. For the AVRC with orthogonal sender components, as specified by (29), we have the following Markov relations,

$$Y_1 - (X'', X_1) - (X', Y)\,, \tag{A62}$$
$$(X'', Y_1) - (X', X_1) - Y\,. \tag{A63}$$

Hence, by (A63), $I(X', X'', X_1; Y) = I(X', X_1; Y)$. As for the second mutual information in the RHS of (A61), by the mutual information chain rule,

$$\begin{aligned} I_q(X', X''; Y, Y_1|X_1) &= I_q(X''; Y_1|X_1) + I_q(X'; Y_1|X'', X_1) + I_q(X', X''; Y|X_1, Y_1) \\ &\overset{(a)}{=} I_q(X''; Y_1|X_1) + I_q(X', X''; Y|X_1, Y_1) \\ &\overset{(b)}{=} I_q(X''; Y_1|X_1) + H_q(Y|X_1, Y_1) - H(Y|X', X_1) \\ &\overset{(c)}{\leq} I_q(X''; Y_1|X_1) + I(X'; Y|X_1) \end{aligned} \tag{A64}$$

where (a) is due to (A62), (b) is due to (A63), and (c) holds since conditioning reduces entropy. Therefore,

$$\mathsf{R}^{\star}_{CS}(\mathcal{L}) \leq \max_{p(x', x'', x_1)} \min \left\{ I(X', X_1; Y)\,,\ \min_{q(s)} I_q(X''; Y_1|X_1) + I(X'; Y|X_1) \right\}. \tag{A65}$$

Without loss of generality, the maximization in (A65) can be restricted to distributions of the form $p(x', x'', x_1) = p(x_1) \cdot p(x'|x_1) \cdot p(x''|x_1)$. □

Appendix K. Proof of Lemma 6

Consider the Gaussian compound relay channel with SFD under input constraints Ω and Ω_1 and state constraint Λ, i.e., $\mathcal{Q} = \{q(s) \,:\, \mathbb{E}S^2 \leq \Lambda\}$.

Appendix K.1. Achievability Proof

Consider the direct part. Although we previously assumed that the input, state and output alphabets are finite, our results for the compound relay channel can be extended to the continuous case as well, using standard discretization techniques [3] (Section 3.4.1); [55,95]. In particular, Lemma 1 can be extended to the compound relay channel $\mathcal{L}^{\mathcal{Q}}$ under input constraints Ω and Ω_1 and state constraint Λ, by choosing a distribution $p(x', x'', x_1)$ such that $\mathbb{E}(X'^2 + X''^2) \leq \Omega$ and $\mathbb{E}X_1^2 \leq \Omega_1$. Then, the capacity of $\mathcal{L}^{\mathcal{Q}}$ is bounded by

$$\begin{aligned} \mathbb{C}(\mathcal{L}^{\mathcal{Q}}) \geq \mathsf{R}_{PDF}(\mathcal{L}^{\mathcal{Q}}) \geq \max_{\substack{p(x'')p(x,x_1):\\ E(X'^2+X''^2)\leq\Omega,\\ \mathbb{E}X_1^2\leq\Omega_1}} \min \Big\{ & \min_{q(s):\,\mathbb{E}S^2\leq\Lambda} I_q(X_1; Y) + \min_{q(s):\,\mathbb{E}S^2\leq\Lambda} I_q(X'; Y|X_1)\,, \\ & I(X''; Y_1) + \min_{q(s):\,\mathbb{E}S^2\leq\Lambda} I_q(X'; Y|X_1) \Big\}, \end{aligned} \tag{A66}$$

which follows from the partial decode-forward lower bound by taking $U = X''$. Lemma 1 further states that there exists a block Markov code that achieves this rate such that the probability of error decays exponentially as the blocklength increases.

Let $0 \leq \alpha, \rho \leq 1$, and let (X', X'', X_1) be jointly Gaussian with

$$X' \sim \mathcal{N}(0, \alpha\Omega)\,,\ X'' \sim \mathcal{N}(0, (1-\alpha)\Omega)\,,\ X_1 \sim \mathcal{N}(0, \Omega_1)\,, \tag{A67}$$

where the correlation coefficient of X' and X_1 is ρ, while X'' is independent of (X', X_1). Hence,

$$I(X'';Y_1) = \frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\sigma^2}\right). \tag{A68}$$

Since Gaussian noise is the worst additive noise under variance constraint [96] (Lemma II.2), and as $\mathbb{V}\mathrm{ar}(X'|X_1=x_1) = (1-\rho^2)\alpha\Omega$ for all $x_1 \in \mathbb{R}$, we have that

$$\min_{q(s):\,\mathbb{E}S^2\le\Lambda} I_q(X';Y|X_1) = \frac{1}{2}\log\left(1+\frac{(1-\rho^2)\alpha\Omega}{\Lambda}\right). \tag{A69}$$

It is left for us to evaluate the first term in the RHS of (A66). Then, by standard whitening transformation, there exist two independent Gaussian random variables T_1 and T_2 such that

$$X' + X_1 = T_1 + T_2\,, \tag{A70}$$

$$T_1 \sim \mathcal{N}(0,(1-\rho^2)\alpha\Omega)\,,\ T_2 \sim \mathcal{N}(0,\Omega_1+\rho^2\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1})\,. \tag{A71}$$

Hence, $Y = T_1 + T_2 + S$, and as $\mathbb{V}\mathrm{ar}(X'|X_1=x_1) = \mathbb{V}\mathrm{ar}(T_1)$ for all $x_1 \in \mathbb{R}$, we have that

$$\begin{aligned} I_q(X_1;Y) =& H_q(Y) - H_q(X'+S|X_1) \\ =& H_q(Y) - H_q(T_1+S) = I_q(T_2;Y) \end{aligned} \tag{A72}$$

Let $\bar{S} \triangleq T_1 + S$. Then, since Gaussian noise is the worst additive noise under variance constraint [96] (Lemma II.2),

$$\begin{aligned} \min_{q(s):\,\mathbb{E}S^2\le\Lambda} I_q(X_1;Y) =& \min_{q(s):\,\mathbb{E}S^2\le\Lambda} I_q(T_2;T_2+\bar{S}) = \frac{1}{2}\log\left(1+\frac{\mathbb{V}\mathrm{ar}(T_2)}{\mathbb{V}\mathrm{ar}(T_1)+\Lambda}\right) \\ =& \frac{1}{2}\log\left(\frac{\Omega_1+\rho^2\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1}+\Lambda}{(1-\rho^2)\alpha\Omega+\Lambda}\right). \end{aligned} \tag{A73}$$

Substituting (A68), (A69) and (A73) in the RHS of (A66), we have that

$$\begin{aligned} \mathbb{C}(\mathcal{L}^{\mathcal{Q}}) \ge \max_{0\le\alpha,\rho\le1} \min\Bigg\{ & \frac{1}{2}\log\left(\frac{\Omega_1+\rho^2\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1}+\Lambda}{(1-\rho^2)\alpha\Omega+\Lambda}\right) + \frac{1}{2}\log\left(1+\frac{(1-\rho^2)\alpha\Omega}{\Lambda}\right), \\ & \frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\sigma^2}\right) + \frac{1}{2}\log\left(1+\frac{(1-\rho^2)\alpha\Omega}{\Lambda}\right)\Bigg\}. \end{aligned} \tag{A74}$$

Observe that the first sum in the RHS of (A74) can be expressed as

$$\begin{aligned} & \frac{1}{2}\log\left(\frac{\Omega_1+\rho^2\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1}+\Lambda}{(1-\rho^2)\alpha\Omega+\Lambda}\right) + \frac{1}{2}\log\left(\frac{(1-\rho^2)\alpha\Omega+\Lambda}{\Lambda}\right) \\ =& \frac{1}{2}\log\left(\frac{\Omega_1+\rho^2\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1}+\Lambda}{\Lambda}\right) = \frac{1}{2}\log\left(1+\frac{\Omega_1+\rho^2\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1}}{\Lambda}\right). \end{aligned} \tag{A75}$$

Hence, the direct part follows from (A74). □

Appendix K.2. Converse Proof

By Lemma 1, $\mathbb{C}^{\star}(\mathcal{L}^{\mathcal{Q}}) \leq \mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}})$. Now, observe that

$$\begin{aligned}
\mathsf{R}_{CS}(\mathcal{L}^{\mathcal{Q}}) &= \min_{q(s)\,:\,\mathbb{E}S^2\leq\Lambda} \max_{\substack{p(x'')p(x,x_1)\,:\\ E(X'^2+X''^2)\leq\Omega,\\ \mathbb{E}X_1^2\leq\Omega}} \min\left\{I_q(X',X_1;Y)\,,\; I(X'';Y_1)+I_q(X';Y|X_1)\right\} \\
&\leq \max_{\substack{p(x'')p(x,x_1)\,:\\ E(X'^2+X''^2)\leq\Omega,\\ \mathbb{E}X_1^2\leq\Omega}} \min\left\{I_q(X',X_1;Y)\,,\; I(X'';Y_1)+I_q(X';Y|X_1)\right\}\Big|_{S\sim\mathcal{N}(0,\Lambda)} \\
&= \max_{0\leq\alpha,\rho\leq 1} \min\left\{\frac{1}{2}\log\left(1+\frac{\Omega_1+\rho^2\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1}}{\Lambda}\right),\right. \\
&\quad \left.\frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\sigma^2}\right)+\frac{1}{2}\log\left(1+\frac{(1-\rho^2)\alpha\Omega}{\Lambda}\right)\right\}, \qquad \text{(A76)}
\end{aligned}$$

where the last equality is due to [5]. □

Appendix L. Proof of Theorem 2

Appendix L.1. Achievability Proof

To show that $\mathbb{C}^{\star}(\mathcal{L}) \geq \mathbb{C}(\mathcal{L}^{\mathcal{Q}})$, we follow the steps in the proof of Theorem 1, where we replace Ahlswede's original RT with the modified version in [85,86] (Lemma 9), plugging $l^n(s^n) = \frac{1}{n}\sum_{i=1}^n s_i^2$. Then, by Lemma 6, it follows that

$$\begin{aligned}
\mathbb{C}^{\star}(\mathcal{L}) \geq \max_{0\leq\alpha,\rho\leq 1} \min\left\{\frac{1}{2}\log\left(1+\frac{(1+\alpha+2\rho\sqrt{\alpha})\Omega}{\Lambda}\right),\right. \\
\left.\frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\sigma^2}\right)+\frac{1}{2}\log\left(1+\frac{(1-\rho^2)\alpha\Omega}{\Lambda}\right)\right\}. \qquad \text{(A77)}
\end{aligned}$$

The details are omitted. □

Appendix L.2. Converse Proof

Assume to the contrary that there exists an achievable rate R such that

$$\begin{aligned}
R > \max_{0\leq\alpha,\rho\leq 1} \min\left\{\frac{1}{2}\log\left(1+\frac{(1+\alpha+2\rho\sqrt{\alpha})\Omega}{\Lambda-\delta}\right),\right. \\
\left.\frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\sigma^2}\right)+\frac{1}{2}\log\left(1+\frac{(1-\rho^2)\alpha\Omega}{\Lambda-\delta}\right)\right\} \qquad \text{(A78)}
\end{aligned}$$

using random codes over the Gaussian AVRC $\mathcal{L}$, under input constraints Ω and Ω_1 and state constraint Λ, where $\delta > 0$ is arbitrarily small. That is, for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR}, n)$ random code $\mathscr{C}^{\Gamma} = (\mu, \Gamma, \{\mathscr{C}_{\gamma}\}_{\gamma\in\Gamma})$ for the Gaussian AVRC $\mathcal{L}$, under input constraints Ω and Ω_1 and state constraint Λ, such that

$$P_{e|\mathbf{s}}(\mathscr{C}^{\Gamma}) \leq \varepsilon\,, \qquad \text{(A79)}$$

for all $m \in [1 : 2^{nR}]$ and $\mathbf{s} \in \mathbb{R}^n$ with $\|\mathbf{s}\|^2 \leq n\Lambda$.

Consider using the random code $\mathscr{C}^{\Gamma}$ over the Gaussian compound relay channel $\mathcal{L}^{\mathcal{Q}}$ under state constraint $(\Lambda - \delta)$, i.e., with

$$\mathcal{Q} = \{q(s) \,:\, \mathbb{E}S^2 \leq \Lambda - \delta\}\,, \tag{A80}$$

under input constraints Ω and Ω_1. Let $\overline{q}(s) \in \mathcal{Q}$ be a given state distribution. Then, define a sequence of i.i.d. random variables $\overline{S}_1, \ldots, \overline{S}_n \sim \overline{q}(s)$. Letting $\overline{q}(s^n) \triangleq \prod_{i=1}^n \overline{q}(s_i)$, the probability of error is bounded by

$$P_e^{(n)}(\overline{q}, \mathscr{C}^{\Gamma}) \leq \sum_{s^n \,:\, l^n(s^n) \leq \Lambda} \overline{q}^n(s^n) P_{e|s^n}^{(n)}(\mathscr{C}^{\Gamma}) + \Pr\left(\frac{1}{n}\sum_{i=1}^n \overline{S}_i^2 > \Lambda\right)\,. \tag{A81}$$

Then, the first sum is bounded by (A79), and the second term vanishes as well by the law of large numbers, since $\overline{q}(s)$ is in (A80). Hence, the rate R in (A78) is achievable for the Gaussian compound relay channel $\mathcal{L}^{\mathcal{Q}}$, in contradiction to Lemma 6. We deduce that the assumption is false, and (A78) cannot be achieved. □

Appendix M. Proof of Theorem 3

Consider the Gaussian AVRC $\mathcal{L}$ with SFD under input constraints Ω and Ω_1 and state constraint Λ. In the proof, we modify the techniques by Csiszár and Narayan [87]. In the direct part, we use their correlation binning technique within the decode-forward coding scheme, and in the converse part, we consider a jamming scheme which simulates the transmission sum by the encoder and the relay.

Appendix M.1. Lower Bound

We construct a block Markov code using backward minimum-distance decoding in two steps. The encoders use B blocks, each consists of n channel uses, to convey $(B-1)$ independent messages to the receiver, where 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.,

$$M_b' \sim \mathrm{Unif}[1 : 2^{nR'}]\,,\ M_b'' \sim \mathrm{Unif}[1 : 2^{nR''}]\,,\ \text{with } R' + R'' = R\,, \tag{A82}$$

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 Construction: Fix $0 \leq \alpha, \rho \leq 1$ with

$$(1-\rho^2)\alpha\Omega > \Lambda\,, \tag{A83}$$

$$\frac{\Omega_1}{\Omega}(\sqrt{\Omega_1} + \rho\sqrt{\alpha\Omega})^2 > \Lambda + (1-\rho^2)\alpha\Omega\,. \tag{A84}$$

We construct B codebooks $\mathcal{F}_b$ of the following form,

$$\mathcal{F}_b = \left\{ \left(\mathbf{x}_1(m_{b-1}'), \mathbf{x}'(m_b', m_b''|m_{b-1}'), \mathbf{x}''(m_b')\right) \,:\, m_{b-1}', m_b' \in [1 : 2^{nR'}]\,,\ m_b'' \in [1 : 2^{nR''}] \right\}, \tag{A85}$$

for $b \in [2 : B-1]$. The codebooks $\mathcal{F}_1$ and $\mathcal{F}_B$ have the same form, with fixed $m_0' = m_B' \equiv 1$ and $m_0'' = m_B'' \equiv 1$.

The sequences $\mathbf{x}''(m_b')$, $m_b' \in [1 : 2^{nR'}]$ are chosen as follows. Observe that the channel from the sender to the relay, $Y_1 = X'' + Z$, does not depend on the state. Thus, by Shannon's well-known result on the point-to-point Gaussian channel [97], the message m_b' can be conveyed to the relay reliably, under input constraint $(1-\alpha)\Omega$, provided that $R' < \frac{1}{2}\log\left(1 + \frac{(1-\alpha)\Omega}{\sigma^2}\right) - \delta_1$, where δ_1 is arbitrarily small (see also [98] (Chapter 9)). That is, for every $\varepsilon > 0$ and sufficiently large n, there exists a $(2^{nR'}, n, \varepsilon)$ code $\mathscr{C}'' = (\mathbf{x}''(m_b'), g_1(\mathbf{y}_{1,b}))$, such that $\left\|\mathbf{x}''(m_b')\right\|^2 \leq n(1-\alpha)\Omega$ for all $m_b' \in [1 : 2^{nR'}]$.

Next, we choose the sequences $\mathbf{x}_1(m'_{b-1})$ and $\mathbf{x}'(m'_b, m''_b|\, m'_{b-1})$, for m'_{b-1}, $m'_b \in [1 : 2^{nR'}]$, $m''_b \in [1 : 2^{nR''}]$. Applying Lemma 7 by [87] repeatedly yields the following.

Lemma A1. *For every $\varepsilon > 0$, $8\sqrt{\varepsilon} < \eta < 1$, $K > 2\varepsilon$, $2\varepsilon \le R' \le K$, $2\varepsilon \le R'' \le K$, and $n \ge n_0(\varepsilon, \eta, K)$,*

1. *there exist $2^{nR'}$ unit vectors,*

$$\mathbf{a}(m'_{b-1}) \in \mathbb{R}^n\,,\ m'_{b-1} \in [1 : 2^{nR'}]\,, \tag{A86}$$

such that for every unit vector $\mathbf{c} \in \mathbb{R}^n$ and $0 \le \theta, \zeta \le 1$,

$$\left|\left\{\widetilde{m}'_{b-1} \in [1 : 2^{nR'}] \,:\ \langle \mathbf{a}(\widetilde{m}'_{b-1}), \mathbf{c}\rangle \ge \theta\right\}\right| \le 2^{n\left([R' + \frac{1}{2}\log(1-\theta^2)]_+ + \varepsilon\right)}\,, \tag{A87}$$

and if $\theta \ge \eta$ and $\theta^2 + \zeta^2 > 1 + \eta - 2^{-2R'}$, then

$$\frac{1}{2^{nR'}}\big|\{m'_{b-1} \in [1 : 2^{nR'}] \,:\ |\langle \mathbf{a}(\widetilde{m}'_{b-1}), \mathbf{a}(m'_{b-1})\rangle| \ge \theta\,,\ |\langle \mathbf{a}(\widetilde{m}'_{b-1}), \mathbf{c}\rangle| \ge \zeta\,, \\ \text{for some } \widetilde{m}'_{b-1} \ne m'_{b-1}\}\big| \le 2^{-n\varepsilon}\,. \tag{A88}$$

2. *Furthermore, for every $m'_b \in [1 : 2^{nR'}]$, there exist $2^{nR''}$ unit vectors,*

$$\mathbf{v}(m'_b, m''_b) \in \mathbb{R}^n\,,\ m''_b \in [1 : 2^{nR''}]\,, \tag{A89}$$

such that for every unit vector $\mathbf{c} \in \mathbb{R}^n$ and $0 \le \theta, \zeta \le 1$,

$$\left|\left\{\widetilde{m}''_b \in [1 : 2^{nR''}] \,:\ \langle \mathbf{v}(m'_b, \widetilde{m}''_b), \mathbf{c}\rangle \ge \theta\right\}\right| \le 2^{n\left([R'' + \frac{1}{2}\log(1-\theta^2)]_+ + \varepsilon\right)}\,, \tag{A90}$$

and if $\theta \ge \eta$ and $\theta^2 + \zeta^2 > 1 + \eta - 2^{-2R''}$, then

$$\frac{1}{2^{nR''}}\big|\{m''_b \in [1 : 2^{nR''}] \,:\ |\langle \mathbf{v}(m'_b, \widetilde{m}''_b), \mathbf{v}(m'_b, m''_b)\rangle| \ge \theta\,,\ |\langle \mathbf{v}(m'_b, \widetilde{m}''_b), \mathbf{c}\rangle| \ge \zeta\,, \\ \text{for some } \widetilde{m}''_b \ne m''_b\}\big| \le 2^{-n\varepsilon}\,. \tag{A91}$$

Then, define

$$\begin{aligned} \mathbf{x}_1(m'_{b-1}) &= \sqrt{n\gamma(\Omega - \delta)} \cdot \mathbf{a}(m'_{b-1})\,, \\ \mathbf{x}'(m'_b, m''_b|m'_{b-1}) &= \rho\sqrt{\alpha\gamma^{-1}} \cdot \mathbf{x}_1(m'_{b-1}) + \beta \cdot \mathbf{v}(m'_b, m''_b)\,, \end{aligned} \tag{A92}$$

where

$$\beta \triangleq \sqrt{n(1-\rho^2)\alpha(\Omega - \delta)}\,,\ \gamma \triangleq \Omega_1/\Omega\,. \tag{A93}$$

Note that $\|\mathbf{x}_1(m'_{b-1})\|^2 = n\gamma(\Omega - \delta) < n\Omega_1$, for all $m'_{b-1} \in [1 : 2^{nR'}]$. On the other hand, $\|\mathbf{x}'(m'_b, m''_b|m'_{b-1})\|^2$ could be greater than $n\alpha\Omega$ due to the possible correlation between $\mathbf{x}_1(m'_{b-1})$ and $\mathbf{v}(m'_b, m''_b)$.

Encoding: Let $(m'_1, m''_1, \ldots, m'_{B-1}, m''_{B-1})$ be a sequence of messages to be sent. In block $b \in [1 : B]$, if $\|\mathbf{x}'(m'_b, m''_b|m'_{b-1})\|^2 \le n\alpha\Omega$, transmit $(\mathbf{x}'(m'_b, m''_b|m'_{b-1}), \mathbf{x}''(m'_b))$. Otherwise, transmit $(\mathbf{0}, \mathbf{x}''(m'_b))$.

Relay Encoding: In block 1, the relay transmits $\mathbf{x}_1(1)$. At the end of block $b \in [1 : B-1]$, the relay receives $\mathbf{y}_{1,b}$, and finds an estimate $\bar{m}'_b = g_1(\mathbf{y}_{1,b})$. In block $b+1$, the relay transmits $\mathbf{x}_1(\bar{m}'_b)$.

Backward Decoding: Once all blocks $(\mathbf{y}_b)_{b=1}^B$ are received, decoding is performed backwards. Set $\hat{m}'_0 = \hat{m}''_0 \equiv 1$. For $b = B-1, B-2, \ldots, 1$, find a unique $\hat{m}'_b \in [1 : 2^{nR'}]$ such that

$$\left\| \mathbf{y}_{b+1} - (1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(\hat{m}'_b) \right\| \leq \left\| \mathbf{y}_{b+1} - (1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(m'_b) \right\| , \text{ for all } m'_b \in [1 : 2^{nR}] . \quad \text{(A94)}$$

If there is 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

$$\|\mathbf{y}_b - \mathbf{x}_1(\hat{m}'_{b-1}) - \mathbf{x}'(\hat{m}'_b, \hat{m}''_b|\hat{m}'_{b-1})\| \leq \|\mathbf{y}_b - \mathbf{x}_1(\hat{m}'_{b-1}) - \mathbf{x}'(\hat{m}'_b, m''_b|\hat{m}'_{b-1})\| , \text{ for all } m''_b \in [1 : 2^{nR''}] . \quad \text{(A95)}$$

If there is more than one such $\hat{m}''_b \in [1 : 2^{nR''}]$, declare an error.

Analysis of Probability of Error: Fix $\mathbf{s} \in \mathcal{S}^n$, and let

$$\mathbf{c}_0 \triangleq \frac{\mathbf{s}}{\|\mathbf{s}\|} . \quad \text{(A96)}$$

The error event is bounded by the union of the following events. For $b \in [1 : B-1]$, define

$$\mathcal{E}_1(b) = \{\bar{M}'_b \neq M'_b\} , \ \mathcal{E}_2(b) = \{\hat{M}'_b \neq M'_b\} , \ \mathcal{E}_3(b) = \{\hat{M}''_b \neq M''_b\} . \quad \text{(A97)}$$

Then, the conditional probability of error given the state sequence $\mathbf{s}$ is bounded by

$$P_{e|\mathbf{s}}(\mathscr{C}) \leq \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_1(b)\right) + \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_2(b) \cap \mathcal{E}_1^c(b)\right) + \sum_{b=1}^{B-1} \Pr\left(\mathcal{E}_3(b) \cap \mathcal{E}_1^c(b-1) \cap \mathcal{E}_2^c(b) \cap \mathcal{E}_2^c(b-1)\right) , \quad \text{(A98)}$$

with $\mathcal{E}_1(0) = \mathcal{E}_2(0) = \emptyset$, where the conditioning on $\mathbf{S} = \mathbf{s}$ is omitted for convenience of notation. Recall that we have defined $\mathscr{C}''$ as a $(2^{nR'}, n, \varepsilon)$ code for the point-to-point Gaussian channel $Y_1 = X'' + Z$. Hence, the first sum in the RHS of (A98) is bounded by $B \cdot \varepsilon$, which is arbitrarily small.

To be more concise, we only give the details for erroneous decoding of M'_b at the receiver. Consider the following events,

$$\begin{aligned}
\mathcal{E}_2(b) &= \{\left\| \mathbf{Y}_{b+1} - (1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(\widetilde{m}'_b) \right\| \leq \left\| \mathbf{Y}_{b+1} - (1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(M'_b) \right\| , \text{ for some } \widetilde{m}'_b \neq M'_b\} , \\
\mathcal{E}_{2,1}(b) &= \{|\langle \mathbf{a}(M'_b), \mathbf{c}_0 \rangle| \geq \eta\} , \\
\mathcal{E}_{2,2}(b) &= \{|\langle \mathbf{a}(M'_b), \mathbf{v}(M_{b+1}) \rangle| \geq \eta\} \\
\mathcal{E}_{2,3}(b) &= \{|\langle \mathbf{v}(M_{b+1}), \mathbf{c}_0 \rangle| \geq \eta\} \\
\widetilde{\mathcal{E}}_2(b) &= \mathcal{E}_2(b) \cap \mathcal{E}_1^c(b) \cap \mathcal{E}_{2,1}^c(b) \cap \mathcal{E}_{2,2}^c(b) \cap \mathcal{E}_{2,3}^c(b) ,
\end{aligned} \quad \text{(A99)}$$

where $M_{b+1} = (M'_{b+1}, M''_{b+1})$. Then,

$$\begin{aligned}
\mathcal{E}_2(b) \cap \mathcal{E}_1^c(b) \subseteq & \mathcal{E}_{2,1}(b) \cup \mathcal{E}_{2,2}(b) \cup \mathcal{E}_{2,3}(b) \cup (\mathcal{E}_2(b) \cap \mathcal{E}_1^c(b)) \\
= & \mathcal{E}_{2,1}(b) \cup \mathcal{E}_{2,2}(b) \cup \mathcal{E}_{2,3}(b) \cup \widetilde{\mathcal{E}}_2(b) .
\end{aligned} \quad \text{(A100)}$$

Hence, by the union of events bound, we have that

$$\Pr\left(\mathcal{E}_2(b) \cap \mathcal{E}_1^c(b)\right) \leq \Pr\left(\mathcal{E}_{2,1}(b)\right) + \Pr\left(\mathcal{E}_{2,2}(b)\right) + \Pr\left(\mathcal{E}_{2,3}(b)\right) + \Pr\left(\widetilde{\mathcal{E}}_2(b)\right) . \quad \text{(A101)}$$

By Lemma A1, given $R' > -\frac{1}{2}\log(1-\eta^2)$, the first term is bounded by

$$\begin{aligned}\Pr\left(\mathcal{E}_{2,1}(b)\right) &= \Pr\left(\langle \mathbf{a}(M'_b), \mathbf{c}_0\rangle \geq \eta\right) + \Pr\left(\langle \mathbf{a}(M'_b), -\mathbf{c}_0\rangle \geq \eta\right) \\ &\leq 2\cdot\frac{1}{2^{nR'}}\cdot 2^{n\left(R'+\frac{1}{2}\log(1-\eta^2)+\varepsilon\right)} \leq 2\cdot 2^{n\left(-\frac{1}{2}\eta^2+\varepsilon\right)}, \end{aligned} \tag{A102}$$

since $\log(1+t) \leq t$ for $t \in \mathbb{R}$. As $\eta^2 \geq 8\varepsilon$, the last expression tends to zero as $n \to \infty$. Similarly, $\Pr\left(\mathcal{E}_{2,2}(b)\right)$ and $\Pr\left(\mathcal{E}_{2,3}(b)\right)$ tend to zero as well. Moving to the fourth term in the RHS of (A101), observe that for a sufficiently small ε and η, the event $\mathcal{E}^c_{2,2}(b)$ implies that $\left\|\mathbf{x}'(M_{b+1}|M'_b)\right\|^2 \leq n\alpha\Omega$, while the event $\mathcal{E}^c_1(b)$ means that $\bar{M}'_b = M'_b$. Hence, the encoder transmits $(\mathbf{x}'(M_{b+1}|M'_b), \mathbf{x}''(M'_{b+1}))$, the relay transmits $\mathbf{x}_1(M'_b)$, and we have that

$$\begin{aligned}&\left\|\mathbf{Y}_{b+1} - (1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(\widetilde{m}'_b)\right\|^2 - \left\|\mathbf{Y}_{b+1} - (1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(M'_b)\right\|^2 \\ &= \left\|(1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(M'_b) + \beta\mathbf{v}(M_{b+1}) + \mathbf{s} - (1+\rho\sqrt{\alpha\gamma^{-1}})\mathbf{x}_1(\widetilde{m}'_b)\right\|^2 - \left\|\beta\mathbf{v}(M_{b+1}) + \mathbf{s}\right\|^2 \\ &= 2(1+\rho\sqrt{\alpha\gamma^{-1}})^2\left(\frac{1}{2}\left\|\mathbf{x}_1(M'_b)\right\|^2 + \frac{1}{2}\left\|\mathbf{x}_1(\widetilde{m}'_b)\right\|^2 - \langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle\right) \\ &\quad + 2(1+\rho\sqrt{\alpha\gamma^{-1}})\left(\langle \mathbf{x}_1(M'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle - \langle \mathbf{x}_1(\widetilde{m}'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle\right)\end{aligned} \tag{A103}$$

Then, since $\left\|\mathbf{x}_1(m'_b)\right\|^2 = n\gamma(\Omega-\delta)$ for all $m'_b \in [1 : 2^{nR'}]$, we have that

$$\begin{aligned}\mathcal{E}_2(b)\cap\mathcal{E}^c_1(b)\cap\mathcal{E}^c_{2,2}(b) \subseteq &\{(1+\rho\sqrt{\alpha\gamma^{-1}})\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle + \langle \mathbf{x}_1(\widetilde{m}'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle \geq \\ &n(1+\rho\sqrt{\alpha\gamma^{-1}})\gamma(\Omega-\delta) + \langle \mathbf{x}_1(M'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle, \text{ for some } \widetilde{m}'_b \neq M'_b\}. \end{aligned} \tag{A104}$$

Observe that for sufficiently small ε and η, the event $\mathcal{E}^c_{2,1}(b)\cap\mathcal{E}^c_{2,2}(b)\cap\mathcal{E}^c_{2,3}(b)$ implies that

$$\langle \mathbf{x}_1(M'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle \geq -\delta, \tag{A105}$$

and

$$\left\|\beta\mathbf{v}(M_{b+1}) + \mathbf{s}\right\|^2 \leq n[(1-\rho^2)\alpha\Omega + \Lambda]. \tag{A106}$$

Hence, by (A104) and (A105),

$$\begin{aligned}\widetilde{\mathcal{E}}_2(b) &= \mathcal{E}_2(b)\cap\mathcal{E}^c_1(b)\cap\mathcal{E}^c_{2,1}(b)\cap\mathcal{E}^c_{2,2}(b)\cap\mathcal{E}^c_{2,3}(b) \\ \subseteq &\{(1+\rho\sqrt{\alpha\gamma^{-1}})\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle + \langle \mathbf{x}_1(\widetilde{m}'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle \geq n(1+\rho\sqrt{\alpha\gamma^{-1}})\gamma(\Omega-2\delta), \\ &\text{for some } \widetilde{m}'_b \neq M'_b\}. \end{aligned} \tag{A107}$$

Dividing both sides of the inequality by $n(1+\rho\sqrt{\alpha\gamma^{-1}})$, we obtain

$$\widetilde{\mathcal{E}}_2(b) \subseteq \left\{\frac{1}{n}\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle + \frac{\langle \mathbf{x}_1(\widetilde{m}'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle}{n(1+\rho\sqrt{\alpha\gamma^{-1}})} \geq \gamma(\Omega-2\delta), \text{ for some } \widetilde{m}'_b \neq M'_b\right\}. \tag{A108}$$

Next, we partition the set of values of $\frac{1}{n}\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle$ to K bins. Let $\tau_1 < \tau_2 < \cdots < \tau_K$ be such partition, where

$$\tau_1 = \gamma(\Omega - 2\delta) - \frac{\sqrt{(\Omega-\delta)[(1-\rho^2)\alpha\Omega + \Lambda]}}{1+\rho\sqrt{\alpha\gamma^{-1}}}\,,\ \tau_K = \gamma(\Omega - 3\delta)\,,$$
$$\tau_{k+1} - \tau_k \le \gamma\cdot\delta\,,\ \text{for } k = [1:K-1]\,, \tag{A109}$$

where K is a finite constant which is independent of n, as in Lemma A1. By (A106) and (A108), given the event $\widetilde{\mathcal{E}}_2(b)$, we have that

$$\frac{1}{n}\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle \ge \tau_1 > 0\,, \tag{A110}$$

where the last inequality is due to (A84), for sufficiently small $\delta > 0$. To see this, observe that the inequality in (A84) is strict, and it implies that

$$\sqrt{\gamma}\cdot(\sqrt{\gamma\Omega} + \rho\sqrt{\alpha\Omega}) > \sqrt{(1-\rho^2)\alpha\Omega + \Lambda}\,. \tag{A111}$$

Hence, for sufficiently small $\delta > 0$, $\tau_1 > 0$ as

$$\tau_1 = \frac{\sqrt{\Omega - 2\delta}}{1+\rho\sqrt{\alpha\gamma^{-1}}}\cdot\left(\sqrt{\gamma}(\sqrt{\gamma(\Omega-2\delta)} + \rho\sqrt{\alpha(\Omega-2\delta)}) - \sqrt{\frac{\Omega-\delta}{\Omega-2\delta}[(1-\rho^2)\alpha\Omega+\Lambda]}\right)\,. \tag{A112}$$

Furthermore, if $\tau_k \le \frac{1}{n}\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle < \tau_{k+1}$, then

$$\frac{\langle \mathbf{x}_1(\widetilde{m}'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle}{n(1+\rho\sqrt{\alpha\gamma^{-1}})} \ge \gamma(\Omega-2\delta) - \tau_{k+1} \ge \gamma(\Omega-3\delta) - \tau_k\,. \tag{A113}$$

Thus,

$$\Pr\left(\widetilde{\mathcal{E}}_2(b)\right) \le \sum_{k=1}^{K-1}\Pr\left(\frac{1}{n}|\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle| \ge \tau_k\,,\ \frac{|\langle \mathbf{x}_1(\widetilde{m}'_b), \beta\mathbf{v}(M_{b+1}) + \mathbf{s}\rangle|}{n(1+\rho\sqrt{\alpha\gamma^{-1}})} \ge \gamma(\Omega-3\delta) - \tau_k\,,\right.$$
$$\left.\text{for some } \widetilde{m}'_b \ne M'_b\right) + \Pr\left(\frac{1}{n}|\langle \mathbf{x}_1(\widetilde{m}'_b), \mathbf{x}_1(M'_b)\rangle| \ge \tau_K\,,\ \text{for some } \widetilde{m}'_b \ne M'_b\right)\,. \tag{A114}$$

By (A106), this can be further bounded by

$$\Pr\left(\widetilde{\mathcal{E}}_2(b)\right) \le \sum_{k=1}^{K}\Pr\left(|\langle \mathbf{a}(\widetilde{m}'_b), \mathbf{a}(M'_b)\rangle| \ge \theta_k\,,\ |\langle \mathbf{a}(\widetilde{m}'_b), \mathbf{c}'(M_{b+1})\rangle| \ge \mu_k\,,\ \text{for some } \widetilde{m}'_b \ne M'_b\right)\,, \tag{A115}$$

where

$$\mathbf{c}'(m_{b+1}) \triangleq \frac{\beta\mathbf{v}(m_{b+1}) + \mathbf{s}}{\|\beta\mathbf{v}(m_{b+1}) + \mathbf{s}\|}\,, \tag{A116}$$

and

$$\theta_k \triangleq \frac{\tau_k}{\gamma(\Omega-\delta)}\,,\ \zeta_k \triangleq \frac{(1+\rho\sqrt{\alpha\gamma^{-1}})\,(\gamma(\Omega-3\delta) - \tau_k)}{\sqrt{\gamma(\Omega-\delta)((1-\rho^2)\alpha\Omega+\Lambda)}}\,,\ \text{for } k\in[1:K-1];\ \theta_K \triangleq \frac{\tau_K}{\Omega-\delta}\,,\ \zeta_K = 0\,. \tag{A117}$$

By Lemma A1, the RHS of (A115) tends to zero as $n\to\infty$ provided that

$$\theta_k \ge \eta \text{ and } \theta_k^2 + \zeta_k^2 > 1 + \eta - e^{-2R'}\,,\ \text{for } k = [1:K]\,. \tag{A118}$$

For sufficiently small ε and η, we have that $\eta \leq \theta_1 = \frac{\tau_1}{\gamma(\Omega-\delta)}$, hence the first condition is met. Then, observe that the second condition is equivalent to $G(\tau_k) > 1+\eta-e^{-2R'}$, for $k \in [1:K-1]$, where

$$G(\tau) = (A\tau)^2 + D^2(L-\tau)^2\,, \tag{A119}$$

with

$$A = \frac{1}{\gamma(\Omega-\delta)}\,,\ D = \frac{1+\rho\sqrt{\alpha\gamma^{-1}}}{\sqrt{\gamma(\Omega-\delta)((1-\rho^2)\alpha\Omega+\Lambda)}}\,,\ L = \gamma(\Omega-3\delta)\,. \tag{A120}$$

By differentiation, we have that the minimum value of this function is given by $\min_{\tau_1\leq\tau\leq\tau_K} G(\tau) = \frac{A^2D^2L^2}{A^2+D^2} = \frac{D^2}{A^2+D^2} - \delta_1$, where $\delta_1 \to 0$ as $\delta \to 0$. Thus, the RHS of (A115) tends to zero as $n \to \infty$, provided that

$$\begin{aligned} R' &< -\frac{1}{2}\log\left(1+\eta-\frac{D^2}{A^2+D^2}+\delta_1\right) \\ &= -\frac{1}{2}\log\left(\eta+\delta_1+\frac{(1-\rho^2)\alpha\Omega+\Lambda}{(\gamma+\alpha+2\rho\sqrt{\alpha\gamma})\Omega+\Lambda-\delta\gamma(1+\rho\sqrt{\alpha\gamma^{-1}})^2}\right)\,. \end{aligned} \tag{A121}$$

This is satisfied for $R' = \mathsf{R}'_\alpha(\mathcal{L}) - \delta'$, with

$$\mathsf{R}'_\alpha(\mathcal{L}) = \frac{1}{2}\log\left(\frac{(\gamma+\alpha+2\rho\sqrt{\alpha\gamma})\Omega+\Lambda}{(1-\rho^2)\alpha\Omega+\Lambda}\right) = -\frac{1}{2}\log\left(\frac{(1-\rho^2)\alpha\Omega+\Lambda}{(\gamma+\alpha+2\rho\sqrt{\alpha\gamma})\Omega+\Lambda}\right)\,. \tag{A122}$$

and arbitrary $\delta' > 0$, if η and δ are sufficiently small.

As for the error event for M''_b, a similar derivation shows that the probability term in the last sum in (A98) exponentially tends to zero as $n \to \infty$, provided that

$$R'' < -\frac{1}{2}\log\left(1+\eta-\frac{\beta^2(1-2\delta)^2}{\beta^2+n\Lambda}\right) < -\frac{1}{2}\log\left(\eta+\frac{\Lambda}{(1-\rho^2)\alpha(\Omega-\delta)+\Lambda}\right)\,. \tag{A123}$$

This is satisfied for $R'' = \mathsf{R}''_\alpha(\mathcal{L}) - \delta''$, with

$$\mathsf{R}''_\alpha(\mathcal{L}) = \frac{1}{2}\log\left(\frac{(1-\rho^2)\alpha\Omega+\Lambda}{\Lambda}\right) = -\frac{1}{2}\log\left(\frac{\Lambda}{(1-\rho^2)\alpha\Omega+\Lambda}\right) \tag{A124}$$

for an arbitrary $\delta'' > 0$, if η and δ are sufficiently small.

We have thus shown achievability of every rate

$$R < \min\left\{\mathsf{R}'_\alpha(\mathcal{L}) + \mathsf{R}''_\alpha(\mathcal{L}), \frac{1}{2}\log\left(1+\frac{(1-\alpha)\Omega}{\sigma^2}\right) + \mathsf{R}''_\alpha(\mathcal{L})\right\}\,, \tag{A125}$$

where

$$\begin{aligned} \mathsf{R}'_\alpha(\mathcal{L}) + \mathsf{R}''_\alpha(\mathcal{L}) &= \frac{1}{2}\log\left(\frac{(\gamma+\alpha+2\rho\sqrt{\alpha\gamma})\Omega+\Lambda}{(1-\rho^2)\alpha\Omega+\Lambda}\right) + \frac{1}{2}\log\left(\frac{(1-\rho^2)\alpha\Omega+\Lambda}{\Lambda}\right) \\ &= \frac{1}{2}\log\left(1+\frac{\Omega_1+\alpha\Omega+2\rho\sqrt{\alpha\Omega\cdot\Omega_1}}{\Lambda}\right) \end{aligned} \tag{A126}$$

(see (A93)). This completes the proof of the lower bound.

Appendix M.2. Upper Bound

Let $R > 0$ be an achievable rate. Then, there exists a sequence of $(2^{nR}, n, \varepsilon_n^*)$ codes $\mathscr{C}_n = (\mathbf{f}, \mathbf{f}_1, g)$ for the Gaussian AVRC $\mathcal{L}$ with SFD such that $\varepsilon_n^* \to 0$ as $n \to \infty$, where the encoder consists of a pair $\mathbf{f} = (\mathbf{f}', \mathbf{f}'')$, with $\mathbf{f}' : [1 : 2^{nR}] \to \mathbb{R}^n$ and $\mathbf{f}'' : [1 : 2^{nR}] \to \mathbb{R}^n$. Assume without loss of generality that the codewords have zero mean, i.e.,

$$\begin{aligned}
&\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \frac{1}{n} \sum_{i=1}^{n} \mathrm{f}_i(m) = 0\,, \\
&\int_{-\infty}^{\infty} d\mathbf{y}_1 \cdot \frac{1}{2^{nR}} \sum_{m \in [1:2^{nR}]} P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m) \cdot \frac{1}{n} \sum_{i=1}^{n} \mathrm{f}_{1,i}(y_{1,1}, y_{1,2}, \ldots, y_{1,i-1}) = 0\,,
\end{aligned} \tag{A127}$$

where $P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\|\mathbf{y}_1 - \mathbf{f}''(m)\|^2/2\sigma^2}$. If this is not the case, redefine the code such that the mean is subtracted from each codeword. Then, define

$$\begin{aligned}
\alpha &\triangleq \frac{1}{n\Omega} \cdot \frac{1}{2^{nR}} \sum_{m \in [1:2^{nR}]} \left\|\mathbf{f}'(m)\right\|^2\,, \\
\alpha_1 &\triangleq \frac{1}{n\Omega_1} \cdot \frac{1}{2^{nR}} \sum_{m \in [1:2^{nR}]} \int_{-\infty}^{\infty} d\mathbf{y}_1 \cdot P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m) \cdot \left\|\mathbf{f}_1(\mathbf{y}_1)\right\|^2\,. \\
\rho &\triangleq \frac{1}{n\sqrt{\alpha\Omega \cdot \alpha_1\Omega_1}} \int_{-\infty}^{\infty} d\mathbf{y}_1 \cdot \frac{1}{2^{nR}} \sum_{m \in [1:2^{nR}]} P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m) \cdot \langle \mathbf{f}'(m), \mathbf{f}_1(\mathbf{y}_1) \rangle\,,
\end{aligned} \tag{A128}$$

Since the code satisfies the input constraints Ω and Ω_1, we have that α, α_1 and ρ are in the interval $[0, 1]$.

First, we show that if

$$\Lambda > \Omega_1 + \alpha\Omega + 2\rho\sqrt{\alpha\Omega \cdot \Omega_1} + \delta\,, \tag{A129}$$

then the capacity is zero, where $\delta > 0$ is arbitrarily small. Consider the following jamming strategy. The jammer draws a message $\widetilde{M} \in [1 : 2^{nR}]$ uniformly at random, and then, generates a sequence $\widetilde{\mathbf{Y}}_1 \in \mathbb{R}^n$ distributed according to $P_{\mathbf{Y}_1|M}(\tilde{\mathbf{y}}_1|\widetilde{m})$. Let $\widetilde{\mathbf{S}} = \mathbf{f}'(\widetilde{M}) + \mathbf{f}_1(\widetilde{\mathbf{Y}}_1)$. If $\frac{1}{n}\left\|\widetilde{\mathbf{S}}\right\|^2 \leq \Lambda$, the jammer chooses $\widetilde{\mathbf{S}}$ to be the state sequence. Otherwise, let the state sequence consist of all zeros. Observe that

$$\begin{aligned}
\mathbb{E}\left\|\widetilde{\mathbf{S}}\right\|^2 =& \mathbb{E}\left\|\mathbf{f}'(\widetilde{M}) + \mathbf{f}_1(\widetilde{\mathbf{Y}}_1)\right\|^2 \\
=& \mathbb{E}\left\|\mathbf{f}'(\widetilde{M})\right\|^2 + \mathbb{E}\left\|\mathbf{f}_1(\widetilde{\mathbf{Y}}_1)\right\|^2 + 2\mathbb{E}\langle \mathbf{f}(\widetilde{M}), \mathbf{f}_1(\widetilde{\mathbf{Y}}_1)\rangle \\
=& n(\alpha\Omega + \alpha_1\Omega_1 + 2\rho\sqrt{\alpha\Omega \cdot \alpha_1\Omega_1}) \\
\leq& n(\alpha\Omega + \Omega_1 + 2\rho\sqrt{\alpha\Omega \cdot \Omega_1}) < n(\Lambda - \delta)\,.
\end{aligned} \tag{A130}$$

where the second equality is due to (A128), and the last inequality is due to (A129). Thus, by Chebyshev's inequality, there exists $\kappa > 0$ such that

$$\Pr\left(\frac{1}{n}\left\|\widetilde{\mathbf{S}}\right\|^2 \leq \Lambda\right) \geq \kappa\,. \tag{A131}$$

The state sequence $\mathbf{S}$ is then distributed according to

$$P_{\mathbf{S}|\{\frac{1}{n}\|\widetilde{\mathbf{S}}\|^2\leq\Lambda\}}(\mathbf{s}) = \frac{1}{2^{nR}}\sum_{\widetilde{m}\in[1:2^{nR}]}\int_{\tilde{\mathbf{y}}_1:\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)=\mathbf{s}} d\mathbf{y}_1 P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m)\,,$$
$$\Pr\left(\mathbf{S}=\mathbf{0}\mid\frac{1}{n}\left\|\widetilde{\mathbf{S}}\right\|^2>\Lambda\right)=1\,. \tag{A132}$$

Assume to the contrary that a positive rate can be achieved when the channel is governed by such state sequence, hence the size of the message set is at least 2, i.e., $\mathsf{M}\triangleq 2^{nR}\geq 2$. The probability of error is then bounded by

$$\begin{aligned}P_e^{(n)}(q,\mathscr{C}) &= \int_{-\infty}^{\infty} d\mathbf{s}\cdot q(\mathbf{s})P_{e|\mathbf{s}}^{(n)}(\mathscr{C}) \geq \Pr\left(\frac{1}{n}\left\|\widetilde{\mathbf{S}}\right\|^2\leq\Lambda\right)\cdot\int_{\mathbf{s}:\frac{1}{n}\|\mathbf{s}\|^2\leq\Lambda} d\mathbf{s}\cdot P_{\mathbf{S}|\{\frac{1}{n}\|\widetilde{\mathbf{S}}\|^2\leq\Lambda\}}(\mathbf{s})\cdot P_{e|\mathbf{s}}^{(n)}(\mathscr{C})\\ &\geq\kappa\cdot\int_{\mathbf{s}:\frac{1}{n}\|\mathbf{s}\|^2\leq\Lambda} d\mathbf{s}\cdot P_{\mathbf{S}|\{\frac{1}{n}\|\widetilde{\mathbf{S}}\|^2\leq\Lambda\}}(\mathbf{s})\cdot P_{e|\mathbf{s}}^{(n)}(\mathscr{C})\end{aligned} \tag{A133}$$

where the inequality holds by (A131). Next, we have that

$$P_{e|\mathbf{s}}^{(n)}(\mathscr{C}) = \frac{1}{\mathsf{M}}\sum_{m=1}^{\mathsf{M}}\int_{-\infty}^{\infty} d\mathbf{y}_1\cdot P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m)\cdot\mathbb{1}\left\{\mathbf{y}_1: g(\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)+\mathbf{s})\neq m\right\}\,, \tag{A134}$$

where we define the indicator function $G(\mathbf{y}_1)=\mathbb{1}\{\mathbf{y}_1\in\mathcal{A}\}$ such that $G(\mathbf{y}_1)=1$ if $\mathbf{y}_1\in\mathcal{A}$, and $G(\mathbf{y}_1)=0$ otherwise. Substituting (A132) and (A134) into (A133) yields

$$\begin{aligned}P_e^{(n)}(q,\mathscr{C}) \geq&\kappa\cdot\int_{\mathbf{s}:\frac{1}{n}\|\mathbf{s}\|^2\leq\Lambda} d\mathbf{s}\cdot\frac{1}{\mathsf{M}}\sum_{\widetilde{m}=1}^{\mathsf{M}}\int_{\tilde{\mathbf{y}}_1:\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)=\mathbf{s}} d\tilde{\mathbf{y}}_1\cdot P_{\mathbf{Y}_1|M}(\tilde{\mathbf{y}}_1|\widetilde{m})\\ &\times\frac{1}{\mathsf{M}}\sum_{m=1}^{\mathsf{M}}\int_{-\infty}^{\infty} d\mathbf{y}_1\cdot P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m)\cdot\mathbb{1}\left\{\mathbf{y}_1: g(\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)+\mathbf{s})\neq m\right\}\,.\end{aligned} \tag{A135}$$

Eliminating $\mathbf{s}=\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)$, and adding the constraint $\|\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)\|^2\leq\Lambda$, we obtain the following,

$$\begin{aligned}P_e^{(n)}(q,\mathscr{C}) \geq&\frac{\kappa}{\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}=1}^{\mathsf{M}}\int_{\substack{(\mathbf{y}_1\,,\,\tilde{\mathbf{y}}_1):\,\frac{1}{n}\|\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)\|^2\leq\Lambda,\\ \frac{1}{n}\|\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)\|^2\leq\Lambda}} d\mathbf{y}_1\, d\tilde{\mathbf{y}}_1\cdot P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m)P_{\mathbf{Y}_1|M}(\tilde{\mathbf{y}}_1|\widetilde{m})\\ &\times\mathbb{1}\left\{\mathbf{y}_1: g(\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)+\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1))\neq m\right\}\,.\end{aligned} \tag{A136}$$

Now, by interchanging the summation variables $(m,\mathbf{y}_1)$ and $(\widetilde{m},\tilde{\mathbf{y}}_1)$, we have that

$$\begin{aligned}P_e^{(n)}(q,\mathscr{C}) \geq&\frac{\kappa}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}=1}^{\mathsf{M}}\int_{\substack{(\mathbf{y}_1\,,\,\tilde{\mathbf{y}}_1):\,\frac{1}{n}\|\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)\|^2\leq\Lambda,\\ \frac{1}{n}\|\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)\|^2\leq\Lambda}} d\mathbf{y}_1\, d\tilde{\mathbf{y}}_1\cdot P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m)P_{\mathbf{Y}_1|M}(\tilde{\mathbf{y}}_1|\widetilde{m})\\ &\times\mathbb{1}\left\{\mathbf{y}_1: g(\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)+\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1))\neq m\right\}\\ &+\frac{\kappa}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}=1}^{\mathsf{M}}\int_{\substack{(\mathbf{y}_1\,,\,\tilde{\mathbf{y}}_1):\,\frac{1}{n}\|\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)\|^2\leq\Lambda,\\ \frac{1}{n}\|\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)\|^2\leq\Lambda}} d\mathbf{y}_1\, d\tilde{\mathbf{y}}_1\cdot P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m)P_{\mathbf{Y}_1|M}(\tilde{\mathbf{y}}_1|\widetilde{m})\\ &\times\mathbb{1}\left\{\mathbf{y}_1: g(\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)+\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1))\neq \widetilde{m}\right\}\,.\end{aligned} \tag{A137}$$

Thus,

$$P_e^{(n)}(q,\mathscr{C}) \geq \frac{\kappa}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}\neq m}\int_{\substack{(\mathbf{y}_1\,,\,\tilde{\mathbf{y}}_1):\,\frac{1}{n}\|\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)\|^2\leq\Lambda\,,\\ \frac{1}{n}\|\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)\|^2\leq\Lambda}} d\mathbf{y}_1\, d\tilde{\mathbf{y}}_1\cdot P_{\mathbf{Y}_1|M}(\mathbf{y}_1|m)P_{\mathbf{Y}_1|M}(\tilde{\mathbf{y}}_1|\widetilde{m})$$
$$\times\Big[\mathbb{1}\left\{\mathbf{y}_1: g(\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)+\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1))\neq m\right\}$$
$$+\mathbb{1}\left\{\mathbf{y}_1: g(\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)+\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1))\neq \widetilde{m}\right\}\Big]. \tag{A138}$$

As the sum in the square brackets is at least 1 for all $\widetilde{m}\neq m$, it follows that

$$P_e^{(n)}(q,\mathscr{C}) \geq \frac{\kappa}{2\mathsf{M}^2}\sum_{m=1}^{\mathsf{M}}\sum_{\widetilde{m}\neq m}\int_{\substack{(\mathbf{y}_1\,,\,\tilde{\mathbf{y}}_1):\,\frac{1}{n}\|\mathbf{f}'(m)+\mathbf{f}_1(\mathbf{y}_1)\|^2\leq\Lambda\,,\\ \frac{1}{n}\|\mathbf{f}'(\widetilde{m})+\mathbf{f}_1(\tilde{\mathbf{y}}_1)\|^2\leq\Lambda}} d\mathbf{y}_1\, d\tilde{\mathbf{y}}_1\cdot P_{\mathbf{Y}_1|M}(y_1^n|m)P_{\mathbf{Y}_1|M}(\tilde{\mathbf{y}}_1|\widetilde{m})$$
$$\geq\frac{\kappa}{4}\cdot\Pr\left(\begin{array}{l}\frac{1}{n}\left\|\mathbf{f}'(M)+\mathbf{f}_1(\mathbf{Y}_1)\right\|^2\leq\Lambda\,,\\ \frac{1}{n}\left\|\mathbf{f}'(\widetilde{M})+\mathbf{f}_1(\widetilde{\mathbf{Y}}_1)\right\|^2\leq\Lambda\,,\ \widetilde{M}\neq M\end{array}\right). \tag{A139}$$

Then, recall that by (A130), the expectation of $\frac{1}{n}\left\|\mathbf{f}'(M)+\mathbf{f}_1(Y_1^n)\right\|^2$ is strictly lower than Λ, and for a sufficiently large n, the conditional expectation of $\frac{1}{n}||\mathbf{f}'(\widetilde{M})+\mathbf{f}_1(\widetilde{\mathbf{Y}}_1)||^2$ given $\widetilde{M}\neq M$ is also strictly lower than Λ. Thus, by Chebyshev's inequality, the probability of error is bounded from below by a positive constant. Following this contradiction, we deduce that if the code is reliable, then $\Lambda\leq(1+\alpha+2\rho\sqrt{\alpha})\Omega$.

It is left for us to show that for α and ρ as defined in (A128), we have that $R<\mathsf{F}_G(\alpha,\rho)$ (see (36)). For a $(2^{nR},n,\varepsilon_n^*)$ code,

$$P_{e|\mathbf{s}}^{(n)}(\mathscr{C})\leq\varepsilon_n^*, \tag{A140}$$

for all $\mathbf{s}\in\mathbb{R}^n$ with $\|\mathbf{s}\|^2\leq n\Lambda$. Then, consider using the code $\mathscr{C}$ over the Gaussian relay channel $W_{Y,Y_1|X,X_1}^{\overline{q}}$, specified by

$$\begin{aligned}Y_1&=X''+Z\,,\\ Y&=X'+X_1+\overline{S}\,,\end{aligned} \tag{A141}$$

where the sequence $\overline{\mathbf{S}}$ is i.i.d. $\sim\overline{q}=\mathcal{N}(0,\Lambda-\delta)$. First, we show that the code $\mathscr{C}$ is reliable for this channel, and then we show that $R<\mathsf{F}_G(\alpha,\rho)$. Using the code $\mathscr{C}$ over the channel $W_{Y,Y_1|X,X_1}^{\overline{q}}$, the probability of error is bounded by

$$P_e^{(n)}(\overline{q},\mathscr{C})=\Pr\left(\frac{1}{n}\left\|\overline{\mathbf{S}}\right\|>\Lambda\right)+\int_{\mathbf{s}:\frac{1}{n}\|\overline{\mathbf{s}}\|\leq\Lambda}d\mathbf{s}\cdot P_{e|\mathbf{s}}^{(n)}(\mathscr{C})\leq\varepsilon_n^*+\varepsilon_n^{**}, \tag{A142}$$

where we have bounded the first term by ε_n^{**} using the law of large numbers and the second term using (A140), where $\varepsilon_n^{**}\to 0$ as $n\to\infty$. Since $W_{Y,Y_1|X,X_1}^{\overline{q}}$ is a channel without a state, we can now show that $R<\mathsf{F}_G(\alpha,\rho)$ by following the lines of [4] and [5]. By Fano's inequality and [4] (Lemma 4), we have that

$$\begin{aligned}R&\leq\frac{1}{n}\sum_{i=1}^{n}I_{\overline{q}}(X_i',X_i'',X_{1,i};Y_i)+\varepsilon_n\,,\\ R&\leq\frac{1}{n}\sum_{i=1}^{n}I_{\overline{q}}(X_i',X_i'';Y_i,Y_{1,i}|X_{1,i})+\varepsilon_n\,,\end{aligned} \tag{A143}$$

where $\bar{q} = \mathcal{N}(0, \Lambda - \delta)$, $\mathbf{X}' = \mathbf{f}'(M)$, $\mathbf{X}'' = \mathbf{f}''(M)$, $\mathbf{X}_1 = \mathbf{f}_1(\mathbf{Y}_1)$, and $\varepsilon_n \to 0$ as $n \to \infty$. For the Gaussian relay channel with SFD, we have the following Markov relations,

$$Y_{1,i} - X_i'' - (X_i', X_{1,i}, Y_{1,i}), \tag{A144}$$

$$(X_i'', Y_{1,i}) - (X_i', X_{1,i}) - Y_i. \tag{A145}$$

Hence, by (A145), $I_{\bar{q}}(X_i', X_i'', X_{1,i}; Y_i) = I_{\bar{q}}(X_i', X_{1,i}; Y_i)$. Moving to the second bound in the RHS of (A143), we follow the lines of [5]. Then, by the mutual information chain rule, we have

$$\begin{aligned} I_{\bar{q}}(X_i', X_i''; Y_i, Y_{1,i}|X_{1,i}) &= I(X_i''; Y_{1,i}|X_{1,i}) + I(X_i'; Y_{1,i}|X_i'', X_{1,i}) + I_{\bar{q}}(X_i', X_i''; Y_i|X_{1,i}, Y_{1,i}) \\ &\overset{(a)}{=} I(X_i''; Y_{1,i}|X_{1,i}) + I_{\bar{q}}(X_i', X_i''; Y_i|X_{1,i}, Y_{1,i}) \\ &\overset{(b)}{=} [H(Y_{1,i}|X_{1,i}) - H(Y_{1,i}|X_i'')] + [H_{\bar{q}}(Y_i|X_{1,i}, Y_{1,i}) - H_{\bar{q}}(Y_i|X_i', X_{1,i})] \\ &\overset{(c)}{\leq} I_{q_1}(X_i''; Y_{1,i}) + I(X_i'; Y_i|X_{1,i}) \end{aligned} \tag{A146}$$

where (a) is due to (A144), (b) is due to (A145), and (c) holds since conditioning reduces entropy. Introducing a time-sharing random variable $K \sim \text{Unif}[1:n]$, which is independent of $\mathbf{X}'$, $\mathbf{X}''$, $\mathbf{X}_1$, $\mathbf{Y}$, $\mathbf{Y}_1$, we have that

$$\begin{aligned} R - \varepsilon_n &\leq I_{\bar{q}}(X_K', X_{1,K}; Y_K|K) \\ R - \varepsilon_n &\leq I(X_K''; Y_{1,K}|K) + I_{\bar{q}}(X_K'; Y_K|X_{1,K}, K). \end{aligned} \tag{A147}$$

Now, by the maximum differential entropy lemma (see e.g., [98] (Theorem 8.6.5)),

$$I_{\bar{q}}(X_K', X_{1,K}; Y_K|K) \leq \frac{1}{2}\log\left(\frac{\mathbb{E}[(X_K' + X_{1,K})^2] + (\Lambda - \delta)}{\Lambda - \delta}\right) = \frac{1}{2}\log\left(1 + \frac{\alpha\Omega + \alpha_1\Omega_1 + 2\rho\sqrt{\alpha\Omega \cdot \alpha_1\Omega_1}}{\Lambda - \delta}\right) \tag{A148}$$

and

$$\begin{aligned} I(X_K''; Y_{1,K}|K) + I_{\bar{q}}(X_K'; Y_K|X_{1,K}, K) &\leq \frac{1}{2}\log\frac{\mathbb{E}X_K''^2 + \sigma^2}{\sigma^2} + \frac{1}{2}\log\frac{\left[1 - \frac{(\mathbb{E}(X_K' \cdot X_{1,K}))^2}{\mathbb{E}X_K'^2 \cdot \mathbb{E}X_{1,K}^2}\right]\mathbb{E}X_K'^2 + (\Lambda - \delta)}{\Lambda - \delta} \\ &= \frac{1}{2}\log\left(1 + \frac{(1-\alpha)\Omega}{\sigma^2}\right) + \frac{1}{2}\log\left(1 + \frac{(1-\rho^2)\alpha\Omega}{\Lambda - \delta}\right), \end{aligned} \tag{A149}$$

where α, α_1 and ρ are given by (A128). Since $\delta > 0$ is arbitrary, and $\alpha_1 \leq 1$, the proof follows from (A147)–(A149). □

References

1. Van der Meulen, E.C. Three-terminal communication channels. *Adv. Appl. Probab.* **1971**, *3*, 120–154. [CrossRef]
2. Kim, Y.H. Coding techniques for primitive relay channels. In Proceedings of the Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 26–28 September 2007; pp. 129–135.
3. El Gamal, A.; Kim, Y. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
4. Cover, T.; Gamal, A.E. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory* **1979**, *25*, 572–584. [CrossRef]
5. Gamal, A.E.; Zahedi, S. Capacity of a class of relay channels with orthogonal components. *IEEE Trans. Inf. Theory* **2005**, *51*, 1815–1817.
6. Xue, F. A New Upper Bound on the Capacity of a Primitive Relay Channel Based on Channel Simulation. *IEEE Trans. Inf. Theory* **2014**, *60*, 4786–4798. [CrossRef]

7. Wu, X.; Özgür, A. Cut-set bound is loose for Gaussian relay networks. In Proceedings of the Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 29 September–2 October 2015; pp. 1135–1142.
8. Chen, Y.; Devroye, N. Zero-Error Relaying for Primitive Relay Channels. *IEEE Trans. Inf. Theory* **2017**, *63*, 7708–7715. [CrossRef]
9. Wu, X.; Özgür, A. Cut-set bound is loose for Gaussian relay networks. *IEEE Trans. Inf. Theory* **2018**, *64*, 1023–1037. [CrossRef]
10. Wu, X.; Barnes, L.P.; Özgür, A. "The Capacity of the Relay Channel": Solution to Cover's Problem in the Gaussian Case. *IEEE Trans. Inf. Theory* **2019**, *65*, 255–275. [CrossRef]
11. Mondelli, M.; Hassani, S.H.; Urbanke, R. A New Coding Paradigm for the Primitive Relay Channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2018), Talisa Hotel in Vail, CO, USA, 17–22 June 2018; pp. 351–355.
12. Ramachandran, V. Gaussian degraded relay channel with lossy state reconstruction. *AEU Int. J. Electr. Commun.* **2018**, *93*, 348–353. [CrossRef]
13. Chen, Z.; Fan, P.; Wu, D.; Xiong, K.; Letaief, K.B. On the achievable rates of full-duplex Gaussian relay channel. In Proceedings of the Global Communication Conference (GLOBECOM'2014), Austin, TX, USA, 8–12 December 2014; pp. 4342–4346.
14. Kolte, R.; Özgür, A.; Gamal, A.E. Capacity Approximations for Gaussian Relay Networks. *IEEE Trans. Inf. Theory* **2015**, *61*, 4721–4734. [CrossRef]
15. Jin, X.; Kim, Y. The Approximate Capacity of the MIMO Relay Channel. *IEEE Trans. Inf. Theory* **2017**, *63*, 1167–1176. [CrossRef]
16. Wu, X.; Barnes, L.P.; Özgür, A. The geometry of the relay channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2017), Aachen, Germany, 25–30 June 2017; pp. 2233–2237.
17. Lai, L.; El Gamal, H. The relay–eavesdropper channel: Cooperation for secrecy. *IEEE Trans. Inf. Theory* **2008**, *54*, 4005–4019. [CrossRef]
18. Yeoh, P.L.; Yang, N.; Kim, K.J. Secrecy outage probability of selective relaying wiretap channels with collaborative eavesdropping. In Proceedings of the Global Communication Conference (GLOBECOM'2015), San Diego, CA, USA, 6–10 December 2015; pp. 1–6.
19. Kramer, G.; van Wijngaarden, A.J. On the white Gaussian multiple-access relay channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2000), Sorrento, Italy, 25–30 June 2000; p. 40.
20. Schein, B.E. Distributed Coordination in Network Information Theory. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2001.
21. Rankov, B.; Wittneben, A. Achievable Rate Regions for the Two-way Relay Channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2006), Seattle, WA, USA, 9–14 July 2006; pp. 1668–1672.
22. Gunduz, D.; Yener, A.; Goldsmith, A.; Poor, H.V. The Multiway Relay Channel. *IEEE Trans. Inf. Theory* **2013**, *59*, 51–63. [CrossRef]
23. Maric, I.; Yates, R.D. Forwarding strategies for Gaussian parallel-relay networks. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2004), Chicago, IL, USA, 27 June–2 July 2004; p. 269.
24. Kochman, Y.; Khina, A.; Erez, U.; Zamir, R. Rematch and forward for parallel relay networks. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2008), Toronto, ON, Canada, 6–11 July 2008; pp. 767–771.
25. Awan, Z.H.; Zaidi, A.; Vandendorpe, L. Secure communication over parallel relay channel. *arXiv* **2010**, arXiv:1011.2115. [CrossRef]
26. Xue, F.; Sandhu, S. Cooperation in a Half-Duplex Gaussian Diamond Relay Channel. *IEEE Trans. Inf. Theory* **2007**, *53*, 3806–3814.
27. Kang, W.; Ulukus, S. Capacity of a class of diamond channels. In Proceedings of the Allerton Conference on Communication, Control and Computing, Urbana-Champaign, IL, USA, 23–26 September 2008; pp. 1426–1431.
28. Chern, B.; Özgür, A. Achieving the Capacity of the N-Relay Gaussian Diamond Network Within log$N$ Bits. *IEEE Trans. Inf. Theory* **2014**, *60*, 7708–7718. [CrossRef]

29. Sigurjónsson, S.; Kim, Y.H. On multiple user channels with state information at the transmitters. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2005), Adelaide, Australia, 4–9 September 2005; pp. 72–76.
30. Simeone, O.; Gündüz, D.; Shamai, S. Compound relay channel with informed relay and destination. In Proceedings of the Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 30 September–2 October 2009; pp. 692–699.
31. Zaidi, A.; Vandendorpe, L. Lower bounds on the capacity of the relay channel with states at the source. *EURASIP J. Wirel. Commun. Netw.* **2009**, *2009*, 1–22. [CrossRef]
32. Zaidi, A.; Kotagiri, S.P.; Laneman, J.N.; Vandendorpe, L. Cooperative Relaying With State Available Noncausally at the Relay. *IEEE Trans. Inf. Theory* **2010**, *56*, 2272–2298. [CrossRef]
33. Zaidi, A.; Shamai, S.; Piantanida, P.; Vandendorpe, L. Bounds on the capacity of the relay channel with noncausal state at the source. *IEEE Trans. Inf. Theory* **2013**, *59*, 2639–2672. [CrossRef]
34. Blackwell, D.; Breiman, L.; Thomasian, A.J. The capacities of certain channel classes under random coding. *Ann. Math. Stat.* **1960**, *31*, 558–567. [CrossRef]
35. Simon, M.K.; Alouini, M.S. *Digital Communication over Fading Channels*; John Wiley & Sons: Hoboken, NJ, USA, 2005; Volume 95.
36. Shamai, S.; Steiner, A. A broadcast approach for a single-user slowly fading MIMO channel. *IEEE Trans. Inf. Theory* **2003**, *49*, 2617–2635. [CrossRef]
37. Abdul Salam, A.; Sheriff, R.; Al-Araji, S.; Mezher, K.; Nasir, Q. Novel Approach for Modeling Wireless Fading Channels Using a Finite State Markov Chain. *ETRI J.* **2017**, *39*, 718–728. [CrossRef]
38. Ozarow, L.H.; Shamai, S.; Wyner, A.D. Information theoretic considerations for cellular mobile radio. *IEEE Trans. Veh. Tech.* **1994**, *43*, 359–378. [CrossRef]
39. Goldsmith, A.J.; Varaiya, P.P. Capacity of fading channels with channel side information. *IEEE Trans. Inf. Theory* **1997**, *43*, 1986–1992. [CrossRef]
40. Caire, G.; Shamai, S. On the capacity of some channels with channel state information. *IEEE Trans. Inf. Theory* **1999**, *45*, 2007–2019. [CrossRef]
41. Zhou, S.; Zhao, M.; Xu, X.; Wang, J.; Yao, Y. Distributed wireless communication system: A new architecture for future public wireless access. *IEEE Commun. Mag.* **2003**, *41*, 108–113. [CrossRef]
42. Xu, Y.; Lu, R.; Shi, P.; Li, H.; Xie, S. Finite-time distributed state estimation over sensor networks with round-robin protocol and fading channels. *IEEE Trans. Cybern.* **2018**, *48*, 336–345. [CrossRef]
43. Kuznetsov, A.V.; Tsybakov, B.S. Coding in a memory with defective cells. *Probl. Peredachi Inf.* **1974**, *10*, 52–60.
44. Heegard, C.; Gamal, A.E. On the capacity of computer memory with defects. *IEEE Trans. Inf. Theory* **1983**, *29*, 731–739. [CrossRef]
45. Kuznetsov, A.V.; Vinck, A.J.H. On the general defective channel with informed encoder and capacities of some constrained memories. *IEEE Trans. Inf. Theory* **1994**, *40*, 1866–1871. [CrossRef]
46. Kim, Y.; Kumar, B.V.K.V. Writing on dirty flash memory. In Proceedings of the Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 30 September–3 October 2014; pp. 513–520.
47. Bunin, A.; Goldfeld, Z.; Permuter, H.H.; Shamai, S.; Cuff, P.; Piantanida, P. Key and message semantic-security over state-dependent channels. *IEEE Trans. Inf. Forensic Secur.* **2018**. [CrossRef]
48. Gungor, O.; Koksal, C.E.; Gamal, H.E. An information theoretic approach to RF fingerprinting. In Proceedings of the Asilomar Conference on Signals, Systems and Computers (ACSSC'2013), Pacific Grove, CA, USA, 3–6 November 2013; pp. 61–65.
49. Ignatenko, T.; Willems, F.M.J. Biometric security from an information-theoretical perspective. *Found. Trends® Commun. Inf. Theory* **2012**, *7*, 135–316. [CrossRef]
50. Han, G.; Xiao, L.; Poor, H.V. Two-dimensional anti-jamming communication based on deep reinforcement learning. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017.
51. Xu, W.; Trappe, W.; Zhang, Y.; Wood, T. The feasibility of launching and detecting jamming attacks in wireless networks. In Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing, Urbana-Champaign, IL, USA, 25–27 May 2005; pp. 46–57.
52. Alnifie, G.; Simon, R. A multi-channel defense against jamming attacks in wireless sensor networks. In Proceedings of the ACM Workshop on QoS Security Wireless Mobile Networks, Crete Island, Greece, 22 October 2007; pp. 95–104.

53. Padmavathi, G.; Shanmugapriya, D. A survey of attacks, security mechanisms and challenges in wireless sensor networks. *arXiv* **2009**, arXiv:0909.0576.
54. Wang, T.; Liang, T.; Wei, X.; Fan, J. Localization of Directional Jammer in Wireless Sensor Networks. In Proceedings of the 2018 International Conference on Robots & Intelligent System (ICRIS) (ICRIS'2018), Changsha, China, 26–27 May 2018; pp. 198–202.
55. Ahlswede, R. Elimination of correlation in random codes for arbitrarily varying channels. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **1978**, *44*, 159–175. [CrossRef]
56. Ericson, T. Exponential error bounds for random codes in the arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1985**, *31*, 42–48. [CrossRef]
57. Csiszár, I.; Narayan, P. The capacity of the arbitrarily varying channel revisited: Positivity, constraints. *IEEE Trans. Inf. Theory* **1988**, *34*, 181–193. [CrossRef]
58. Ahlswede, R. Coloring hypergraphs: A new approach to multi-user source coding, Part 2. *J. Comb.* **1980**, *5*, 220–268.
59. Ahlswede, R. Arbitrarily varying channels with states sequence known to the sender. *IEEE Trans. Inf. Theory* **1986**, *32*, 621–629. [CrossRef]
60. Jahn, J.H. Coding of arbitrarily varying multiuser channels. *IEEE Trans. Inf. Theory* **1981**, *27*, 212–226. [CrossRef]
61. Hof, E.; Bross, S.I. On the deterministic-code capacity of the two-user discrete memoryless Arbitrarily Varying General Broadcast channel with degraded message sets. *IEEE Trans. Inf. Theory* **2006**, *52*, 5023–5044. [CrossRef]
62. Winshtok, A.; Steinberg, Y. The arbitrarily varying degraded broadcast channel with states known at the encoder. In Proceedings of the 2006 IEEE International Symposium on Information Theory, Seattle, WA, USA, 9–14 July 2006; pp. 2156–2160.
63. He, X.; Khisti, A.; Yener, A. MIMO Multiple Access Channel With an Arbitrarily Varying Eavesdropper: Secrecy Degrees of Freedom. *IEEE Trans. Inf. Theory* **2013**, *59*, 4733–4745.
64. Pereg, U.; Steinberg, Y. The arbitrarily varying degraded broadcast channel with causal side information at the encoder. In Proceedings of the 2018 International Zurich Seminar Information Communication (IZS'2018), Aachen, Germany, 25–30 June 2018; pp. 20–24.
65. Keresztfalvi, T.; Lapidoth, A. Partially-robust communications over a noisy channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2018), Vail, CO, USA, 17–22 June 2018; pp. 2003–2006.
66. Gubner, J.A. Deterministic Codes for Arbitrarily Varying Multiple-Access Channels. Ph.D. Dissertation, University of Maryland, College Park, MD, USA, 1988.
67. Gubner, J.A. On the deterministic-code capacity of the multiple-access arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1990**, *36*, 262–275. [CrossRef]
68. Gubner, J.A. State constraints for the multiple-access arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1991**, *37*, 27–35. [CrossRef]
69. Gubner, J.A. On the capacity region of the discrete additive multiple-access arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1992**, *38*, 1344–1347. [CrossRef]
70. Gubner, J.A.; Hughes, B.L. Nonconvexity of the capacity region of the multiple-access arbitrarily varying channel subject to constraints. *IEEE Trans. Inf. Theory* **1995**, *41*, 3–13. [CrossRef]
71. Ahlswede, R.; Cai, N. *Arbitrarily Varying Multiple-Access Channels*; Universität Bielefeld: Bielefeld, Germany, 1996.
72. Ahlswede, R.; Cai, N. Arbitrarily varying multiple-access channels. I. Ericson's symmetrizability is adequate, Gubner's conjecture is true. *IEEE Trans. Inf. Theory* **1999**, *45*, 742–749. [CrossRef]
73. He, X.; Khisti, A.; Yener, A. MIMO multiple access channel with an arbitrarily varying eavesdropper. In Proceedings of the Allerton Conference on Communication, Control and Computing (Allerton'2011), Monticello, IL, USA, 28–30 September 2011; pp. 1182–1189.
74. Wiese, M.; Boche, H. The arbitrarily varying multiple-access channel with conferencing encoders. *IEEE Trans. Inf. Theory* **2013**, *59*, 1405–1416. [CrossRef]
75. Nitinawarat, S. On the Deterministic Code Capacity Region of an Arbitrarily Varying Multiple-Access Channel Under List Decoding. *IEEE Trans. Inf. Theory* **2013**, *59*, 2683–2693. [CrossRef]

76. MolavianJazi, E.; Bloch, M.; Laneman, J.N. Arbitrary jamming can preclude secure communication. In Proceedings of the Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 30 September–2 October 2009; pp. 1069–1075.
77. Boche, H.; Schaefer, R.F. Capacity results and super-activation for wiretap channels with active wiretappers. *IEEE Trans. Inf. Theory* **2013**, *8*, 1482–1496. [CrossRef]
78. Aydinian, H.; Cicalese, F.; Deppe, C. *Information Theory, Combinatorics, and Search Theory*; Springer: Berlin/Heidelberg, Germany, 2013; Chapter 5.
79. Boche, H.; Schaefer, R.F.; Poor, H.V. On arbitrarily varying wiretap channels for different classes of secrecy measures. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2014), Honolulu, HI, USA, 29 June–4 July 2014; pp. 2376–2380.
80. Boche, H.; Schaefer, R.F.; Poor, H.V. On the continuity of the secrecy capacity of compound and arbitrarily varying wiretap channels. *IEEE Trans. Inf. Forensic Secur.* **2015**, *10*, 2531–2546. [CrossRef]
81. Nötzel, J.; Wiese, M.; Boche, H. The arbitrarily varying wiretap channel—Secret randomness, stability, and super-activation. *IEEE Trans. Inf. Theory* **2016**, *62*, 3504–3531. [CrossRef]
82. Goldfeld, Z.; Cuff, P.; Permuter, H.H. Arbitrarily Varying Wiretap Channels With Type Constrained States. *IEEE Trans. Inf. Theory* **2016**, *62*, 7216–7244. [CrossRef]
83. He, D.; Luo, Y. Arbitrarily varying wiretap channel with state sequence known or unknown at the receiver. *arXiv* **2017**, arXiv:1701.02043.
84. Boche, H.; Deppe, C. Secure identification for wiretap channels; Robustness, super-additivity and continuity. *IEEE Trans. Inf. Forensic Secur.* **2018**, *13*, 1641–1655. [CrossRef]
85. Pereg, U.; Steinberg, Y. The Arbitrarily Varying Channel Under Constraints With Side Information at the Encoder. *IEEE Trans. Inf. Theory* **2019**, *65*, 861–887. [CrossRef]
86. Pereg, U.; Steinberg, Y. The arbitrarily varying channel under constraints with causal side information at the encoder. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2017), Aachen, Germany, 25–30 June 2017; pp. 2805–2809.
87. Csiszár, I.; Narayan, P. Capacity of the Gaussian arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1991**, *37*, 18–26. [CrossRef]
88. Behboodi, A.; Piantanida, P. On the simultaneous relay channel with informed receivers. In Proceedings of the IEEE International Symposium on Information Theory (ISIT'2009), Seoul, Korea, 28 June–3 July 2009; pp. 1179–1183.
89. Bjelaković, I.; Boche, H.; Sommerfeld, J. Capacity results for arbitrarily varying wiretap channels. In *Information Theory, Combinatorics, and Search Theory*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 123–144.
90. Sion, M. On General Minimax Theorems. *Pac. J. Math.* **1958**, *8*, 171–176. [CrossRef]
91. Pereg, U.; Steinberg, Y. The arbitrarily varying gaussian relay channel with sender frequency division. *arXiv* **2018**, arXiv:1805.12595.
92. Hughes, B.; Narayan, P. Gaussian arbitrarily varying channels. *IEEE Trans. Inf. Theory* **1987**, *33*, 267–284. [CrossRef]
93. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.
94. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
95. Blackwell, D.; Breiman, L.; Thomasian, A.J. The capacity of a class of channels. *Ann. Math. Stat.* **1959**, *30*, 1229–1241. [CrossRef]
96. Diggavi, S.N.; Cover, T.M. The worst additive noise under a covariance constraint. *IEEE Trans. Inf. Theory* **2001**, *47*, 3072–3081. [CrossRef]
97. Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656. [CrossRef]
98. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.