Article

Impact of Interference in Coexisting Wireless Networks with Applications to Arbitrarily Varying Bidirectional Broadcast Channels

Rafael F. Wyrembelski * and Holger Boche

Lehrstuhl für Theoretische Informationstechnik, Technische Universität München, Arcisstraße 21, 80333 Munich, Germany; E-Mail: boche@tum.de

* Author to whom correspondence should be addressed; E-Mail: wyrembelski@tum.de; Tel.: +49-89-289-23246; Fax: +49-89-289-23242.

Abstract: The paradigm shift from an exclusive allocation of frequency bands, one for each system, to a shared use of frequencies comes along with the need of new concepts since interference will be an ubiquitous phenomenon. In this paper, we use the concept of *arbitrarily varying channels* to model the impact of unknown interference caused by coexisting wireless systems which operate on the same frequencies. Within this framework, capacity can be zero if pre-specified encoders and decoders are used. This necessitates the use of more sophisticated coordination schemes where the choice of encoders and decoders is additionally coordinated based on common randomness. As an application we study the *arbitrarily varying bidirectional broadcast channel* and derive the capacity regions for different coordination strategies. This problem is motivated by decode-and-forward bidirectional or two-way relaying, where a relay establishes a bidirectional communication between two other nodes while sharing the resources with other coexisting wireless networks.

Notation

In this paper we denote discrete random variables by non-italic capital letters and their corresponding realizations and ranges by lower case italic letters and script letters, e.g., X, x, and $\mathcal{X}$, respectively; the notation X^n stands for the sequence $X_1, X_2, ..., X_n$ of length n; $\mathbb{N}$ and $\mathbb{R}_+$ denote the set of positive integers and non-negative real numbers; all logarithms, exponentials, and information quantities are taken to the basis 2; $I(\cdot;\cdot)$, $H(\cdot)$, and $D(\cdot\|\cdot)$ are the mutual information, entropy, and Kullback–Leibler (information) divergence; $\mathbb{E}[\cdot]$ and $\mathbb{P}\{\cdot\}$ denote the expectation and probability; $\langle\cdot,\cdot\rangle$ is the inner product and $|\cdot|^+ = \max\{\cdot, 0\}$; $\mathcal{P}(\cdot)$ is the set of all probability distributions and $(\cdot)^c$ is the complement of a set; $W^{\otimes n}$ is the n-th memoryless extension of the stochastic matrix W; lhs := rhs means the value of the right hand side (rhs) is assigned to the left hand side (lhs); lhs =: rhs is defined accordingly.

1. Introduction

The ongoing research progress reveals a paradigm shift from an exclusive allocation of certain frequency bands to a shared use of frequencies. While most current systems such as conventional cellular systems usually operate on exclusive frequency bands, several future systems such as ad-hoc or sensor networks will operate on shared resources in an uncoordinated and self-organizing way. The main issue that comes along with this development is that interference becomes an ubiquitous phenomenon and that it will be one of the major impairments in future wireless networks. Since the induced interference can no longer be coordinated between the coexisting networks, new concepts are needed especially for the frequency usage.

As an example, Figure 1 depicts a wireless network that consists of several uncoordinated transmitter-receiver pairs or links, where each receiver receives the signal he is interested in but is also confronted with interfering signals from other transmitting nodes. If there is no a priori knowledge about applied transmit strategies of all other transmitting nodes such as coding or modulation schemes, there is no knowledge about the induced interference. Thus, users are confronted with channels that may vary from symbol to symbol in an unknown and arbitrary manner. The concept of *arbitrarily varying channels* (AVC) [1–4] provides a suitable and robust model for such communication scenarios.

Figure 1. Wireless network with several transmitter-receiver pairs. Each receiver receives a desired signal (solid) and simultaneously receives interference from all other transmitters (dashed).
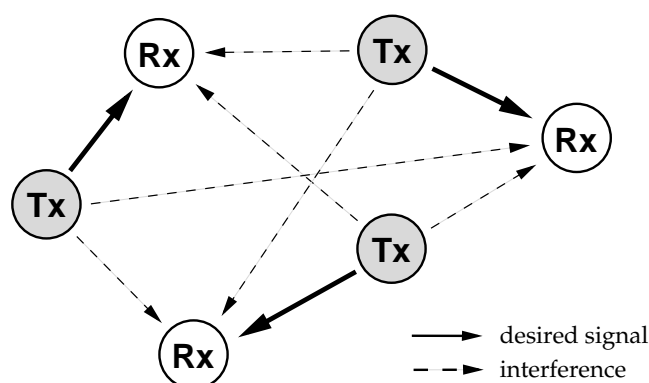
Interestingly, it is shown for the single-user AVC that its capacity highly depends on how encoder and decoder are coordinated within one transmitter-receiver link: the *deterministic code capacity*, *i.e.*, the traditional approach with pre-specified encoder and decoder, either equals the *random code capacity*, *i.e.*, additional encoder-decoder coordination based on common randomness, or is otherwise zero [2]. It is shown that symmetrizable AVCs prevent reliable communication for the traditional approach without additional coordination. Roughly speaking, in this case a symmetrizable AVC can *emulate* a valid input, which makes it impossible for the decoder to decide on the correct codeword. Unfortunately, many channels of practical importance fall in the category of symmetrizable channels [4].

The situation changes significantly, if constraints on the permissible codewords and channel states are imposed. Such restrictions are motivated by the fact that in real communication systems the transmitter as well as possible interferers are usually limited in their transmit powers. For the single-user AVC under input and state constraints, it is shown that due to the imposed constraints the deterministic code capacity may be positive even for symmetrizable channels, but may be less than its random code capacity [4,5].

Besides the single-user AVC there are several important extensions to multi-user settings as well. The arbitrarily varying wiretap channel is analyzed in [6,7]. The arbitrarily varying multiple access channel (AVMAC) is analyzed in [8–10], where its deterministic code and random code capacity regions are established. The AVMAC with constraints on input and states is considered in [11,12], where in the latter it is shown that the random code capacity region is non-convex in general. The AVMAC with conferencing encoders is analyzed in detail in [13,14]. While the AVMAC is well understood, there are only partial results known so far for the arbitrarily varying general broadcast channel. Achievable deterministic code rate regions are analyzed in [8,15], where the latter further imposes the assumption of degraded message sets. But unfortunately, no converses or outer bounds on the capacity region are given.

In this paper we analyze *bidirectional relaying*, or two-way relaying, for arbitrarily varying channels. The concept of bidirectional relaying has the potential to significantly improve the overall performance and coverage in wireless networks such as ad-hoc, sensor, and even cellular systems. This is mainly based on the fact that it advantageously exploits the bidirectional information flow of the communication to reduce the inherent loss in spectral efficiency induced by half-duplex relays [16–19].

Bidirectional relaying applies to three-node networks, where a half-duplex relay node establishes a bidirectional communication between two other nodes using a two-phase decode-and-forward protocol. There, in the initial multiple access (MAC) phase two nodes transmit their messages to the relay node which decodes them. In the succeeding broadcast phase the relay re-encodes and transmits both messages in such a way that both receiving nodes can decode their intended message using their own message from the previous phase as side information. Note that due to the complementary side information at the receiving nodes this scenario differs from the classical broadcast channel and is therefore known as bidirectional broadcast channel (BBC). It is shown in [20–23] for discrete memoryless channels and in [24] for MIMO Gaussian channels that capacity is achieved by a single data stream that combines both messages based on the network coding idea. Optimal transmit strategies for the multi-antenna BBC are then analyzed in [25,26]. Bidirectional relaying for compound channels is studied in [27,28], while [29] discusses adaptive bidirectional relaying with quantized channel state information. Besides the decode-and-forward protocol [20–32] there are also amplify-and-forward [32–36] or compress-and-forward [37–39] approaches similarly as for the classical relay channel. A

newer approach is compute-and-forward [40–46], where the relay decodes a certain function of both individual messages. Another approach is given in [47] which is based on the noisy network coding idea [48–50].

Here, we use the concept of arbitrarily varying channels to study bidirectional relaying that operates on the same (frequency) resources as other coexisting wireless networks. Then the initial MAC phase is specified by the AVMAC and is therefore well understood [8–12]. Thus, it remains to study the BBC phase for arbitrarily varying channels. The *arbitrarily varying bidirectional broadcast channel (AVBBC)* is analyzed in [51–53], where it is shown that the AVBBC displays a dichotomy behavior similar to the single-user AVC: its deterministic code capacity region either equals its random code capacity region or else has an empty interior. Having practical limitations on transmit powers in mind, in this paper we impose constraints on the permissible codewords and state sequences and derive the corresponding deterministic code and random code capacity regions of the AVBBC under input and state constraints.

The rest of this paper is organized as follows. In Section 2 we briefly review the concept of types from Csiszár and Körner and state some information theoretic and combinatoric preliminaries. In Section 3 we introduce the concept of arbitrarily varying channels as a suitable model for communication in wireless networks, which share the resources with other coexisting systems in an uncoordinated way, and review the impact of coordination within one transmitter-receiver link on the capacity. As an application for this framework we then study bidirectional relaying under such conditions. We impose constraints on the permissible input and state sequences and analyze bidirectional relaying for arbitrarily varying channels in Section 4. This requires the study of the AVBBC under input and state constraints for which we derive its deterministic code and random code capacity regions. Finally, we conclude the paper in Section 5.

2. Preliminaries

We denote the mutual information between the input random variable X and the output random variable Y by $I(X; Y)$. To emphasize the dependency of the mutual information on the input distribution $p \in \mathcal{P}(\mathcal{X})$ and the channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, we also write $I(X; Y) = I(p, W)$ interchangeably.

Furthermore, we extensively use the concept of *types* from Csiszár and Körner [3], which is briefly reviewed in the following. The *type* of a sequence $x^n = (x_1, ..., x_n) \in \mathcal{X}^n$ of length n is a distribution $P_{x^n} \in \mathcal{P}(\mathcal{X})$ defined by $P_{x^n}(a) := \frac{1}{n} N(a|x^n)$ for every $a \in \mathcal{X}$. Thereby, $N(a|x^n)$ denotes the number of indices i such that $x_i = a$, $i = 1, ..., n$. The set of all types of sequences in $\mathcal{X}^n$ is denoted by $\mathcal{P}_0(n, \mathcal{X})$. The notation extends to *joint types* in a natural way. For example the joint type of sequences $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$ is the distribution $P_{x^n,y^n} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ where $P_{x^n,y^n}(a, b) = \frac{1}{n} N(a, b|x^n, y^n)$ for every $a \in \mathcal{X}, b \in \mathcal{Y}$, where $N(a, b|x^n, y^n)$ is the number of indices i such that $(x_i, y_i) = (a, b)$, $i = 1, ..., n$.

For notational convenience, we represent (joint) types of sequences of length n by (joint) distributions of dummy variables. For instance, the random variables X and Y represent a joint type, e.g., $P_{XY} = P_{x^n,y^n}$ for some $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$. The set of all sequences of type P_{x^n} is denoted by $\mathcal{T}_X^n = \{x^n : x^n \in \mathcal{X}^n, P_{x^n} = P_X\}$. Of course, this notation extends to joint types and sections in a self-explanatory way, e.g., $\mathcal{T}_{XY}^n = \{(x^n, y^n) : x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n, P_{x^n,y^n} = P_{XY}\}$ or $\mathcal{T}_{Y|X}^n(x^n) = \{y^n : (x^n, y^n) \in \mathcal{T}_{XY}^n\}$.

Remark 1. *To avoid notational ambiguity we usually use small letters to denote arbitrary probability distributions, e.g., $p \in \mathcal{P}(\mathcal{X})$, and capital letters to highlight types, e.g., $P_X \in \mathcal{P}_0(n, \mathcal{X})$.*

Next, we state as facts some bounds on types which we will need for our proofs, *cf.* for example Csiszár and Körner ([3], Section 1.2).

Fact 1: The number of possible types of sequences of length n is a polynomial in n, *i.e.*,

$$|\mathcal{P}_0(n, \mathcal{X})| \le (n+1)^{|\mathcal{X}|}$$

Fact 2: We have

$$(n+1)^{-|\mathcal{X}|} \exp\big(nH(X)\big) \le |\mathcal{T}_X^n| \le \exp\big(nH(X)\big), \quad \text{if } \mathcal{T}_X^n \ne \emptyset$$
$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\big(nH(Y|X)\big) \le |\mathcal{T}_{Y|X}^n(x^n)| \le \exp\big(nH(Y|X)\big), \quad \text{if } \mathcal{T}_{Y|X}^n(x^n) \ne \emptyset$$

*Fact 3:* For any channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$,

$$\sum_{x^n \in \mathcal{T}_{Y|X}^n(x^n)} W^{\otimes n}(y^n|x^n) = W^{\otimes n}(\mathcal{T}_{Y|X}^n(x^n)|x^n) \le \exp\big(-nD(P_{XY}\|P_X \otimes W)\big)$$

where $W^{\otimes n}(y^n|x^n) := \prod_{k=1}^n W(y_k|x_k)$ and $P_X \otimes W$ denotes the distribution on $\mathcal{X} \times \mathcal{Y}$ with probability mass function $P_X(x)W(y|x)$.

3. Modeling of Communication in Coexisting Wireless Networks

Here we introduce the concept of arbitrarily varying channels as a suitable model for communication in coexisting wireless networks. To highlight the crucial points we consider the simplest interference scenario with two transmitter-receiver pairs (or links) as shown in Figure 2. Here, each receiver receives signals from both transmitters, but is only interested in the information from its own transmitter.

Figure 2. Interference channel with two transmitters and receivers. Each receiver receives the desired signal (solid) from the intended transmitter but simultaneously receives also interference (dashed) the other transmitter.
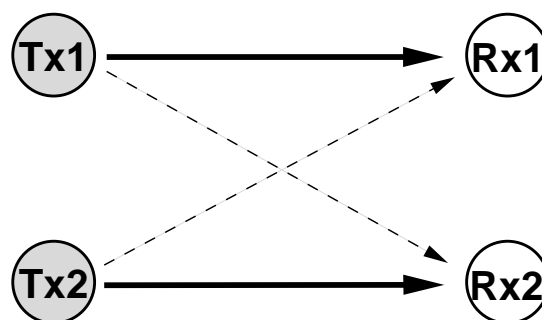
Since in practical systems a transmitter usually uses a finite modulation scheme and a receiver quantizes the received signal before further processing, it is reasonable to assume finite input and output alphabets denoted by $\mathcal{X}_i$ and $\mathcal{Y}_i$ for link i, $i = 1, 2$, respectively. Then, for input and output sequences

$x_i^n \in \mathcal{X}_i^n$ and $y_i^n \in \mathcal{Y}_i^n$ of length n, the transmission over the discrete memoryless channel is completely characterized by a stochastic matrix

$$W_i^{\otimes n}(y_i^n | x_1^n, x_2^n) := \prod_{k=1}^n W_i(y_{i,k} | x_{1,k}, x_{2,k}), \quad i = 1, 2 \tag{1}$$

Thereby, the additive noise at the receivers is taken into account by considering stochastic matrices and not deterministic ones. Interestingly, the transmission model in Equation (1) looks like a multiple access channel, since the received signal depends on both the codeword of the intended message and the codeword of the interfering message from the other link.

Remark 2. *If we treat the received signal from the other transmitter as additional noise, we would end up with a modified stochastic matrix $\widetilde{W}_i^{\otimes n}(y_i^n | x_i^n)$, $i = 1, 2$, where the received signal depends only on the codeword of the intended message.*

We consider the standard model with block codes of arbitrary but fixed length n. Let $\mathcal{M}_i := \{1, ..., M_{i,n}\}$, $i = 1, 2$, be the set of messages to transmit. The traditional coding strategy for each transmitter-receiver pair is specified by the following definition of deterministic codes.

Definition 1. *A* deterministic $(n, M_{i,n})$-code *or* codebook $\mathcal{C}_{i,det}$ *for transmitter-receiver pair i is a family*

$$\mathcal{C}_{i,det} := \left\{ (x_{m_i}^n, \mathcal{D}_{m_i}^{(i)}) : m_i \in \mathcal{M}_i \right\}$$

with codewords $x_{m_i}^n \in \mathcal{X}_i^n$, one for each message $m_i \in \mathcal{M}_i$, and decoding sets $\mathcal{D}_{m_i}^{(i)} \subseteq \mathcal{Y}_i^n$ for all $m_i \in \mathcal{M}_i$ with $\mathcal{D}_{m_k}^{(i)} \cap \mathcal{D}_{m_l}^{(i)} = \emptyset$ for $m_k \neq m_l$, $i = 1, 2$.

When $x_{m_1}^n$ and $x_{m_2}^n$ have been sent according to fixed codebooks $\mathcal{C}_{1,det}$ and $\mathcal{C}_{2,det}$, and y_1^n and y_2^n have been received, the decoder of receiver i is in error if $y_i^n \notin \mathcal{D}_{m_i}^{(i)}$, $i = 1, 2$. With this, we can define the probability of error at receiver 1 for given messages m_1 and m_2 as

$$e_1(m_1, x_{m_2}^n) := W_1^{\otimes n}\left((\mathcal{D}_{m_1}^{(1)})^c | x_{m_1}^n, x_{m_2}^n \right)$$

and the *average probability of error* at receiver 1 as

$$\bar{e}_1(x_{m_2}^n) := \frac{1}{M_{1,n}} \sum_{m_1=1}^{M_{1,n}} e_1(m_1, x_{m_2}^n)$$

with similar expressions $e_2(m_2, x_{m_1}^n)$ and $\bar{e}_2(x_{m_1}^n)$ for receiver 2.

Important to note is that the probability of error depends on the codebooks that both transmitter-receiver pairs use as well as on the specific message the interfering transmitter sends.

Definition 2. *A rate $R_i \in \mathbb{R}_+$ is said to be* deterministically achievable *if for any $\delta > 0$ there exists an $n(\delta) \in \mathbb{N}$ and a sequence of deterministic $(n, M_{i,n})$-codes $\{\mathcal{C}_{i,det}^{(n)}\}_{n \in \mathbb{N}}$, $i = 1, 2$, such that for all $n \geq n(\delta)$ we have*

$$\frac{1}{n} \log M_{i,n} \geq R_i - \delta$$

while $\max_{x_{m_2}^n} \bar{e}_1(x_{m_2}^n), \max_{x_{m_1}^n} \bar{e}_2(x_{m_1}^n) \to 0$ as $n \to \infty$. The deterministic code capacity is the largest deterministically achievable rate.

If we assume no coordination between both transmitter-receiver pairs, there is no a priori knowledge about the used codebooks and codewords that are chosen by the interfering transmitter. Consequently, the receiver can be confronted with arbitrary interfering sequences. This corresponds to the concept of *arbitrarily varying channels* (AVC) [1–4] and the only way to guarantee a successful transmission is to find a universal strategy that works for all possible codebooks and interfering codewords simultaneously.

To model the appearance of arbitrary interfering sequences, we introduce a finite state set $\mathcal{S}$. Then, for a fixed state sequence $s^n \in \mathcal{S}^n$ of length n and input and output sequences $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, the discrete memoryless channel is given by $W^{\otimes n}(y^n|x^n, s^n) := \prod_{k=1}^{n} W(y_k|x_k, s_k)$. (In the following we drop the index $(\cdot)_i$ indicating the transmitter-receiver pair, since obviously the argumentation holds for all i.)

Note that the input sequence and interfering sequence originate from different and, in particular, uncoordinated transmitters, so that they are independent of each other. But of course, the codebook has to be designed in such a way that each codeword works for all possible interfering sequences simultaneously.

Definition 3. *The discrete memoryless* arbitrarily varying channel *(AVC)* $\mathfrak{W}^n$ *is the family*

$$\mathfrak{W}^n := \left\{ W^{\otimes n} : \mathcal{X}^n \times \mathcal{S}^n \to \mathcal{P}(\mathcal{Y}^n) \right\}_{n \in \mathbb{N}, s^n \in \mathcal{S}^n}$$

Further, for any probability distribution $q \in \mathcal{P}(\mathcal{S})$ we denote the averaged channel by $\overline{W}_q(y|x) := \sum_{s \in \mathcal{S}} W(y|x, s)q(s)$.

3.1. Impact of Coordination within Transmitter-Receiver Pair

In the following we analyze and review different approaches of coordination in one transmitter-receiver pair and specify their impact on the transmission. Therefore, we characterize all achievable rates at which reliable communication is possible for three different types of coordination: the traditional approach as well as additional encoder-decoder coordination based on common randomness or based on correlated side information.

3.1.1. No Additional Coordination

The system design of the traditional or conventional approach without additional coordination is defined by a deterministic coding strategy, where transmitter and receiver use a pre-specified encoder and decoder as given in Definition 1. We further need the concept of symmetrizability to state the main result for this approach.

Definition 4. *An AVC* $\mathfrak{W}^n$ *is* symmetrizable *if for some channel* $U : \mathcal{X} \to \mathcal{P}(\mathcal{S})$

$$\sum_{s \in \mathcal{S}} W(y|x, s)U(s|x') = \sum_{s \in \mathcal{S}} W(y|x', s)U(s|x) \tag{2}$$

holds for every $x, x' \in \mathcal{X}$ *and* $y \in \mathcal{Y}$. *This means the channel* $\widetilde{W}(y|x, x') = \sum_{s \in \mathcal{S}} W(y|x, s)U(s|x')$ *is symmetric in* x, x' *for all* $x, x' \in \mathcal{X}$ *and* $y \in \mathcal{Y}$.

For the traditional approach the capacity is known [2–4] and summarized in the following theorem.

Theorem 1. *The deterministic code capacity $C_{det}(\mathfrak{W}^n)$ of the AVC $\mathfrak{W}^n$ is*

$$C_{det}(\mathfrak{W}^n) = \begin{cases} \max_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{S})} I(p, \overline{W}_q) & \text{if } \mathfrak{W}^n \text{ is non-symmetrizable} \\ 0 & \text{if } \mathfrak{W}^n \text{ is symmetrizable} \end{cases}$$

The complete proof can be found for example in [4]. In the following we only want to highlight the key insight why we have a zero capacity if the AVC $\mathfrak{W}^n$ is symmetrizable.

Let $x_m^n \in \mathcal{X}^n$, $m = 1, ..., M_n$ with $M_n \geq 2$ be arbitrary codewords. For a symmetrizable AVC $\mathfrak{W}^n$, we can consider interfering sequences that look like valid codewords, more precisely we set $s_m^n = x_m^n$, $m = 1, ..., M_n$. Now, for each pair of codewords $(k, l) \in \mathcal{M} \times \mathcal{M}$ with $k \neq l$ we have for the probability of error

$$\begin{aligned} \mathbb{E}\big[e(k, \mathrm{S}_l^n)\big] + \mathbb{E}\big[e(l, \mathrm{S}_k^n)\big] &= \mathbb{E}\big[W^{\otimes n}((\mathcal{D}_k)^c | x_k^n, \mathrm{S}_l^n)\big] + \mathbb{E}\big[W^{\otimes n}((\mathcal{D}_l)^c | x_l^n, \mathrm{S}_k^n)\big] \\ &= \mathbb{E}\big[W^{\otimes n}((\mathcal{D}_k)^c | x_k^n, \mathrm{S}_l^n)\big] + \mathbb{E}\big[W^{\otimes n}((\mathcal{D}_l)^c | x_k^n, \mathrm{S}_l^n)\big] \\ &\geq \mathbb{E}\big[W^{\otimes n}((\mathcal{D}_k)^c | x_k^n, \mathrm{S}_l^n)\big] + \mathbb{E}\big[W^{\otimes n}(\mathcal{D}_k | x_k^n, \mathrm{S}_l^n)\big] \\ &= \mathbb{E}\big[W^{\otimes n}((\mathcal{D}_k)^c \cup \mathcal{D}_k | x_k^n, \mathrm{S}_l^n)\big] = 1 \end{aligned}$$

where the second equality follows from the fact that the AVC $\mathfrak{W}^n$ is symmetrizable, *cf.* Definition 4. Hence, this leads for the average probability of error to

$$\begin{aligned} \frac{1}{M_n} \sum_{l=1}^{M_n} \mathbb{E}\big[\bar{e}(\mathrm{S}_l^n)\big] &= \frac{1}{(M_n)^2} \sum_{k=1}^{M_n} \sum_{l=1}^{M_n} \mathbb{E}\big[e(k, \mathrm{S}_l^n)\big] \\ &\geq \frac{1}{(M_n)^2} \cdot \frac{M_n(M_n - 1)}{2} \\ &= \frac{M_n - 1}{2 M_n} \geq \frac{1}{4} \end{aligned}$$

which implies that $\mathbb{E}[\bar{e}(\mathrm{S}_l^n)] \geq \frac{1}{4}$ for at least one $l \in \mathcal{M}$. Since the average probability of error is bounded from below by a positive constant, a reliable transmission is not possible, so that we have $C_{\det}(\mathfrak{W}^n) = 0$ if the AVC is symmetrizable.

This becomes intuitively clear, if one realizes the following. Since the AVC $\mathfrak{W}^n$ is symmetrizable, *cf.* (2), it can happen that the interfering sequence looks like another valid codeword. Then, the receiver receives a superimposed version of two valid codewords and cannot distinguish which one comes from the intended transmitter and which one is the interfering sequence. Thus, reliable communication can no longer be guaranteed.

3.1.2. Encoder-Decoder Coordination Based on Common Randomness

Since the traditional interference coordination with predetermined encoder and decoder fails in the case of symmetrizable channels, we are interested in strategies that work well also in this case. Therefore, we consider in the following a strategy with a more involved coordination, where we additionally allow transmitter and receiver to coordinate their choice of encoder and decoder based on an access to a common resource independent of the current message. This leads directly to the following definition.

Definition 5. *A random (n, M_n, Z)-code $\mathcal{C}_{ran}(\mathfrak{W}^n)$ for the AVC $\mathfrak{W}^n$ is given by a family $\mathcal{C}_{ran}(\mathfrak{W}^n) \coloneqq \{\mathcal{C}(z) : z \in \mathcal{Z}\}$ of deterministic (n, M_n)-codes*

$$\mathcal{C}(z) \coloneqq \left\{ \left(x_m^n(z), \mathcal{D}_m(z) \right) : m \in \mathcal{M} \right\}$$

together with a random variable $Z \in \mathcal{Z}$ distributed according to $p_Z \in \mathcal{P}(\mathcal{Z})$.

This means that codewords and decoding sets are chosen according to a common random experiment, realized in Definition 5 by the random variable Z, whose outcome has to be known to the transmitter and receiver in advance. The definitions of probability of error, a *randomly achievable* rate, and the *random code capacity* $C_{\mathrm{ran}}(\mathfrak{W}^n)$ follow accordingly as in Section 3.1.1.

The access to the common resource can be realized for example by an external source such as a satellite signal. Moreover, we know from [2] that if we transmit at rate R with exponentially many messages, *i.e.*, $\exp(nR)$, it suffices to use a random code which consists of n^2 encoder-decoder pairs and a uniformly distributed random variable whose value indicates which encoder and decoder the transmitter and receiver have to use.

Due to the additional coordination within one transmitter-receiver pair, we expect an improvement in the performance compared to the traditional approach especially for symmetrizable channels. The following result confirms our intuition [1,3].

Theorem 2. *The random code capacity $C_{ran}(\mathfrak{W}^n)$ of the AVC $\mathfrak{W}^n$ is*

$$C_{ran}(\mathfrak{W}^n) = \max_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{S})} I(p, \overline{W}_q)$$

It shows that the random code capacity $C_{\mathrm{ran}}(\mathfrak{W}^n)$ has the same value as for the traditional interference coordination but is also achieved in the case of symmetrizable channels.

3.1.3. Encoder-Decoder Coordination Based on Correlated Side Information

For the previous additional encoder-decoder coordination we assumed that both transmitter and receiver have access to a common random experiment. This seems to be a hard condition and one can think of a weaker version. Therefore, we allow transmitter and receiver each to have access to an own random experiment which are both correlated. In more detail, the correlated side information strategy is given by the following definition.

Definition 6. *A correlated $(n, M_n, n, Z_{\mathcal{X}}, Z_{\mathcal{Y}})$-code $\mathcal{C}_{corr}(\mathfrak{W}^n)$ for the AVC $\mathfrak{W}^n$ is given by $\mathcal{C}_{corr}(\mathfrak{W}^n) \coloneqq \{\mathcal{C}(z_{\mathcal{X}}^n, z_{\mathcal{Y}}^n) : z_{\mathcal{X}}^n \in \mathcal{Z}_{\mathcal{X}}^n, z_{\mathcal{Y}}^n \in \mathcal{Z}_{\mathcal{Y}}^n\}$ of deterministic (n, M_n)-codes*

$$\mathcal{C}(z_{\mathcal{X}}^n, z_{\mathcal{Y}}^n) \coloneqq \left\{ (x_m^n(z_{\mathcal{X}}^n), \mathcal{D}_m(z_{\mathcal{Y}}^n)) : m \in \mathcal{M} \right\}$$

together with random variables $Z_{\mathcal{X}} \in \mathcal{Z}_{\mathcal{X}}$ and $Z_{\mathcal{Y}} \in \mathcal{Z}_{\mathcal{Y}}$ distributed according to $p_{Z_{\mathcal{X}}} \in \mathcal{P}(\mathcal{Z}_{\mathcal{X}})$ and $p_{Z_{\mathcal{Y}}} \in \mathcal{P}(\mathcal{Z}_{\mathcal{Y}})$ with $I(Z_{\mathcal{X}}; Z_{\mathcal{Y}}) > 0$.

Thereby, the fact that the random variables $Z_{\mathcal{X}}$ and $Z_{\mathcal{Y}}$ are correlated is guaranteed by the (weak) condition $I(Z_{\mathcal{X}}; Z_{\mathcal{Y}}) > 0$. Note that in contrast to the additional encoder-decoder coordination based

on common randomness, the codewords and decoding sets now depend on a whole sequence of the random variables.

The next result states the capacity for the case of additional encoder-decoder coordination based on correlated side information at transmitter and receiver [54].

Theorem 3. *The correlated side information capacity $C_{corr}(\mathfrak{W}^n)$ of the AVC $\mathfrak{W}^n$ is*

$$C_{corr}(\mathfrak{W}^n) = \max_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{S})} I(p, \overline{W}_q)$$

The theorem shows that even if transmitter and receiver only have access to correlated versions of a random experiment, such side information is already sufficient to achieve the same rates as for the encoder-decoder coordination based on common randomness. Thus, correlated side information suffices to overcome symmetrizable channel conditions.

4. Bidirectional Relaying under Arbitrarily Varying Channels

In the previous section we established the concept of arbitrarily varying channels as a suitable model for communication in wireless networks which operate on the same resources as other coexisting systems. Here we use this framework and apply it to bidirectional relaying. There, a relay node establishes a bidirectional communication between two other nodes using a two-phase decode-and-forward protocol as shown in Figure 3. The initial MAC phase for arbitrarily varying channels is characterized by the AVMAC and therefore well understood, *cf.* [8–12,14]. Thus, it remains to study the succeeding BBC phase. Since in practical systems transmitters are usually limited in their transmit power, this requires the study of the *AVBBC under input and state constraints*, which is the main contribution of this paper.

Figure 3. Bidirectional relaying in a three-node network, where nodes 1 and 2 exchange their messages m_1 and m_2 with the help of the relay node using a decode-and-forward protocol.

4.1. Arbitrarily Varying Bidirectional Broadcast Channel

For the bidirectional broadcast phase we assume that the relay has successfully decoded both messages from the previous MAC phase. Now, the relay broadcasts an optimal re-encoded message in such a way that both nodes can decode the intended message using their own message from the previous phase as side information. The transmission is affected by a channel which varies arbitrarily in an unknown manner from symbol to symbol during the whole transmission of a codeword. We model this behavior with the help of a finite state set $\mathcal{S}$. Further, let $\mathcal{X}$ and $\mathcal{Y}_i$, $i = 1, 2$, be finite input and output sets. Then, for a fixed state sequence $s^n \in \mathcal{S}^n$ of length n and input and output sequences $x^n \in \mathcal{X}^n$ and $y_i^n \in \mathcal{Y}_i^n$, $i = 1, 2$, the discrete memoryless broadcast channel is given by $W^{\otimes n}(y_1^n, y_2^n | x^n, s^n) \coloneqq \prod_{k=1}^{n} W(y_{1,k}, y_{2,k} | x_k, s_k)$.

Definition 7. *The discrete memoryless* arbitrarily varying broadcast channel $\mathfrak{W}^n$ *is the family*

$$\mathfrak{W}^n := \left\{ W^{\otimes n} : \mathcal{X}^n \times \mathcal{S}^n \to \mathcal{P}(\mathcal{Y}_1^n \times \mathcal{Y}_2^n) \right\}_{n \in \mathbb{N}, s^n \in \mathcal{S}^n}$$

Since we do not allow any cooperation between the receiving nodes, it is sufficient to consider the marginal transition probabilities $W_i^{\otimes n}(y_i^n | x^n, s^n) = \prod_{k=1}^{n} W_{i,k}(y_{i,k} | x_k, s_k)$, $i = 1, 2$, only. Further, for any probability distribution $q \in \mathcal{P}(\mathcal{S})$ we denote the averaged broadcast channel by

$$\overline{W}_q(y_1, y_2 | x) := \sum_{s \in \mathcal{S}} W(y_1, y_2 | x, s) q(s) \tag{3}$$

and the corresponding averaged marginal channels by $\overline{W}_{1,q}(y_1 | x)$ and $\overline{W}_{2,q}(y_2 | x)$.

Further, we will need the concept of symmetrizability for the AVBBC, which is an extension of the one for the single-user AVC introduced in [4], *cf.* also Definition 4.

Definition 8. *An AVBBC $\mathfrak{W}^n$ is $\mathcal{Y}_i$-symmetrizable if for some channel $U_i : \mathcal{X} \to \mathcal{P}(\mathcal{S})$*

$$\sum_{s \in \mathcal{S}} W_i(y_i | x, s) U_i(s | x') = \sum_{s \in \mathcal{S}} W_i(y_i | x', s) U_i(s | x) \tag{4}$$

holds for every $x, x' \in \mathcal{X}$ and $y_i \in \mathcal{Y}_i$, $i = 1, 2$.

4.1.1. Input and State Constraints

Since transmitter and possible interferers are usually limited in their transmit powers, we impose constraints on the permissible input and state sequences. We follow [4] and define cost functions $g(x)$ and $l(s)$ on $\mathcal{X}$ and $\mathcal{S}$, respectively. For convenience, we assume that $\min_{x \in \mathcal{X}} g(x) = \min_{s \in \mathcal{S}} l(s) = 0$ and define $g_{\max} := \max_{x \in \mathcal{X}} g(x)$ and $l_{\max} := \max_{s \in \mathcal{S}} l(s)$. For given $x^n = (x_1, ..., x_n)$ and $s^n = (s_1, ..., s_n)$ we set

$$g(x^n) := \frac{1}{n} \sum_{k=1}^{n} g(x_k) \tag{5a}$$

$$l(s^n) := \frac{1}{n} \sum_{k=1}^{n} l(s_k) \tag{5b}$$

Further, for notational convenience we define the costs caused by given probability distributions $p \in \mathcal{P}(\mathcal{X})$ and $q \in \mathcal{P}(\mathcal{S})$ as

$$g(p) = \sum_{x \in \mathcal{X}} p(x) g(x) = \mathbb{E}_p[g(p)] \quad \text{and} \quad l(q) = \sum_{s \in \mathcal{S}} q(s) l(s) = \mathbb{E}_q[l(q)]$$

and observe that, if we consider types, these definitions immediately yield

$$g(x^n) = g(P_{x^n}) \quad \text{and} \quad l(s^n) = l(P_{s^n})$$

for every $x^n \in \mathcal{X}^n$ and every $s^n \in \mathcal{S}^n$, respectively, *cf.* also [4].

This allows us to define the set of all state sequences of length n that satisfy a given state constraint Λ by

$$\mathcal{S}_\Lambda^n := \left\{ s^n \in \mathcal{S}^n : \frac{1}{n} \sum_{k=1}^{n} l(s_k) = \mathbb{E}_{P_{s^n}}[l(s^n)] \leq \Lambda \right\}$$

Furthermore, the set of all probability distributions $q \in \mathcal{P}(\mathcal{S})$ that satisfy $\mathbb{E}_q[l(q)] \leq \Lambda$ is given by

$$\mathcal{P}(\mathcal{S}, \Lambda) := \left\{ q : q \in \mathcal{P}(\mathcal{S}), \mathbb{E}_q[l(q)] \leq \Lambda \right\}$$

In [52] it is shown that an AVBBC $\mathfrak{W}^n$ (without state constraint) has a capacity region whose interior is empty if the AVBBC $\mathfrak{W}^n$ is $\mathcal{Y}_1$-symmetrizable or $\mathcal{Y}_2$-symmetrizable. If we impose a state constraint, the situation changes significantly. Now, it is possible that the interior of the capacity region is non-empty even if the AVBBC $\mathfrak{W}^n$ is $\mathcal{Y}_i$-symmetrizable in the sense of Definition 8. Rather, $\mathcal{Y}_i$-symmetrizability enters the picture via

$$\Lambda_i(P_X) = \begin{cases} \min\limits_{U_i \in \mathcal{U}_i} \sum\limits_{x \in \mathcal{X}} \sum\limits_{s \in \mathcal{S}} P_X(x) U_i(s|x) l(s) & \text{if } \mathcal{U}_i \neq \emptyset \\ \infty & \text{if } \mathcal{U}_i = \emptyset \end{cases} \tag{6}$$

$i = 1, 2$, which indicates whether the symmetrization violates the imposed state constraint or not. Thereby, $\mathcal{U}_i$ is the set of all channels $U_i : \mathcal{X} \to \mathcal{P}(\mathcal{S})$ which satisfy (4). For given type P_X the quantity $\Lambda_i(P_X)$ is called *symmetrizability costs* and can be interpreted as the minimum costs that are needed to symmetrize the AVBBC $\mathfrak{W}^n$. Clearly, if $\mathfrak{W}^n$ is $\mathcal{Y}_i$-symmetrizable, then $\mathcal{U}_i \neq \emptyset$ and $\Lambda_i(P_X)$ is finite. Further, if $\mathfrak{W}^n$ is non-$\mathcal{Y}_i$-symmetrizable, then $\mathcal{U}_i = \emptyset$, and we set $\Lambda_i(P_X) = \infty$ for convenience.

4.1.2. Coordination Strategies

We consider the standard model with a block code of arbitrary but sufficient fixed length n. Let $\mathcal{M}_i := \{1, ..., M_{i,n}\}$ be the message set of node i, $i = 1, 2$, which is also known at the relay node. Further, we use the abbreviation $\mathcal{M} := \mathcal{M}_1 \times \mathcal{M}_2$.

First, we introduce the traditional approach without additional coordination which is based on a deterministic coding strategy with pre-specified encoder and decoders at the relay and receivers.

Definition 9. *A deterministic $(n, M_{1,n}, M_{2,n})$-code $\mathcal{C}_{det}(\mathfrak{W}^n)$ of length n for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ is a family*

$$\mathcal{C}_{det}(\mathfrak{W}^n) := \left\{ (x_m^n, \mathcal{D}_{m_2|m_1}^{(1)}, \mathcal{D}_{m_1|m_2}^{(2)}) : m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2 \right\}$$

with codewords

$$x_m^n \in \mathcal{X}^n \qquad \text{with } g(x_m^n) \leq \Gamma$$

one for each message $m = (m_1, m_2)$, satisfying the input constraint Γ, and decoding sets at nodes 1 and 2

$$\mathcal{D}_{m_2|m_1}^{(1)} \subseteq \mathcal{Y}_1^n \quad \text{and} \quad \mathcal{D}_{m_1|m_2}^{(2)} \subseteq \mathcal{Y}_2^n$$

for all $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$. For given m_1 at node 1 the decoding sets must be disjoint, i.e., $\mathcal{D}_{m_2|m_1}^{(1)} \cap \mathcal{D}_{m_2'|m_1}^{(1)} = \emptyset$ for $m_2' \neq m_2$, and similarly for given m_2 at node 2 the decoding sets must satisfy $\mathcal{D}_{m_1|m_2}^{(2)} \cap \mathcal{D}_{m_1'|m_2}^{(2)} = \emptyset$ for $m_1' \neq m_1$.

When x_m^n with $m = (m_1, m_2)$ and $g(x_m^n) \leq \Gamma$ has been sent, and y_1^n and y_2^n have been received at nodes 1 and 2, the decoder at node 1 is in error if y_1^n is not in $\mathcal{D}_{m_2|m_1}^{(1)}$. Accordingly, the decoder at node 2

is in error if y_2^n is not in $\mathcal{D}_{m_1|m_2}^{(2)}$. This allows us to define the probability of error for the deterministic code $\mathcal{C}_{\text{det}}(\mathfrak{W}^n)$ for given message $m = (m_1, m_2)$ and state sequence $s^n \in \mathcal{S}_\Lambda^n$, *i.e.*, it satisfies the state constraint Λ, as

$$e(m, s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) := W^{\otimes n}\big((\mathcal{D}_{m_2|m_1}^{(1)} \times \mathcal{D}_{m_1|m_2}^{(2)})^c|x_m^n, s^n\big)$$

and the corresponding marginal probabilities of error at nodes 1 and 2 as $e_1(m, s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) :=$ $W_1^{\otimes n}((\mathcal{D}_{m_2|m_1}^{(1)})^c|x_m^n, s^n)$ and $e_2(m, s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) := W_2^{\otimes n}((\mathcal{D}_{m_1|m_2}^{(2)})^c|x_m^n, s^n)$, respectively. Thus, the *average probability of error* for state sequence $s^n \in \mathcal{S}_\Lambda^n$ is given by

$$\bar{e}(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) := \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e(m, s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))$$

and the corresponding marginal average probability of error at node i by $\bar{e}_i(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) :=$ $\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_i(m, s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))$, $i = 1, 2$. Clearly, we always have $\bar{e}(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \leq \bar{e}_1(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) + \bar{e}_2(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n))$.

For given $0 < \lambda_n < 1$, the code $\mathcal{C}_{\text{det}}(\mathfrak{W}^n)$ is called a $(n, M_{1,n}, M_{2,n}, \lambda_n)$-code (with average probability of error λ_n) for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ if

$$\bar{e}(s^n|\mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \leq \lambda_n \qquad \text{for all } s^n \in \mathcal{S}_\Lambda^n$$

Definition 10. *A rate pair* $(R_1, R_2) \in \mathbb{R}_+^2$ *is said to be* deterministically achievable *for the AVBBC* $\mathfrak{W}^n$ *under input constraint* Γ *and state constraint* Λ *if for any* $\delta > 0$ *there exists an* $n(\delta) \in \mathbb{N}$ *and a sequence of deterministic* $(n, M_{1,n}, M_{2,n}, \lambda_n)$-*codes* $\{\mathcal{C}_{det}^{(n)}(\mathfrak{W}^n)\}_{n \in \mathbb{N}}$ *with codewords* x_{m_1, m_2}^n, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$, *each satisfying* $g(x_{m_1,m_2}^n) \leq \Gamma$, *such that for all* $n \geq n(\delta)$ *we have*

$$\frac{1}{n} \log M_{1,n} \geq R_2 - \delta \quad and \quad \frac{1}{n} \log M_{2,n} \geq R_1 - \delta$$

while

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}(s^n|\mathcal{C}_{det}(\mathfrak{W}^n)) \leq \lambda_n$$

with $\lambda_n \to 0$ *as* $n \to \infty$. *The set of all achievable rate pairs is the* deterministic code capacity region *of the AVBBC* $\mathfrak{W}^n$ *under input constraint* Γ *and state constraint* Λ *and is denoted by* $\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma, \Lambda)$.

If $\Gamma \geq g_{\max}$ or $\Lambda \geq l_{\max}$, then the input or state sequences are not restricted by the corresponding constraint, respectively. Consequently, we denote the capacity region with state constraint and no input constraint by $\mathcal{R}_{\text{det}}(\mathfrak{W}^n|g_{\max}, \Lambda)$ and the capacity region with input constraint and no state constraint by $\mathcal{R}_{\text{det}}(\mathfrak{W}^n|\Gamma, l_{\max})$.

Remark 3. *The definitions above require that we have to find codes such that the average probability of error goes to zero as the block length tends to infinity simultaneously for all state sequences that fulfill the state constraint. This means the codes are universal with respect to the state sequence.*

Next, we introduce the encoder-decoder coordination based on common randomness which is specified by a random code, where the encoder and the decoders are chosen according to a common random experiment whose outcome has to be known at all nodes in advance.

Definition 11. *A random $(n, M_{1,n}, M_{2,n}, Z)$-code $\mathcal{C}_{ran}(\mathfrak{W}^n)$ of length n for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ is given by a family $\mathcal{C}_{ran}(\mathfrak{W}^n) := \{\mathcal{C}(z) : z \in \mathcal{Z}\}$ of deterministic $(n, M_{1,n}, M_{2,n})$-codes*

$$\mathcal{C}(z) := \big\{\big(x_m^n(z), \mathcal{D}_{m_2|m_1}^{(1)}(z), \mathcal{D}_{m_1|m_2}^{(2)}(z)\big) : m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2\big\}$$

together with a random variable $Z \in \mathcal{Z}$ distributed according to $p_Z \in \mathcal{P}(\mathcal{Z})$. Thereby, each $\mathcal{C}(z)$ is a deterministic code in the sense of Definition 9, which means that each $\mathcal{C}(z)$ satisfies the input and state constraints individually.

Then, the average probability of error of the random code $\mathcal{C}_{ran}(\mathfrak{W}^n)$ for given state sequence $s^n \in \mathcal{S}_\Lambda^n$ is given by

$$\bar{e}(s^n|\mathcal{C}_{ran}(\mathfrak{W}^n)) := \mathbb{E}_Z[\bar{e}(s^n|\mathcal{C}(Z))]$$

and accordingly the corresponding marginal average probability of error at node i by $\bar{e}_i(s^n|\mathcal{C}_{ran}(\mathfrak{W}^n)) := \mathbb{E}_Z[\bar{e}_i(s^n|\mathcal{C}(Z))]$, $i = 1, 2$. For given $0 < \lambda_n < 1$, the random code $\mathcal{C}_{ran}(\mathfrak{W}^n)$ is called a $(n, M_{1,n}, M_{2,n}, Z, \lambda_n)$-code (with average probability of error λ_n) for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ if

$$\bar{e}(s^n|\mathcal{C}_{ran}(\mathfrak{W}^n)) \leq \lambda_n \qquad \text{for all } s^n \in \mathcal{S}_\Lambda^n$$

The definitions of a *randomly achievable rate pair* under input and state constraints and the *random code capacity region* $\mathcal{R}_{ran}(\mathfrak{W}^n|\Gamma, \Lambda)$ under input and state constraints follow accordingly.

4.2. Encoder-Decoder Coordination Based on Common Randomness

Here, we derive the random code capacity region of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ. This characterizes the scenario, where transmitter and receivers can coordinate their choice of encoder and decoders based on common randomness. For this purpose we define the region

$$\mathcal{R}(P_X|\Lambda) := \big\{(R_1, R_2) \in \mathbb{R}_+^2 : R_1 \leq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_X, \overline{W}_{1,q}) \tag{7}$$
$$R_2 \leq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_X, \overline{W}_{2,q})\big\}$$

for joint probability distributions $\{P_X(x)\overline{W}_q(y_1, y_2|x)\}_{q \in \mathcal{P}(\mathcal{S}, \Lambda)}$.

Theorem 4. *The random code capacity region $\mathcal{R}_{ran}(\mathfrak{W}^n|\Gamma, \Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ is*

$$\mathcal{R}_{ran}(\mathfrak{W}^n|\Gamma, \Lambda) = \bigcup_{P_X : g(P_X) \leq \Gamma} \mathcal{R}(P_X|\Lambda)$$

In the following we give the proof of the random code capacity region where the achievability part is mainly based on an extension of *Ahlswede's robustification technique* [55,56].

4.2.1. Compound Bidirectional Broadcast Channel

As in [51] for the AVBBC $\mathfrak{W}^n$ without constraints on input and states, we start with a construction of a suitable compound BBC, where the key idea is to restrict it in an appropriate way. Having the state constraint Λ in mind, it is reasonable to restrict our attention to all probability distributions $q \in \mathcal{P}(\mathcal{S}, \Lambda)$. Let us consider the family of averaged broadcast channels, *cf.* (3),

$$\left\{ \overline{W}_q(y_1, y_2|x) \right\}_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} \tag{8}$$

and observe that this already corresponds to a compound BBC where each permissible probability distribution $q \in \mathcal{P}(\mathcal{S}, \Lambda)$ parametrizes one element of the compound channel which we denote by $\overline{\mathfrak{W}}$ in the following. The capacity region of the compound BBC is known and can be found in [27]. It is shown that for given input distribution P_X all rate pairs (R_1, R_2) satisfying $(R_1, R_2) \in \mathcal{R}(P_X|\Lambda)$, *cf.* (7), are deterministically achievable. In particular, this is valid for a input distribution P_X that satisfies the input constraint $g(P_X) \leq \Gamma$.

In more detail, in [27] it is shown that there exists a deterministic code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ such that all rate pairs $(R_1, R_2) \in \mathcal{R}(P_X|\Lambda)$ are achievable while the average probability of error can be bounded from above by

$$\bar{e}(q|\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})) := \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \overline{W}_q^{\otimes n}\left((\mathcal{D}_{m_2|m_1}^{(1)} \times \mathcal{D}_{m_1|m_2}^{(2)})^c | x_m^n\right) \leq \lambda_{\overline{\mathfrak{W}}, n} \qquad \text{for all } q \in \mathcal{P}(\mathcal{S}, \Lambda)$$

with $\lambda_{\overline{\mathfrak{W}}, n} = \lambda_{\overline{\mathfrak{W}}, 1n} + \lambda_{\overline{\mathfrak{W}}, 2n}$ where $\lambda_{\overline{\mathfrak{W}}, in}$ is the average probability of error at node i, $i = 1, 2$. Moreover, for n large enough, we have

$$\lambda_{\overline{\mathfrak{W}}, in} = (n+1)^{|\mathcal{X}||\mathcal{Y}_i|} 2^{-n\frac{c\delta^2}{2}} + \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_i|}}{1 - (n+1)^{|\mathcal{X}|} 2^{-nc\delta^2}} 2^{-n\frac{\tau}{8}}$$

which decreases exponentially fast for increasing block length n. Thereby, $\delta > 0$, $\tau > 0$, and $c > 0$ are constants independent of n, *cf.* [27].

Together with (3) this immediately implies that for $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ the average probability of a successful transmission over the compound BBC $\overline{\mathfrak{W}}$ is bounded from below by

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \overline{W}_q^{\otimes n}(\mathcal{D}_{m_2|m_1}^{(1)} \times \mathcal{D}_{m_1|m_2}^{(2)} | x_m^n) > 1 - \lambda_{\overline{\mathfrak{W}}, n}$$

or equivalently by

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{s^n \in \mathcal{S}^n} W^{\otimes n}(\mathcal{D}_{m_2|m_1}^{(1)} \times \mathcal{D}_{m_1|m_2}^{(2)} | x_m^n, s^n) q^{\otimes n}(s^n) > 1 - \lambda_{\overline{\mathfrak{W}}, n}$$

for all $q^{\otimes n} = \prod_{k=1}^n q$ and $q \in \mathcal{P}(\mathcal{S}, \Lambda)$.

4.2.2. Robustification

As in [51] for the AVBBC without state constraints, we use the deterministic code $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ to construct a random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ.

Let Π_n be the group of permutations acting on $\{1, 2, ..., n\}$. For given sequence $s^n = (s_1, s_2, ..., s_n) \in \mathcal{S}^n$ and permutation $\pi \in \Pi_n : \mathcal{S}^n \to \mathcal{S}^n$, we denote the permuted sequence $(s_{\pi(1)}, s_{\pi(2)}, ..., s_{\pi(n)}) \in \mathcal{S}^n$ by $\pi(s^n)$. Further, we denote the inverse permutation by π^{-1} so that $\pi^{-1}(\pi(s^n)) = s^n$ since π is bijective.

Theorem 5 (Robustification technique). *Let $f : \mathcal{S}^n \to [0, 1]$ be a function such that for some $\alpha \in (0, 1)$ the inequality*

$$\sum_{s^n \in \mathcal{S}^n} f(s^n) q^{\otimes n}(s^n) > 1 - \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda) \tag{9}$$

holds where $\mathcal{P}_0(n, \mathcal{S}, \Lambda) := \{q \in \mathcal{P}_0(n, \mathcal{S}) : \mathbb{E}_q[l(q)] \leq \Lambda\}$. Then it also holds

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi(s^n)) > 1 - (n+1)^{|\mathcal{S}|} \alpha \qquad \text{for all } s^n \in \mathcal{S}_\Lambda^n \tag{10}$$

Proof. The proof is a modification of the corresponding proof in [56], where a similar result is given without constraints on the sequences of states. First, we observe that (9) is equivalent to

$$\sum_{s^n \in \mathcal{S}^n} \left(1 - f(s^n)\right) q^{\otimes n}(s^n) \leq \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda) \tag{11}$$

Since each $\pi \in \Pi_n$ is bijective and because $q^{\otimes n}(\pi(s^n)) = q^{\otimes n}(s^n)$ for all $s^n \in \mathcal{S}^n$, we obtain from (11)

$$\begin{aligned}
\alpha &\geq \sum_{s^n \in \mathcal{S}^n} \left(1 - f(\pi(s^n))\right) q^{\otimes n}(\pi(s^n)) \\
&= \sum_{s^n \in \mathcal{S}^n} \left(1 - f(\pi(s^n))\right) q^{\otimes n}(s^n) \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda) \text{ and all } \pi \in \Pi_n
\end{aligned} \tag{12}$$

Therefore, averaging (12) over Π_n yields

$$\sum_{s^n \in \mathcal{S}^n} \frac{1}{n!} \sum_{\pi \in \Pi_n} \left(1 - f(\pi(s^n))\right) q^{\otimes n}(s^n) \leq \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda) \tag{13}$$

Since $1 - \frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi(s^n)) \geq 0$, restricting the state sequences to $\mathcal{T}_q^n$ we get from (13)

$$\sum_{s^n \in \mathcal{T}_q^n} \frac{1}{n!} \sum_{\pi \in \Pi_n} \left(1 - f(\pi(s^n))\right) q^{\otimes n}(s^n) \leq \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda)$$

which is equivalent to

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} \left(1 - f(\pi(s^n))\right) q^{\otimes n}(\mathcal{T}_q^n) \leq \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda) \text{ and all } s^n \in \mathcal{T}_q^n \tag{14}$$

because for $s^n \in \mathcal{T}_q^n$, the term $\frac{1}{n!} \sum_{\pi \in \Pi_n}(1 - f(\pi(s^n)))$ does not depend on s^n. Since $\mathcal{T}_q^n \geq (n+1)^{-|\mathcal{S}|}$, *cf.* [3], Equation (14) implies

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} \left(1 - f(\pi(s^n))\right) \leq (n+1)^{|\mathcal{S}|} \alpha \qquad \text{for all } q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda) \text{ and all } s^n \in \mathcal{T}_q^n \tag{15}$$

Obviously, we have $\mathcal{S}_\Lambda^n = \bigcup_{q \in \mathcal{P}_0(n, \mathcal{S}, \Lambda)} \mathcal{T}_q^n$ so that (15) shows that

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi(s^n)) > 1 - (n+1)^{|\mathcal{S}|} \alpha \qquad \text{for all } s^n \in \mathcal{S}_\Lambda^n$$

which completes the proof of the theorem. $\qquad \square$

With the robustification technique and

$$f\big(\pi(s^n)\big) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W^{\otimes n}\big(\mathcal{D}^{(1)}_{m_2|m_1} \times \mathcal{D}^{(2)}_{m_1|m_2}|x^n_m, \pi(s^n)\big)$$

we immediately obtain a random $(n, M_{1,n}, M_{2,n}, \Pi_n)$-code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ, which is given by the family

$$\mathcal{C}_{\text{ran}}(\mathfrak{W}^n) = \big\{(\pi^{-1}(x^n_m), \pi^{-1}(\mathcal{D}^{(1)}_{m_2|m_1}), \pi^{-1}(\mathcal{D}^{(2)}_{m_1|m_2})) : m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2, \pi \in \Pi_n\big\} \quad (16)$$

where the permutations π are uniformly distributed on Π_n and

$$\pi^{-1}(\mathcal{D}^{(1)}_{m_2|m_1}) = \bigcup_{y^n_1 \in \mathcal{D}^{(1)}_{m_2|m_1}} \pi^{-1}(y^n_1) \quad \text{and} \quad \pi^{-1}(\mathcal{D}^{(2)}_{m_1|m_2}) = \bigcup_{y^n_2 \in \mathcal{D}^{(2)}_{m_1|m_2}} \pi^{-1}(y^n_2)$$

Since Π_n is the group of permutations of size n, the cardinality of Π_n is $n!$ so that the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ consists of $n!$ deterministic $(n, M_{1,n}, M_{2,n})$-codes.

From the robustification technique follows that the average probability of error of $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ is bounded from above by

$$\bar{e}(s^n|\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)) \leq (n+1)^{|\mathcal{S}|} \lambda_{\overline{\mathfrak{W}},n} =: \lambda_{\mathfrak{W}^n\text{ran},n} \qquad \text{for all } s^n \in \mathcal{S}^n_\Lambda \quad (17)$$

Moreover, from the construction it is clear that for given input P_X, the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ achieves for the AVBBC $\mathfrak{W}^n$ the same rate pairs as $\mathcal{C}_{\text{det}}(\overline{\mathfrak{W}})$ for the compound BBC $\overline{\mathfrak{W}}$ as specified in (7). Finally, taking the union over all input distributions P_X that satisfy the input constraint $g(P_X) \leq \Gamma$ establishes the achievability of the random code capacity $\mathcal{R}_{\text{ran}}(\mathfrak{W}^n|\Gamma, \Lambda)$ as stated in Theorem 4.

4.2.3. Converse

It remains to show that the presented random coding strategy actually achieves all possible rate pairs so that it is optimal in the sense that no other rate pairs are achievable.

As a first step, it is easy to show that the average probability of error for the random code $\mathcal{C}_{\text{ran}}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ equals the average probability of error for the random code for the compound BBC $\overline{\mathfrak{W}}$. Hence, it is clear that we cannot achieve higher rates as for the constructed compound BBC $\overline{\mathfrak{W}}$ with random codes. The deterministic rates of the compound channel can be found in [27]. Additionally, as in [57] for the single-user compound channel, it can be easily shown that for the compound BBC $\overline{\mathfrak{W}}$ the achievable rates for deterministic and random codes are equal. Since the constructed random code for the AVBBC $\mathfrak{W}^n$ already achieves these rates, the converse is established.

This finishes the proof of Theorem 4 and therewith the random code capacity region $\mathcal{R}_{\text{ran}}(\mathfrak{W}^n|\Gamma, \Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ.

4.3. No Additional Coordination

A random coding strategy as constructed in the previous section requires *common randomness* between all nodes, since the encoder and the decoders depend all on the same random permutation which has to be known at all nodes in advance. If this kind of resource is not available, one is interested in deterministic strategies. In this section, we derive the deterministic code capacity region of the AVBBC with constraints on input and states.

Theorem 6. *If* $\max_{P_X:g(P_X)\leq\Gamma}\Lambda_i(P_X) > \Lambda$, $i = 1,2$, *then the deterministic code capacity region* $\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma,\Lambda)$ *of the AVBBC* $\mathfrak{W}^n$ *under input constraint* Γ *and state constraint* Λ *is*

$$\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma,\Lambda) = \bigcup_{\substack{P_X:\,g(P_X)\leq\Gamma \\ \Lambda_i(P_X)>\Lambda,\,i=1,2}} \mathcal{R}(P_X|\Lambda)$$

If $\max_{P_X:g(P_X)\leq\Gamma}\Lambda_1(P_X) < \Lambda$ *or* $\max_{P_X:g(P_X)\leq\Gamma}\Lambda_2(P_X) < \Lambda$, *then* $int(\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma,\Lambda)) = \emptyset$.

From the theorem we immediately obtain the deterministic code capacity region of the AVBBC $\mathfrak{W}^n$ with state constraint Λ and no input constraint, *i.e.*, $\mathcal{R}_{det}(\mathfrak{W}^n|g_{\max},\Lambda)$.

Corollary 1. *If* $\max_{P_X}\Lambda_i(P_X) > \Lambda$, $i = 1,2$, *then the deterministic code capacity region* $\mathcal{R}_{det}(\mathfrak{W}^n|g_{max},\Lambda)$ *of the AVBBC* $\mathfrak{W}^n$ *with state constraint* Λ *and no input constraint is given by*

$$\mathcal{R}_{det}(\mathfrak{W}^n|g_{max},\Lambda) = \bigcup_{P_X:\,\Lambda_i(P_X)>\Lambda,i=1,2} \mathcal{R}(P_X|\Lambda)$$

If $\max_{P_X}\Lambda_1(P_X) < \Lambda$ *or* $\max_{P_X}\Lambda_2(P_X) < \Lambda$, *then* $int(\mathcal{R}_{det}(\mathfrak{W}^n|g_{max},\Lambda)) = \emptyset$.

We observe that the deterministic code capacity region $\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma,\Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ displays a dichotomy behavior similarly as in the unconstrained case [51]: it either equals a non-empty region or has an empty interior. Unfortunately, this knowledge cannot be exploited to prove the corresponding deterministic code capacity region since, as already observed in [4] for the single-user AVC, *Ahlswede's elimination technique* [2] does not work anymore if constraints are imposed on the permissible codewords and sequences of states. Consequently, to prove Theorem 6 we need a proof idea which does not rely on this technique. In the following subsections we present the proof which is therefore mainly based on an extension of [4].

4.3.1. Symmetrizability

The following lemma shows that under state constraint Λ no code with codewords of type P_X satisfying $\Lambda_1(P_X) < \Lambda$ or $\Lambda_2(P_X) < \Lambda$ can be good.

Lemma 1. *For a* $\mathcal{Y}_1$-*symmetrizable AVBBC* $\mathfrak{W}^n$ *any deterministic code* $\mathcal{C}_{det}(\mathfrak{W}^n)$ *of block length* n *with codewords* $x^n_{m_1,m_2}$, $m_1 = 1,...,M_{1,n}$, $m_2 = 1,...,M_{2,n}$, *each of type* P_X *with* $\Lambda_1(P_X) < \Lambda$, *and* $M_{2,n} \geq 2$ *has*

$$\max_{s^n:l(s^n)\leq\Lambda} \bar{e}_1(s^n|\mathcal{C}_{det}(\mathfrak{W}^n)) \geq \frac{M_{2,n}-1}{2M_{2,n}} - \frac{1}{n}\frac{l^2_{max}}{(\Lambda-\Lambda_1(P_X))^2}$$

Similarly, for a $\mathcal{Y}_2$-*symmetrizable AVBBC* $\mathfrak{W}^n$ *any deterministic code* $\mathcal{C}_{det}(\mathfrak{W}^n)$ *of block length* n *with codewords* $x^n_{m_1,m_2}$, $m_1 = 1,...,M_{1,n}$, $m_2 = 1,...,M_{2,n}$, *each of type* P_X *with* $\Lambda_2(P_X) < \Lambda$, *and* $M_{1,n} \geq 2$ *has*

$$\max_{s^n:l(s^n)\leq\Lambda} \bar{e}_2(s^n|\mathcal{C}_{det}(\mathfrak{W}^n)) \geq \frac{M_{1,n}-1}{2M_{1,n}} - \frac{1}{n}\frac{l^2_{max}}{(\Lambda-\Lambda_2(P_X))^2}$$

Proof. The proof can be found in Appendix A.1. □

Remark 4. *The lemma indicates that for a successful transmission using codewords of type P_X the symmetrizability costs $\Lambda_i(P_X)$, $i = 1, 2$, have to exceed the permissible (or available) costs Λ, since otherwise the AVBBC $\mathfrak{W}^n$ can be symmetrized, which prohibits any reliable or error-free communication. This already establishes the second part of Theorem 6 and therewith characterizes when $int(\mathcal{R}_{det}(\mathfrak{W}^n|\Gamma, \Lambda)) = \emptyset$.*

4.3.2. Positive Rates

Next, we present a coding strategy with codewords of type P_X that achieves the desired rates as specified in Theorem 6 if the symmetrizability costs exceed the permissible costs, *i.e.*, $\Lambda_1(P_X) > \Lambda$ and $\Lambda_2(P_X) > \Lambda$. Fortunately, we are in the same position as in the single-user AVC [4]: the coding strategy for the AVBBC without constraints [52] must only be slightly modified to apply also to the AVBBC with constraints.

We need codewords $x^n_{m_1, m_2}$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$ with the following properties.

Lemma 2. *For any $\epsilon > 0$, $n \geq n_0(\epsilon)$, $M_{i,n} \geq 2^{n\epsilon}$, $i = 1, 2$, and given type P_X, there exist codewords $x^n_{m_1, m_2} \in \mathcal{T}^n_X$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$ such that for every $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n_\Lambda$, and every joint type $P_{XX'S}$, with $R_1 = \frac{1}{n} \log M_{2,n}$ and $R_2 = \frac{1}{n} \log M_{1,n}$, we have for each fixed $m_1 \in \mathcal{M}_1$ the following properties*

$$\left|\{m'_2 : (x^n, x^n_{m_1, m'_2}, s^n) \in \mathcal{T}^n_{XX'S}\}\right| \leq \exp\left(n(|R_1 - I(X'; XS)|^+ + \epsilon)\right) \tag{18a}$$

$$\frac{1}{M_{2,n}} \left|\{m_2 : (x^n_{m_1, m_2}, s^n) \in \mathcal{T}^n_{XS}\}\right| \leq \exp\left(-n\frac{\epsilon}{2}\right) \qquad \text{if } I(X; S) > \epsilon \tag{18b}$$

$$\frac{1}{M_{2,n}} \left|\{m_2 : (x^n_{m_1, m_2}, x^n_{m_1, m'_2}, s^n) \in \mathcal{T}^n_{XX'S} \text{ for some } m'_2 \neq m_2\}\right| \leq \exp\left(-n\frac{\epsilon}{2}\right)$$

$$\text{if } I(X; X'S) - |R_1 - I(X'; S)|^+ > \epsilon \tag{18c}$$

where $|x|^+ = \max\{x, 0\}$, and further for each fixed $m_2 \in \mathcal{M}_2$

$$\left|\{m'_1 : (x^n, x^n_{m'_1, m_2}, s^n) \in \mathcal{T}^n_{XX'S}\}\right| \leq \exp\left(n(|R_2 - I(X'; XS)|^+ + \epsilon)\right) \tag{18d}$$

$$\frac{1}{M_{1,n}} \left|\{m_1 : (x^n_{m_1, m_2}, s^n) \in \mathcal{T}^n_{XS}\}\right| \leq \exp\left(-n\frac{\epsilon}{2}\right) \qquad \text{if } I(X; S) > \epsilon \tag{18e}$$

$$\frac{1}{M_{1,n}} \left|\{m_1 : (x^n_{m_1, m_2}, x^n_{m'_1, m_2}, s^n) \in \mathcal{T}^n_{XX'S} \text{ for some } m'_1 \neq m_1\}\right| \leq \exp\left(-n\frac{\epsilon}{2}\right)$$

$$\text{if } I(X; X'S) - |R_2 - I(X'; S)|^+ > \epsilon \tag{18f}$$

Proof. The proof can be found in Appendix A.2. □

We follow [4] and define the decoding sets similarly as for the single-user AVC under input and state constraints. Therefore, we define the set

$$\mathcal{D}_{\eta_i}(\Lambda) = \left\{P_{XSY_i} : D(P_{XSY_i} \| P_X \otimes P_S \otimes W_i) \leq \eta_i, l(P_S) \leq \Lambda\right\}, \qquad i = 1, 2$$

Then, the decoding sets at node 1 are specified as follows.

Definition 12. *For given codewords $x^n_{m_1,m_2} \in \mathcal{T}^n_X$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$, and $\eta_1 > 0$ we have $y^n_1 \in \mathcal{D}^{(1)}_{m_2|m_1}$ if and only if*

(i) *there exists an $s^n \in \mathcal{S}^n_\Lambda$ such that $P_{x^n_{m_1,m_2},s^n,y^n_1} \in \mathcal{D}_{\eta_1}(\Lambda)$*

(ii) *for each codeword $x^n_{m_1,m'_2}$ with $m'_2 \neq m_2$ which satisfies $P_{x^n_{m_1,m'_2},s'^n,y^n_1} \in \mathcal{D}_{\eta_1}(\Lambda)$ for some $s'^n \in \mathcal{S}^n_\Lambda$, we have $I(XY_1; X'|S) \leq \eta_1$ where X, X', S, Y_1 are dummy random variables such that $P_{XX'SY_1}$ equals the joint type of $(x^n_{m_1,m_2}, x^n_{m_1,m'_2}, s^n, y^n_1)$.*

The decoding sets at node 2 are defined accordingly with $\eta_2 > 0$. A key part is now to ensure that these decoding sets are unambiguously defined. This means that they are disjoint for small enough η_1 and η_2, which can be shown analogously to the single-user case [4]. Here is where the conditions on the symmetrizability costs, $\Lambda_i(P_X) > \Lambda$, $i = 1, 2$, come in.

Lemma 3. *Let $\alpha > 0$ and $\beta > 0$, then for a sufficiently small η_i, $i = 1, 2$, no quintuple of random variables X, X', S, S', and Y_i can simultaneously satisfy $P_X = P_{X'}$ with*

$$\Lambda_i(P_X) \geq \Lambda + \alpha \qquad and \qquad \min_{x \in \mathcal{X}} P_X(x) \geq \beta$$

and

$$P_{XSY_i} \in \mathcal{D}_{\eta_i}(\Lambda), \qquad P_{X'S'Y_i} \in \mathcal{D}_{\eta_i}(\Lambda), \tag{19a}$$

$$I(XY_i; X'|S) \leq \eta_i, \qquad I(X'Y_i; X|S') \leq \eta_i \tag{19b}$$

Proof. The proof can be found in Appendix A.3. □

So far we defined coding and decoding rules. Next, we show that codewords of type P_X with properties as given in Lemma 2 and decoding sets as given in Definition 12 suffices to achieve all rate pairs as specified by the region $\mathcal{R}(P_X|\Lambda)$, *cf.* (7).

Lemma 4. *Given $\Lambda > 0$ and arbitrarily small $\alpha > 0$, $\beta > 0$, and $\delta > 0$, for any type P_X satisfying*

$$\Lambda_i(P_X) \geq \Lambda + \alpha, \; i = 1, 2, \qquad \min_{x \in \mathcal{X}} P_X(x) \geq \beta$$

there exist a code $\mathcal{C}_{det}(\mathfrak{W}^n)$ of block length $n \geq n_0$ with codewords $x^n_{m_1,m_2} \in \mathcal{T}^n_X$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$, such that

$$\frac{1}{n} \log M_{1,n} > \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_X, \overline{W}_{2,q}) - \delta \qquad and \qquad \frac{1}{n} \log M_{2,n} > \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_X, \overline{W}_{1,q}) - \delta$$

while

$$\max_{s^n:l(s^n)\leq\Lambda} \bar{e}_i\big(s^n|\mathcal{C}_{det}(\mathfrak{W}^n)\big) \leq \exp(-n\gamma_i), \qquad i = 1, 2 \tag{20}$$

where n_0 and $\gamma_i > 0$ depend only on α, β, δ, and the AVBBC $\mathfrak{W}^n$.

Proof. The proof follows [4] (Lemma 5) where a similar result is shown for the single-user AVC.

Let $x_{m_1,m_2}^n \in \mathcal{T}_{\mathrm{X}}^n$, $m_1 = 1,...,M_{1,n}$, $m_2 = 1,...,M_{2,n}$, each satisfying the input constraint $g(x_{m_1,m_2}^n) \leq \Gamma$, be codewords with properties as specified in Lemma 2 (ϵ will be chosen later) and $R_1 = \frac{1}{n} \log M_{2,n}$, $R_2 = \frac{1}{n} \log M_{1,n}$ satisfying

$$\inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) - \delta < R_1 < \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) - \frac{2}{3}\delta \tag{21a}$$

$$\inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{2,q}) - \delta < R_2 < \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{2,q}) - \frac{2}{3}\delta \tag{21b}$$

Let the decoding sets $\mathcal{D}_{m_2|m_1}^{(1)}$ and $\mathcal{D}_{m_1|m_2}^{(2)}$ be as given in Definition 12. Then Lemma 3 ensures that η_1 and η_2 can be chosen small enough to ensure that the decoding sets are well defined.

Furthermore, $I(\mathrm{X}; \mathrm{Y}_i)$ is uniformly continuous in P_{XY_i} and divergence dominates the variational distance [3] so that we can choose η_i small enough to ensure that $P_{\mathrm{XSY}_i} \in \mathcal{D}_{\eta_i}(\Lambda)$ implies

$$I(\mathrm{X}; \mathrm{Y}_i) \geq \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{i,q}) - \frac{\delta}{3}, \qquad i = 1, 2 \tag{22}$$

In the following we carry out the analysis for the probability of error at node 1. Then the analysis for node 2 follows accordingly using the same arguments. Now, we establish an exponentially decreasing upper bound on the probability of error as postulated in (20) for node 1 for a fixed state sequence $s^n \in \mathcal{S}_{\Lambda}^n$.

For each $m_1 \in \mathcal{M}_1$ we first observe by Definition 12 of the decoding sets that y_1^n is erroneously decoded if decoding rule (i) or decoding rule (ii) is violated. More precisely, when message $m = (m_1, m_2)$ has been sent, then the decoder makes an error if $P_{x_m^n, s^n, y_1^n} \notin \mathcal{D}_{\eta_1}(\Lambda)$ or there exists a joint type $P_{\mathrm{XX'SY}_1}$ with $(x_{m_1,m_2}^n, x_{m_1,m_2'}^n, s^n, y_1^n) \in \mathcal{T}_{\mathrm{XX'SY}_1}^n$ for some $m_2' \neq m_2$ such that (a) $P_{\mathrm{XSY}_1} \in \mathcal{D}_{\eta_1}(\Lambda)$; (b) $P_{\mathrm{X'S'Y}_1} \in \mathcal{D}_{\eta_1}(\Lambda)$ for some S'; and (c) $I(\mathrm{XY}_1; \mathrm{X}'|\mathrm{S}) > \eta_1$. Let $\mathcal{E}_{\eta_1}(\Lambda)$ denote the set of all types $P_{\mathrm{XX'SY}_1}$ which satisfy the aforementioned conditions (a)–(c). Consequently, the probability of error for message m and state sequence $s^n \in \mathcal{S}_{\Lambda}^n$ is bounded by

$$e_1(m, s^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \leq \sum_{y_1^n : P_{x_m^n, s^n, y_1^n} \notin \mathcal{D}_{\eta_1}(\Lambda)} W_1^{\otimes n}(y_1^n|x_m^n, s^n) + \sum_{P_{\mathrm{XX'SY}_1} \in \mathcal{E}_{\eta_1}(\Lambda)} e_{\mathrm{XX'SY}_1}(m, s^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \tag{23}$$

where

$$e_{\mathrm{XX'SY}_1}(m, s^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) := \sum_{\substack{y_1^n : (x_{m_1,m_2}^n, x_{m_1,m_2'}^n, s^n, y_1^n) \in \mathcal{T}_{\mathrm{XX'SY}_1}^n \\ \text{for some } m_2' \neq m_2}} W_1^{\otimes n}(y_1^n|x_{m_1,m_2}^n, s^n) \tag{24}$$

Next, for given $m_1 \in \mathcal{M}_1$ we define the set

$$\mathcal{A}_{m_1} := \left\{ m_2 : (x_{m_1,m_2}^n, s^n) \in \bigcup_{I(\mathrm{X};\mathrm{S}) > \epsilon} \mathcal{T}_{\mathrm{XS}}^n \right\}$$

and use the trivial bound $e_1((m_1, m_2), s^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \leq 1$ for all such $m_2 \in \mathcal{A}_{m_1}$. With this and (23) we get for the average probability of error

$$\bar{e}_1(s^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \leq \frac{|\mathcal{A}_{m_1}|}{|\mathcal{M}|} + \frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} \sum_{m_2 \in \mathcal{A}_{m_1}^c} \sum_{y_1^n : P_{x_m^n, s^n, y_1^n} \notin \mathcal{D}_{\eta_1}(\Lambda)} W_1^{\otimes n}(y_1^n|x_m^n, s^n)$$

$$+ \frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} \sum_{m_2 \in \mathcal{A}_{m_1}^c} \sum_{P_{\mathrm{XX'SY}_1} \in \mathcal{E}_{\eta_1}(\Lambda)} e_{\mathrm{XX'SY}_1}(m, s^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))$$

Property (18b) of the codewords and Fact 1 from Section 2 imply for the first term that

$$\frac{|\mathcal{A}_{m_1}|}{|\mathcal{M}|} \leq (n+1)^{|\mathcal{X}||\mathcal{S}|} \exp\big(-n\frac{\epsilon}{2}\big) \leq \exp\big(-n\frac{\epsilon}{3}\big) \tag{25}$$

where the last inequality holds for sufficiently large n.

To bound the second term we observe that for any $m_2 \in \mathcal{A}_{m_1}^c$

$$\begin{aligned}
\sum_{y_1^n : P_{x_m^n, s^n, y_1^n} \notin \mathcal{D}_{\eta_1}(\Lambda)} W_1^{\otimes n}(y_1^n | x_m^n, s^n) &\leq \sum_{P_{\mathrm{XSY}_1} \notin \mathcal{D}_{\eta_1}(\Lambda)} W_1^{\otimes n}(\mathcal{T}_{\mathrm{Y}_1|\mathrm{XS}}^n(x_m^n, s^n)|x_m^n, s^n) \\
&\leq \sum_{P_{\mathrm{XSY}_1} \notin \mathcal{D}_{\eta_1}(\Lambda)} \exp\big(-nD(P_{\mathrm{XSY}_1}\|P_{\mathrm{XS}} \otimes W_1)\big) \\
&\leq (n+1)^{|\mathcal{X}||\mathcal{S}||\mathcal{Y}_1|} \exp\big(-n(\eta_1 - \epsilon)\big) \\
&\leq \exp\big(-n(\eta_1 - 2\epsilon)\big)
\end{aligned} \tag{26}$$

where the second inequality follows from Fact 3 and the third inequality from Fact 1 and

$$\begin{aligned}
D(P_{\mathrm{XSY}_1}\|P_{\mathrm{XS}} \otimes W_1) &= D(P_{\mathrm{XSY}_1}\|P_{\mathrm{X}} \otimes P_{\mathrm{S}} \otimes W_1) - I(\mathrm{X}; \mathrm{S}) \\
&> \eta_1 - \epsilon.
\end{aligned}$$

It remains to bound for $P_{\mathrm{XX'SY}_1} \in \mathcal{E}_{\eta_1}(\Lambda)$ the term

$$\frac{1}{|\mathcal{M}|} \sum_{m_1 \in \mathcal{M}_1} \sum_{m_2 \in \mathcal{A}_{m_1}^c} e_{\mathrm{XX'SY}_1}(m, s^n | \mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \tag{27}$$

Before we proceed to bound (27) we observe that if $I(\mathrm{X}; \mathrm{X'S}) > |R_1 - I(\mathrm{X'}; \mathrm{S})|^+ + \epsilon$, then by (18c),

$$\frac{1}{M_{2,n}}\big|\{m_2 : (x_{m_1,m_2}^n, x_{m_1,m_2'}^n, s^n) \in \mathcal{T}_{\mathrm{XX'S}}^n \text{ for some } m_2' \neq m_2\}\big| \leq \exp\big(-n\frac{\epsilon}{2}\big) \tag{28}$$

Consequently, it suffices to proceed when $P_{\mathrm{XX'SY}_1} \in \mathcal{E}_{\eta_1}(\Lambda)$ satisfies

$$I(\mathrm{X}; \mathrm{X'S}) \leq |R_1 - I(\mathrm{X'}; \mathrm{S})|^+ + \epsilon$$

From (24) we may write

$$\begin{aligned}
&e_{\mathrm{XX'SY}_1}(m, s^n | \mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \\
&\qquad\leq \sum_{m_2' : (x_{m_1,m_2}^n, x_{m_1,m_2'}^n, s^n) \in \mathcal{T}_{\mathrm{XX'S}}^n} W_1^{\otimes n}(\mathcal{T}_{\mathrm{Y}_1|\mathrm{XX'S}}^n(x_{m_1,m_2}^n, x_{m_1,m_2'}^n, s^n)|x_{m_1,m_2}^n, s^n)
\end{aligned} \tag{29}$$

Since $W_1^{\otimes n}(y_1^n | x_{m_1,m_2}^n, s^n)$ is constant for $y_1^n \in \mathcal{T}_{\mathrm{Y}_1|\mathrm{XS}}^n(x_{m_1,m_2}^n, s^n)$ and $W_1^{\otimes n}(y_1^n | x_{m_1,m_2}^n, s^n) \leq 1/(|\mathcal{T}_{\mathrm{Y}_1|\mathrm{XS}}^n(x_{m_1,m_2}^n, s^n)|)$ the inner term in (29) is bounded from above by

$$\begin{aligned}
W_1^{\otimes n}(\mathcal{T}_{\mathrm{Y}_1|\mathrm{XX'S}}^n(x_{m_1,m_2}^n, x_{m_1,m_2'}^n, s^n)|x_{m_1,m_2}^n, s^n) &\leq \frac{|\mathcal{T}_{\mathrm{Y}_1|\mathrm{XX'S}}^n(x_{m_1,m_2}^n, x_{m_1,m_2'}^n, s^n)|}{|\mathcal{T}_{\mathrm{Y}_1|\mathrm{XS}}^n(x_{m_1,m_2}^n, s^n)|} \\
&\leq \exp\big(-n(I(\mathrm{Y}_1; \mathrm{X'}|\mathrm{XS}) - \epsilon)\big)
\end{aligned} \tag{30}$$

where the last inequality follows immediately from Fact 2. Next, using (18a), it follows from (29) together with (30) that

$$e_{\mathrm{XX'SY_1}}(m, s^n | \mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \leq \exp\big(-n(I(\mathrm{Y_1}; \mathrm{X'}|\mathrm{XS}) - |R_1 - I(\mathrm{X'}; \mathrm{XS})|^+ - 2\epsilon)\big) \qquad (31)$$

Since

$$|R_1 - I(\mathrm{X'}; \mathrm{XS})|^+ \geq R_1 - I(\mathrm{X'}; \mathrm{XS}) - \epsilon$$

is obviously fulfilled, we can substitute this into (31) and obtain

$$e_{\mathrm{XX'SY_1}}(m, s^n | \mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \leq \exp\big(-n(I(\mathrm{X'}; \mathrm{XSY_1}) - R_1 - \epsilon)\big)$$
$$\leq \exp\big(-n(I(\mathrm{X'}; \mathrm{Y_1}) - R_1 - \epsilon)\big)$$

Since $P_{\mathrm{X'S'Y_1}} \in \mathcal{D}_{\eta_1}(\Lambda)$ for some S', it follows from (21) and (22) that

$$I(\mathrm{X'}; \mathrm{Y_1}) - R_1 \geq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) - R_1 - \frac{\delta}{3} > \frac{\delta}{3}$$

and therewith

$$e_{\mathrm{XX'SY_1}}(m, s^n | \mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n)) \leq \exp\big(-n(\frac{\delta}{3} - \epsilon)\big) \qquad (32)$$

Now, we choose $\epsilon < \min\{\frac{\delta}{3}, \frac{\eta_1}{2}\}$ so that (25), (26), (28), and (32) imply that the average probability of error decreases exponentially fast for sufficiently large n. Since the derived bounds hold uniformly for all $s^n \in \mathcal{S}_\Lambda^n$, the first part of the proof is complete. Similarly, we can now bound the average probability of error at node 2 using the same argumentation. $\qquad \square$

4.3.3. Converse

It remains to show that there are no other rate pairs achievable than these rate pairs which are already characterized by Theorem 6. If $\Lambda_i(P_{\mathrm{X}}) < \Lambda$, $i = 1, 2$, the converse is already established by Lemma 1. Consequently, we only need to consider the case where $\Lambda_i(P_{\mathrm{X}}) > \Lambda$, $i = 1, 2$, in the following.

Lemma 5. *For any $\Lambda > 0$, $\delta > 0$, and $\epsilon < 1$, there exists n_0 such that for any deterministic code $\mathcal{C}_{det}(\mathfrak{W}^n)$ of block length $n \geq n_0$ with $M_{1,n} M_{2,n}$ codewords, each of type P_{X}, satisfying*

$$\frac{1}{n} \log M_{2,n} \geq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) + \delta$$

implies

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_1(s^n | \mathcal{C}_{det}(\mathfrak{W}^n)) > \epsilon$$

And similarly, if the codewords satisfy $\frac{1}{n} \log M_{1,n} \geq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_{\mathrm{X}}, \overline{W}_{2,q}) + \delta$, then $\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_2(s^n | \mathcal{C}_{det}(\mathfrak{W}^n)) > \epsilon$.

Proof. The proof follows [4] (Lemma 2) where a similar converse result is shown for the single-user case. We carry out the analysis for receiving node 1, then the result for receiving node 2 follows accordingly using the same argumentation.

Let us consider a joint probability distribution

$$P_{\mathrm{XSY_1}}(x, s, y_1) = P_{\mathrm{X}}(x) q(s) W_1(y_1 | x, s) \qquad (33)$$

If some probability distribution $q \in \mathcal{P}(\mathcal{S}, \Lambda)$ satisfies

$$\mathbb{E}_q[l(q)] \leq \Lambda(1-\eta) \tag{34}$$

for some $\eta > 0$ which depends on δ but not on P_X, then

$$I(X; Y_1) \leq \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_X, \overline{W}_{1,q}) + \frac{\delta}{2} \tag{35}$$

To prove (35) let $q^* \in \mathcal{P}(\mathcal{S}, \Lambda)$ be a probability distribution which achieves the infimum in $\inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_X, \overline{W}_{1,q})$ so that we have $I(X; Y_1^*) = \inf_{q \in \mathcal{P}(\mathcal{S}, \Lambda)} I(P_X, \overline{W}_{1,q})$ for $P_{XS^*Y_1^*}$ as given in (33) with $\mathbb{E}_{q^*}[l(q^*)] \leq \Lambda$. Next, we use q^* to construct a new probability distribution with slightly smaller costs than Λ as required in (34). Therefore, let $s_0 \in \mathcal{S}$ with $l(s_0) = 0$ and define

$$q(s) := \begin{cases} (1-\eta)q^*(s) & \text{if } s \neq s_0 \\ \eta + (1-\eta)q^*(s) & \text{if } s = s_0 \end{cases}$$

Clearly, $q(s)$ satisfies (34), and therefore (35) holds for sufficiently small η, since $I(X; Y_1)$ is a uniformly continuous in (P_X, q) if P_{XSY_1} is given as in (33).

Similarly as in [4] (Lemma 2), we consider now any deterministic code $\mathcal{C}_{\det}(\mathfrak{W}^n)$ with codewords x_{m_1,m_2}^n, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$, and decoding sets $\mathcal{D}_{m_2|m_1}^{(1)}$ and $\mathcal{D}_{m_1|m_2}^{(2)}$ for all $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$, *cf.* Definition 9. Further, let $S^n = (S_1, ..., S_n) \in \mathcal{S}^n$ be a sequence, where each element is independent and identically distributed according to q as constructed above. Then for receiving node 1 we get for each fixed $m_1 \in \mathcal{M}_1$ for the probability of error

$$\begin{aligned} \mathbb{E}_q\big[\bar{e}_1(S^n | \mathcal{C}_{\det}(\mathfrak{W}^n))\big] &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathbb{E}_q\big[e_1((m_1, m_2), S^n | \mathcal{C}_{\det}(\mathfrak{W}^n))\big] \\ &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{y_1^n \notin \mathcal{D}_{m_2|m_1}^{(1)}} \mathbb{E}_q\big[W_1^n(y_1^n | x_{m_1,m_2}^n, S^n)\big] \\ &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{y_1^n \notin \mathcal{D}_{m_2|m_1}^{(1)}} \prod_{k=1}^n \mathbb{E}_q\big[W_1(y_{1,k} | x_{m_1,m_2,k}, S_k)\big] \end{aligned} \tag{36}$$

Next, we set

$$\overline{W}_{1,q}(y_1 | x) = \mathbb{E}_q[W_1(y_1 | x, s)] \tag{37}$$

which is, in fact, a discrete memoryless channel (DMC). For each $m_1 \in \mathcal{M}_1$, (36) yields that $\mathbb{E}_q[\bar{e}_1(S^n | \mathcal{C}_{\det}(\mathfrak{W}^n))] = \bar{e}_1(\overline{W}_{1,q} | \mathcal{C}_{\det}(\mathfrak{W}^n))$ where $\bar{e}_1(\overline{W}_{1,q} | \mathcal{C}_{\det}(\mathfrak{W}^n))$ is the average probability of error when the deterministic code $\mathcal{C}_{\det}(\mathfrak{W}^n)$ is used on the DMC $\overline{W}_{1,q}$. Next, observe that

$$\begin{aligned} \mathbb{P}\{l(S^n) > \Lambda\} &= \mathbb{P}\left\{\frac{1}{n} \sum_{k=1}^n l(S_k) > \mathbb{E}_q[l(q)] + \eta\Lambda\right\} \\ &\leq \frac{\big(\text{var}[l(q)]\big)^2}{n(\eta\Lambda)^2} \\ &\leq \frac{l_{\max}^2}{n\eta^2\Lambda^2} \end{aligned}$$

which follows from (34), (5b), and Chebyshev's inequality so that we get

$$\max_{s^n : l(s^n) \leq \Lambda} \bar{e}_1(s^n | \mathcal{C}_{\det}(\mathfrak{W}^n)) \geq \mathbb{E}_q[\bar{e}_1(S^n)] - \mathbb{P}\{l(S^n) > \Lambda\}$$

$$\geq \bar{e}_1(\overline{W}_{1,q} | \mathcal{C}_{\det}(\mathfrak{W}^n)) - \frac{l_{\max}^2}{n\eta^2\Lambda^2} \tag{38}$$

Now, we are almost done. We observe that the definition of $P_{\mathrm{XSY_1}}$ as given in (35) implies that Y_1 is connected with X by the channel $\overline{W}_{1,q}$ as defined in (37). For such a DMC a strong converse in terms of maximal error can be found in [3], which immediately yields also a strong converse for the DMC in terms of average probability of error as needed here. In more detail, (36) implies, by the strong converse for a DMC with codewords of type P_{X}, that if all codewords x_{m_1,m_2}^n, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$, each of type P_{X}, then, for each $m_1 \in \mathcal{M}_1$, the average probability of error $\bar{e}_1(\overline{W}_{1,q} | \mathcal{C}_{\det}(\mathfrak{W}^n))$ is arbitrarily close to 1 if $\frac{1}{n} \log M_{2,n} \geq \inf_{q \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_{\mathrm{X}}, \overline{W}_{1,q}) + \delta$ and n sufficiently large enough. Finally, this together with (38) complete the first part of the proof.

The result for receiving node 2 follows accordingly using the same argumentation which completes the proof of the lemma. □

4.3.4. Capacity Region

Now we are in the position to finally establish the deterministic code capacity region, which is one of the main contributions of this work. Thus, summarizing the results obtained so far, we see that for given input distribution P_{X} the achievable rates for the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ are given by $\mathcal{R}(P_{\mathrm{X}}|\Lambda)$ if $\Lambda_i(P_{\mathrm{X}}) > \Lambda$, $i = 1, 2$. Taking the union over all such valid inputs we finally obtain

$$\mathcal{R}_{\det}(\mathfrak{W}^n | \Gamma, \Lambda) = \bigcup_{\substack{P_{\mathrm{X}} : g(P_{\mathrm{X}}) \leq \Gamma, \\ \Lambda_i(P_{\mathrm{X}}) > \Lambda, i=1,2}} \mathcal{R}(P_{\mathrm{X}}|\Lambda)$$

On the other hand, we have $\mathrm{int}(\mathcal{R}_{\det}(\mathfrak{W}^n | \Gamma, \Lambda)) = \emptyset$ if $\max_{P_{\mathrm{X}} : g(P_{\mathrm{X}}) \leq \Gamma} \Lambda_1(P_{\mathrm{X}}) < \Lambda$ or $\max_{P_{\mathrm{X}} : g(P_{\mathrm{X}}) \leq \Gamma} \Lambda_2(P_{\mathrm{X}}) < \Lambda$, which follows immediately from Lemma 1. This, indeed, establishes the deterministic code capacity region $\mathcal{R}_{\det}(\mathfrak{W}^n | \Gamma, \Lambda)$ of the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ as stated in Theorem 6.

Remark 5. *The case where $\Lambda_i(P_{\mathrm{X}}) = \Lambda$, $i \in \{1, 2\}$, remains unsolved in a similar way as for the single-user AVC [4]. Likewise, we expect that $\mathrm{int}(\mathcal{R}(P_{\mathrm{X}}|\Lambda)) = \emptyset$ in that case.*

4.4. Unknown Varying Additive Interference

So far we considered discrete memoryless channels and analyzed the corresponding arbitrarily varying bidirectional broadcast channel. Here, we assume channels with additive white Gaussian noise, where the transmission in the bidirectional broadcast phase is further corrupted by unknown varying additive interference. Therefore, we also call this a *BBC with unknown varying interference*. Clearly, the interference at both receivers may differ so that we introduce two artificial interferers or *jammers*, one for each receiver, to model this scenario. Then the BBC with unknown varying interference is specified by the flat fading input-output relation between the relay node and node i, $i = 1, 2$, which is given by

$$y_i = x + j_i + n_i$$

Here, $y_i \in \mathbb{R}$ denotes the output at node i, $x \in \mathbb{R}$ the input, $j_i \in \mathbb{R}$ the additive interference, and $n_i \in \mathbb{R}$ the additive Gaussian noise distributed according to $\mathcal{N}(0, \sigma^2)$.

The transmit powers of the relay and of the artificial jammers are restricted by average power constraints Γ and Λ_i, $i = 1, 2$, respectively. This means, all permissible input sequences $x^n = (x_1, x_2, ..., x_n)$ of length n must satisfy

$$\frac{1}{n} \sum_{k=1}^{n} x_k^2 \leq \Gamma \tag{39}$$

and all permissible jamming sequences $j_i^n = (j_{i,1}, j_{i,2}, ..., j_{i,n})$, $i = 1, 2$, of length n must satisfy

$$\frac{1}{n} \sum_{k=1}^{n} j_{i,k}^2 \leq \Lambda_i \tag{40}$$

From conditions (39) and (40) it follows that all permissible codewords and interfering sequences lie on or within an n-dimensional sphere of radius $\sqrt{n\Gamma}$ or $\sqrt{n\Lambda_i}$, $i = 1, 2$, respectively.

Similarly as for the discrete memoryless AVBBC, it makes a difference for the BBC with unknown varying interference, if we consider deterministic or random coding strategies. Hence, we want specify their different impact on the transmission in the following.

4.4.1. No Additional Coordination

The traditional approach without additional coordination is in general based on a system design which ensures that the interference at the receivers does not exceed a certain threshold. For example in current cellular networks, this is realized by separating cells in space which operate at the same frequency.

Theorem 7. *The deterministic code capacity region $\mathcal{R}_{det}(\mathfrak{W}^n)$ of the BBC with unknown varying interference with input constraint Γ and jamming constraints Λ_1 and Λ_2 is the set of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy*

$$R_i \leq \begin{cases} \frac{1}{2} \log \left(1 + \frac{\Gamma}{\Lambda_i + \sigma^2} \right) & \text{if } \Gamma > \Lambda_i \\ 0 & \text{if } \Gamma \leq \Lambda_i \end{cases} \tag{41}$$

$i = 1, 2$. This means $int(\mathcal{R}_{det}(\mathfrak{W}^n)) \neq \emptyset$ if and only if $\Gamma > \Lambda_1$ and $\Gamma > \Lambda_2$.

Sketch of Proof. First, we consider the case when $\Gamma \leq \Lambda_1$ or $\Gamma \leq \Lambda_2$. Let $x_{m_1,m_2}^n \in \mathbb{R}^n$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$ with $M_{1,n} \geq 2$ and $M_{2,n} \geq 2$ be arbitrary codewords satisfying the input constraint (39). For $\Gamma \leq \Lambda_1$ we can consider the jamming sequences $j_{1,m_1,m_2}^n = x_{m_1,m_2}^n$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$. Then for each $m_1 \in \mathcal{M}_1$ at node 1 the following holds. For each pair $(k, l) \in \mathcal{M}_2 \times \mathcal{M}_2$ with $k \neq l$ we have for the probability of error at node 1

$$\mathbb{E}\left[e_1((m_1, k), j_l^n | \mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \right] + \mathbb{E}\left[e_1((m_1, l), j_k^n | \mathcal{C}_{\text{det}}(\mathfrak{W}^n)) \right]$$
$$= \mathbb{P}\{ x_{m_1,k}^n + j_{1,m_1,l}^n + n_1^n \notin \mathcal{D}_{k|m_1}^{(1)} \} + \mathbb{P}\{ x_{m_1,l}^n + j_{1,m_1,k}^n + n_1^n \notin \mathcal{D}_{l|m_1}^{(1)} \}$$
$$= \mathbb{P}\{ x_{m_1,k}^n + j_{1,m_1,l}^n + n_1^n \in (\mathcal{D}_{k|m_1}^{(1)})^c \} + \mathbb{P}\{ x_{m_1,k}^n + j_{1,m_1,l}^n + n_1^n \notin \mathcal{D}_{l|m_1}^{(1)} \}$$
$$\geq \mathbb{P}\{ x_{m_1,k}^n + j_{1,m_1,l}^n + n_1^n \in (\mathcal{D}_{k|m_1}^{(1)})^c \} + \mathbb{P}\{ x_{m_1,k}^n + j_{1,m_1,l}^n + n_1^n \in \mathcal{D}_{k|m_1}^{(1)} \}$$
$$= \mathbb{P}\{ x_{m_1,k}^n + j_{1,m_1,l}^n + n_1^n \in (\mathcal{D}_{k|m_1}^{(1)})^c \cup \mathcal{D}_{k|m_1}^{(1)} \} = 1$$

Hence, for a fixed $m_1 \in \mathcal{M}_1$ this leads for the average probability of error to

$$\frac{1}{M_{2,n}} \sum_{k=1}^{M_{2,n}} \mathbb{E}\big[\bar{e}_1(j_{1,m_1,k}^n | \mathcal{C}_{\text{det}}(\mathfrak{W}^n))\big]$$

$$= \frac{1}{M_{2,n}} \frac{1}{M_{1,n}M_{2,n}} \sum_{k=1}^{M_{2,n}} \sum_{m_1'=1}^{M_{1,n}} \sum_{m_2'=1}^{M_{2,n}} \mathbb{E}\big[e_1((m_1', m_2'), j_{1,m_1,k}^n | \mathcal{C}_{\text{det}}(\mathfrak{W}^n))\big]$$

$$\geq \frac{1}{M_{1,n}(M_{2,n})^2} \sum_{m_1'}^{M_{1,n}} \frac{M_{2,n}(M_{2,n}-1)}{2}$$

$$= \frac{M_{1,n}M_{2,n}(M_{2,n}-1)}{2M_{1,n}(M_{2,n})^2} = \frac{M_{2,n}-1}{2M_{2,n}} \geq \frac{1}{4}$$

This implies that $\mathbb{E}[\bar{e}_1(j_{1,m_1,m_2}^n | \mathcal{C}_{\text{det}}(\mathfrak{W}^n))] \geq \frac{1}{4}$ for at least one $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$. Since the average probability of error is bounded from below by a positive constant, a reliable transmission from the relay to node 1 is not possible so that we end up with $R_1 = 0$. The case $\Gamma \leq \Lambda_2$ similarly leads to $R_2 = 0$.

Remark 6. *Interestingly, Theorem 7 shows that the existence of positive rates only depends on the interference and is completely independent of the noise. Consequently, the goal of the traditional approach is to ensure that the received interference will be small enough. Otherwise, there is no communication possible, not even at very low rates.*

Now, we turn to the case when $\Gamma > \Lambda_1$ and $\Gamma > \Lambda_2$. To show that the rates given in (41) are actually achievable, we follow [58] where a similar result is proved for the corresponding single-user scenario. The strategy is outlined in the following.

Without loss of generality we assume that $\Gamma = 1$ and further $0 < \Lambda_i < 1$, $i = 1, 2$. Then it suffices to show that for every small $\delta > 0$ and sufficiently large n there exist $M_{1,n}M_{2,n}$ codewords x_{m_1,m_2}^n (on the unit sphere) with $M_{1,n} = \exp(nR_2)$ and $M_{2,n} = \exp(nR_1)$ and $C_i - 2\delta < R_i < C_i - \delta$ with $C_i := \frac{1}{2}\log(1 + \frac{1}{\Lambda_i + \sigma^2})$, $i = 1, 2$, *cf.* (41), such that the average probability is arbitrarily small for all j_i^n satisfying (40). To ensure that the probability of error gets arbitrarily small, the codewords must possess certain properties which are guaranteed by the following lemma. This is a straightforward extension of the single-user case [58] (Lemma 1) to the BBC with unknown varying interference.

Lemma 6. *For every $\epsilon > 0$, $8\sqrt{\epsilon} < \eta < 1$, $K > 2\epsilon$, and $M_{1,n} = \exp(nR_2)$, $M_{2,n} = \exp(nR_1)$ with $2\epsilon \leq R_i \leq K$, $i = 1, 2$, for $n \geq n_0(\epsilon, \eta, K)$ there exist unit vectors x_{m_1,m_2}^n, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$ such that for every unit vector u^n and constants α, β in $[0, 1]$, we have for each $m_1 \in \mathcal{M}_1$*

$$\left|\left\{m_2 : \langle x_{m_1,m_2}^n, u^n \rangle \geq \alpha\right\}\right| \leq \exp\left(n(|R_1 + \tfrac{1}{2}\log(1 - \alpha^2)|^+ + \epsilon)\right)$$

and, if $\alpha \geq \eta$, $\alpha^2 + \beta^2 > 1 + \eta - \exp(-2R_1)$

$$\frac{1}{M_{2,n}}\left|\left\{m_2' : |\langle x_{m_1,m_2}^n, x_{m_1,m_2'}^n \rangle| \geq \alpha, |\langle x_{m_1,m_2}^n, u^n \rangle| \geq \beta, \text{ for some } m_2' \neq m_2\right\}\right| \leq \exp(-n\epsilon)$$

and similarly for each $m_2 \in \mathcal{M}_2$

$$\left|\left\{m_1 : \langle x_{m_1,m_2}^n, u^n \rangle \geq \alpha\right\}\right| \leq \exp\left(n(|R_2 + \tfrac{1}{2}\log(1 - \alpha^2)|^+ + \epsilon)\right)$$

and, if $\alpha \geq \eta$, $\alpha^2 + \beta^2 > 1 + \eta - \exp(-2R_2)$

$$\frac{1}{M_{1,n}} \left| \left\{ m_1' : | \left\langle x_{m_1,m_2}^n, x_{m_1',m_2}^n \right\rangle | \geq \alpha, | \left\langle x_{m_1,m_2}^n, u^n \right\rangle | \geq \beta, \text{ for some } m_1' \neq m_1 \right\} \right| \leq \exp(-n\epsilon)$$

Proof. The proof is a straightforward extension of the corresponding single-user result given in [58] (Lemma 1) and is therefore omitted for brevity. □

At the receiving nodes it suffices to use a minimum-distance decoder. Then for each $m_1 \in \mathcal{M}_1$ the decoding sets at node 1 and for each $m_2 \in \mathcal{M}_2$ at node 2 are given by

$$\mathcal{D}_{m_2|m_1}^{(1)} := \left\{ y_1^n : \|y_1^n - x_{m_1,m_2}^n\|^2 < \|y_1^n - x_{m_1,m_2'}^n\|^2 \text{ for all } m_2' \neq m_2 \right\} \tag{42a}$$

$$\mathcal{D}_{m_1|m_2}^{(2)} := \left\{ y_2^n : \|y_2^n - x_{m_1,m_2}^n\|^2 < \|y_2^n - x_{m_1',m_2}^n\|^2 \text{ for all } m_1' \neq m_1 \right\} \tag{42b}$$

With the presented coding and decoding rule, the probability of error gets arbitrarily small for increasing block length, which can be shown analogously to [58]. The details are omitted for brevity.

It remains to show that the described strategy is optimal, which means that no other rate pairs are achievable. From the previous discussions, we already know that the capacity region of the deterministic code capacity region is included in the capacity region of the random code capacity region. In the next subsection, from Theorem 8 we see that for $\Gamma > \Lambda_i$, $i = 1, 2$, the maximal achievable rates for both strategies are equal. Since the described strategy already achieves these rates, the optimality is proved.

4.4.2. Encoder-Decoder Coordination Based on Common Randomness

Next, we study a more involved coordination scheme. We assume that the relay and the receivers are synchronized in such a manner that they can coordinate their choice of the encoder and decoders based on an access to a common resource independent of the current message.

This can be realized by using a random code. If we transmit at rates R_1 and R_2 with exponentially many messages, *i.e.*, $\exp(nR_1)$ and $\exp(nR_2)$, we know from [2] that it suffices to use a random code, which consists of n^2 pairs of encoder and decoders and a uniformly distributed random variable whose value indicates which of the pair all nodes have to use. The access to the common random variable can be realized by an external source, e.g., a satellite signal, or a preamble prior to the transmission. Clearly, for sufficiently large block length the (polynomial) costs for the coordination are negligible. We call this *additional encoder-decoder coordination based on common randomness*. Due to the more involved coordination we expect an improvement in the performance compared to the traditional approach, especially for high interference.

Theorem 8. *The random code capacity region* $\mathcal{R}_{ran}(\mathfrak{W}^n)$ *of the BBC with unknown varying interference with input constraint* Γ *and jamming constraints* Λ_1 *and* Λ_2 *is the set of all rate pairs* $(R_1, R_2) \in \mathbb{R}_+^2$ *that satisfy*

$$R_i \leq \frac{1}{2} \log \left(1 + \frac{\Gamma}{\Lambda_i + \sigma^2} \right), \quad i = 1, 2 \tag{43}$$

Sketch of Proof. The theorem can be proved analogously to [59] where a similar result is proved for the single-user case. The random strategy which achieves the rates given in (43) is outlined in the following.

The codewords $x^n_{m_1,m_2}$ are uniformly distributed on the n-sphere of radius $\sqrt{n\Gamma}$. Similar to the traditional approach, a minimum-distance decoder as given in (42) at the receiving nodes is sufficient. It remains to show that for all rate pairs satisfying (43) the probability of error gets arbitrarily small for increasing block length. This can be done similarly to [59].

The optimality of the presented random strategy, which means that no other rate pairs are achievable, follows immediately from [59] and can be shown by standard arguments.

Remark 7. *The capacity region $\mathcal{R}_{ran}(\mathfrak{W}^n)$ is identical to the one if the interfering sequences would consist of iid Gaussian symbols distributed according to $\mathcal{N}(0, \Lambda_i)$, $i = 1, 2$. This means, the arbitrary, possibly non-Gaussian, unknown interference do not affect the achievable rates more than Gaussian noise of the same power.*

5. Discussion

The concept of *arbitrarily varying channels* has been shown to be a suitable and robust model for communication in wireless networks, which share their resources with other coexisting systems in an uncoordinated way. The main issue that comes along with this development is that interference becomes an ubiquitous phenomenon and that it will be one of the limiting factors in future wireless networks.

It has been shown that unknown varying interference has a dramatic impact on the communication in such wireless systems. If the traditional approach without additional coordination is applied, unknown varying interference can lead to situations that completely prohibit any reliable communication. This is mainly based on the assumption that the traditional approach treats the interference as some kind of additional noise. As we have seen, this is in general to imprecise and leads to a performance loss especially if the interference is caused by other transmitters that use the same or a similar codebook. Then, interference can look like other valid codewords and receivers cannot reliably distinguish between the intended signal and the interference anymore. Consequently, a traditional approach based on a deterministic coding strategy is only reasonable if the interference can be made small enough. For Gaussian channels this means that the power of the interference signal must be ensured to be smaller than the power of the transmit signal. Thus, especially in the high interference case where the interference power exceeds the transmit power, a more sophisticated coordination based on a random coding strategy is needed for reliable communication. It is shown that an additional coordination of the encoder and decoder based on a common resource, such as common randomness or correlated side information, is sufficient to handle the interference even if it is stronger than the desired signal.

To date only the single-user AVC is analyzed under additional encoder-decoder coordination based on correlated side information in [54]. It would be interesting to extend it also to other (multi-user) settings.

In this paper we used the concept of arbitrarily varying channels to analyze bidirectional relaying in coexistence with other wireless networks. This required the study of the *arbitrarily varying bidirectional broadcast channel (AVBBC)*. Based on Ahlswede's elimination technique [2] the following dichotomy of the deterministic code capacity region of an AVBBC was revealed in [51,52]: it either equals its random code capacity region or else has an empty interior. Unfortunately, many channels of practical interest are symmetrizable, which results in an ambiguity of the codewords at the receivers. Such channels prohibit any reliable communication and therewith fall in the latter category.

Imposing constraints on the permissible sequences of channel states reveals further phenomena. Now, even when the channel is symmetrizable, the deterministic code capacity region of the AVBBC under input and state constraints may be non-empty but less than its random code capacity region. Thereby, we observed that the constraints on the state sequences may reduce the deterministic code capacity region so that it is in general strictly smaller than the corresponding random code capacity region, but they preserve the general dichotomy behavior of the deterministic code capacity region: it still either equals a non-empty region or else has an empty interior. Although the deterministic code capacity region displays a dichotomy behavior, it cannot be exploited to prove the corresponding capacity region since Ahlswede's elimination technique [2] does not work anymore in the presence of constraints on input and states, *cf.* also [60]. This necessitated a proof technique which does not rely on the dichotomy behavior and is based on an idea of Csiszár and Narayan [4].

Besides the concept of arbitrarily varying channels, there are also other approaches to tackle the problem of interference or channel uncertainty in wireless networks. One approach to model the interference is based on the framework of interference functions, *cf.* for example [61] or [62,63]. In this axiomatic approach the interference functions are assumed to have some basic properties such as non-negativity, scale-invariance, and monotonicity. It is shown that under these assumptions the performance of wireless systems depends continuously on the interference functions. These assumptions are valid and reasonable for conventional cellular systems which are coordinated in a centralized way. But if such systems compete with other coexisting systems on the same wireless resources, the concept of arbitrarily varying channels show that these assumptions are no longer valid.

In the signal processing community, a common approach to tackle the problem of channel uncertainty is the robust design of wireless systems based on robust optimization techniques. There are statistical approaches which assume the channel to be random but according to a certain statistic that is known. For example heuristics are developed for the multi-antenna downlink scenario from a signal processing point of view in [64,65]. These approaches are developed for conventional cellular systems, and it would be interesting for future work to analyze if these approaches can be extended to the case with unknown interference from other coexisting wireless networks.

Another approach is based on the worst noise analysis as studied in [66–69]. Here, the impact of interference and channel uncertainty is analyzed for conventional single cell systems and, again, it would be interesting to analyze if this approach can be extended to scenarios with interference from coexisting wireless networks.

Acknowledgments

References

1. Blackwell, D.; Breiman, L.; Thomasian, A.J. The capacities of certain channel classes under random coding. *Ann. Math. Stat.* **1960**, *31*, 558–567.

2. Ahlswede, R. Elimination of correlation in random codes for arbitrarily varying channels. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **1978**, *44*, 159–175.

3. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2 ed.; Cambridge University Press: Cambridge, UK, 2011.

4. Csiszár, I.; Narayan, P. The capacity of the arbitrarily varying channel revisited: Positivity, constraints. *IEEE Trans. Inf. Theory* **1988**, *34*, 181–193.

5. Csiszár, I.; Narayan, P. Arbitrarily varying channels with constrained inputs and states. *IEEE Trans. Inf. Theory* **1988**, *34*, 27–34.

6. MolavianJazi, E.; Bloch, M.; Laneman, J.N. Arbitrary jamming can preclude secure communication. In Proceedings of the Allerton Conference Communication, Control, Computing, Urbana-Champaign, IL, USA, 30 September–02 October 2009; pp. 1069–1075.

7. Bjelaković, I.; Boche, H.; Sommerfeld, J. Strong secrecy in arbitrarily varying wiretap channels. In Proceedings of the IEEE Information Theory Workshop, Lausanne, Switzerland, 3–7 September 2012.

8. Jahn, J.H. Coding of arbitrarily varying multiuser channels. *IEEE Trans. Inf. Theory* **1981**, *27*, 212–226.

9. Gubner, J.A. On the deterministic-code capacity of the multiple-access arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1990**, *36*, 262–275.

10. Ahlswede, R.; Cai, N. Arbitrarily varying multiple-access channels part I—Ericson's symmetrizability is adequate, gubner's conjecture is true. *IEEE Trans. Inf. Theory* **1999**, *45*, 742–749.

11. Gubner, J.A. State constraints for the multiple-access arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1991**, *37*, 27–35.

12. 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.

13. Wiese, M.; Boche, H.; Bjelaković, I.; Jungnickel, V. The compound multiple access channel with partially cooperating encoders. *IEEE Trans. Inf. Theory* **2011**, *57*, 3045–3066.

14. Wiese, M.; Boche, H. The arbitrarily varying multiple-access channel with conferencing encoders. In Proceedings of the IEEE International Symposium Information Theory, Saint Petersburg, Russia, 31 July–5 August 2011; pp. 993–997.

15. 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.

16. Rankov, B.; Wittneben, A. Spectral efficient protocols for half-duplex fading relay channels. *IEEE J. Sel. Areas Commun.* **2007**, *25*, 379–389.

17. Larsson, P.; Johansson, N.; Sunell, K.E. Coded bi-directional relaying. In Proceedings of the 5th Scandinavian Workshop on Ad Hoc Networks, Stockholm, Sweden, 3–4 May 2005; pp. 851–855.

18. Wu, Y.; Chou, P.; Kung, S.Y. Information exchange in wireless networks with network coding and physical-layer broadcast. In Proceedings of the Conference Information Sciences and Systems, Baltimore, MD, USA, March 2005; pp. 1–6.

19. Knopp, R. Two-way radio networks with a star topology. In Proceedings of the International Zurich Seminar on Communication, Zurich, Switzerland, February 2006; pp. 154–157.

20. Oechtering, T.J.; Schnurr, C.; Bjelaković, I.; Boche, H. Broadcast capacity region of two-phase bidirectional relaying. *IEEE Trans. Inf. Theory* **2008**, *54*, 454–458.

21. Kim, S.J.; Mitran, P.; Tarokh, V. Performance bounds for bidirectional coded cooperation protocols. *IEEE Trans. Inf. Theory* **2008**, *54*, 5235–5241.

22. Kramer, G.; Shamai (Shitz), S. Capacity for classes of broadcast channels with receiver side information. In Proceedings of the IEEE Information Theory Workshop, Tahoe City, CA, USA, 2–6 September 2007; pp. 313–318.

23. Xie, L.L. Network coding and random binning for multi-user channels. In Proceedings of the Canadian Workshop on Information Theory, 6-8 June 2007; pp. 85–88.

24. Wyrembelski, R.F.; Oechtering, T.J.; Bjelaković, I.; Schnurr, C.; Boche, H. Capacity of Gaussian MIMO bidirectional broadcast channels. In Proceedings of the IEEE International Symposium Information Theory, Toronto, Canada, 6-11 July 2008; pp. 584–588.

25. Oechtering, T.J.; Wyrembelski, R.F.; Boche, H. Multiantenna bidirectional broadcast channels–optimal transmit strategies. *IEEE Trans. Signal Process.* **2009**, *57*, 1948–1958.

26. Oechtering, T.J.; Jorswieck, E.A.; Wyrembelski, R.F.; Boche, H. On the optimal transmit strategy for the MIMO bidirectional broadcast channel. *IEEE Trans. Commun.* **2009**, *57*, 3817–3826.

27. Wyrembelski, R.F.; Bjelaković, I.; Oechtering, T.J.; Boche, H. Optimal coding strategies for bidirectional broadcast channels under channel uncertainty. *IEEE Trans. Commun.* **2010**, *58*, 2984–2994.

28. Wyrembelski, R.F.; Oechtering, T.J.; Boche, H.; Skoglund, M. Robust transmit strategies for multiantenna bidirectional broadcast channels. In Proceedings of the ITG Workshop Smart Antennas, Dresden, Germany, 7–8 March 2012; pp. 46–53.

29. Kim, T.T.; Poor, H.V. Diversity-multiplexing trade-off in adaptive two-way relaying. *IEEE Trans. Inf. Theory* **2011**, *57*, 4235–4254.

30. Zhao, J.; Kuhn, M.; Wittneben, A.; Bauch, G. Optimum time-division in MIMO two-way decode-and-forward relaying systems. In Proceedings of the Asilomar Conf. Signals, Systems, Computers, Pacific Grove, CA, USA, 26–29 November 2008; pp. 1494–1500.

31. Sezgin, A.; Khajehnejad, M.A.; Avestimehr, A.S.; Hassibi, B. Approximate capacity region of the two-pair bidirectional Gaussian relay network. In Proceedings of the IEEE International Symposium Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 2018 – 2022.

32. Popovski, P.; Koike-Akino, T. Coded bidirectional relaying in wireless networks. In *New Directions in Wireless Communications Research*; Springer: Berlin/Heidelberg, Germany, 2009; Chapter 11, pp. 291–316.

33. Zhang, R.; Liang, Y.C.; Chai, C.C.; Cui, S. Optimal beamforming for two-way multi-antenna relay channel with analogue network coding. *IEEE J. Sel. Areas Commun.* **2009**, *27*, 699–712.

34. Ngo, H.Q.; Quek, T.Q.S.; Shin, H. Amplify-and-forward two-way relay networks: Error exponents and resource allocation. *IEEE Trans. Commun.* **2009**, *58*, 2653–2666.

35. Roemer, F.; Haardt, M. Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse. *IEEE Trans. Signal Process.* **2010**, *58*, 5720–5735.

36. Yilmaz, E.; Zakhour, R.; Gesbert, D.; Knopp, R. Multi-pair two-way relay channel with multiple Antenna relay station. In Proceedings of the IEEE International Conference Communication, Cape Town, South Africa, 23–27 May 2010; pp. 1–5.

37. Schnurr, C.; Oechtering, T.J.; Stańczak, S. Achievable rates for the restricted half-duplex two-way relay channel. In Proceedings of the Asilomar Conference Signals, Systems, Computers, Pacific Grove, CA, USA, 4–7 November 2007; pp. 1468–1472.

38. Gündüz, D.; Tuncel, E.; Nayak, J. Rate regions for the separated two-way relay channel. In Proceedings of the Allerton Conference Communication, Control, Computing, Urbana-Champaign, IL, USA, 23–26 September 2008; pp. 1333–1340.

39. Zhong, P.; Vu, M. Compress-forward without Wyner-Ziv binning for the one-way and two-way relay channels. In Proceedings of the Allerton Conference Communication, Control, Computing, Urbana-Champaign, IL, USA, 28–30 September 2011; pp. 426–433.

40. Ong, L.; Kellett, C.M.; Johnson, S.J. Functional-decode-forward for the general discrete memoryless two-way relay channel. In Proceedings of the IEEE International Conference Communication Systems, Singapore, 17–19 November 2010; pp. 351–355.

41. Wilson, M.P.; Narayanan, K.; Pfister, H.D.; Sprintson, A. Joint physical layer coding and network coding for bidirectional relaying. *IEEE Trans. Inf. Theory* **2010**, *56*, 5641–5654.

42. Nam, W.; Chung, S.Y.; Lee, Y.H. Capacity of the gaussian two-way relay channel to within $\frac{1}{2}$ bit. *IEEE Trans. Inf. Theory* **2010**, *56*, 5488–5494.

43. Baik, I.J.; Chung, S.Y. Network coding for two-way relay channels using lattices. *Telecommun. Rev.* **2007**, *17*, 1009–1021.

44. Nazer, B.; Gastpar, M. Compute-and-forward: Harnessing interference through structured codes. *IEEE Trans. Inf. Theory* **2011**, *57*, 6463–6486.

45. Song, Y.; Devroye, N. A lattice compress-and-forward scheme. In Proceedings of the IEEE Infomation Theory Workshop, Paraty, Brasil, 16–20 October 2011; pp. 110–114.

46. Kim, S.J.; Smida, B.; Devroye, N. Lattice strategies for a multi-pair bi-directional relay network. In Proceedings of the IEEE International Symposium Infomation Theory, Saint Petersburg, Russia, 31 July–5 August 2011; pp. 2243–2247.

47. Lim, S.H.; Kim, Y.H.; El Gamal, A.; Chung, S.Y. Layered noisy network coding. In Proceedings of the IEEE Wireless Network Coding Conference, Boston, MA, USA, 21 June 2010; pp. 1–6.

48. Lim, S.H.; Kim, Y.H.; El Gamal, A.; Chung, S.Y. Noisy network coding. *IEEE Trans. Inf. Theory* **2011**, *57*, 3132–3152.

49. Kramer, G.; Hou, J. Short-message quantize-forward network coding. In Proceedings of the 8th International Workshop on Multi-Carrier Systems & Solutions, Herrsching, Germany, 3–4 May 2011; pp. 1–3.

50. Kramer, G.; Hou, J. On message lengths for noisy network coding. In Proceedings of the IEEE Infomation Theory Workshop, Paraty, Brazil, 16–20 October 2011; pp. 430–431.

51. Wyrembelski, R.F.; Bjelaković, I.; Boche, H. Coding strategies for bidirectional relaying for arbitrarily varying channels. In Proceedings of the IEEE Global Communication Conference, Honolulu, HI, USA, 30 November–4 December 2009; pp. 1–6.

52. Wyrembelski, R.F.; Bjelaković, I.; Boche, H. On the capacity of bidirectional relaying with unknown varying channels. In Proceedings of the IEEE Workshop Computational Advances Multi-Sensor Adaptive Processing, Aruba, Dutch Antilles, 13–16 December 2009; pp. 269–272.

53. Wyrembelski, R.F.; Bjelaković, I.; Boche, H. List decoding for bidirectional broadcast channels with unknown varying channels. In Proceedings of the IEEE International Conference Communication, Cape Town, South Africa, 23–27 May 2010; pp. 1–6.

54. Ahlswede, R.; Cai, N. Correlated sources help transmission over an arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1997**, *43*, 1254–1255.

55. Ahlswede, R. Coloring hypergraphs: A new approach to multi-user source coding—II. *J. Comb. Inform. Syst. Sci.* **1980**, *5*, 220–268.

56. Ahlswede, R. Arbitrarily varying channels with states sequence known to the sender. *IEEE Trans. Inf. Theory* **1986**, *32*, 621–629.

57. Ahlswede, R.; Wolfowitz, J. The structure of capacity functions for compound channels. In Proceedings of the International Symposium on Probability and Information Theory, McMaster University, Hamilton, Canada, April 1969; pp. 12–54.

58. Csiszár, I.; Narayan, P. Capacity of the gaussian arbitrarily varying channel. *IEEE Trans. Inf. Theory* **1991**, *37*, 18–26.

59. Hughes, B.; Narayan, P. Gaussian arbitrarily varying channels. *IEEE Trans. Inf. Theory* **1987**, *33*, 267–284.

60. Lapidoth, A.; Narayan, P. Reliable communication under channel uncertainty. *IEEE Trans. Inf. Theory* **1998**, *44*, 2148–2177.

61. Yates, R.D. A framework for uplink power control in cellular radio systems. *IEEE J. Sel. Areas Commun.* **1995**, *13*, 1341–1347.

62. Boche, H.; Schubert, M. A unifying approach to interference modeling for wireless networks. *IEEE Trans. Signal Process.* **2010**, *58*, 3282–3297.

63. Boche, H.; Schubert, M. Concave and convex interference functions—General characterizations and applications. *IEEE Trans. Signal Process.* **2008**, *56*, 4951–4965.

64. Vucic, N.; Boche, H. Robust QoS-constrained optimization of downlink multiuser MISO systems. *IEEE Trans. Signal Process.* **2009**, *57*, 714–725.

65. Vucic, N.; Boche, H.; Shi, S. Robust transceiver optimization in downlink multiuser MIMO systems. *IEEE Trans. Signal Process.* **2009**, *57*, 3576–3587.

66. Jorswieck, E.A.; Boche, H. Majorization and matrix-monotone functions in wireless communications. **2007**, *3*, 553–701.

67. Boche, H.; Jorswieck, E.A. Outage probability of multiple antenna systems: Optimal transmission and impact of correlation. In Proceedings of the International Zurich Seminar on Communication, Zurich, Switzerland, 18–20 February 2004; pp. 116–119.

68. Jorswieck, E.A.; Boche, H. Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information. *IEEE Trans. Signal Process.* **2004**, *52*, 3440–3453.

69. Jorswieck, E.A.; Boche, H. Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *IEEE Trans. Wireless Commun.* **2004**, *3*, 1543–1553.

70. Pinsker, M.S. *Information and Information Stability of Random Variables and Processes*; Holden-Day: San Francisco, CA, USA, 1964.

Appendices

A. Additional Proofs

A.1. Proof of Lemma 1

The lemma follows immediately from [4] (Lemma 1), where a similar result for the single-user AVC is proved. Using the same ideas we are able to extend the proof to the AVBBC $\mathfrak{W}^n$ under input constraint Γ and state constraint Λ. Thereby, we carry out the analysis for the case where $\Lambda_1(P_X) < \Lambda$ for given type P_X, then the case $\Lambda_2(P_X) < \Lambda$ follows accordingly.

We consider any deterministic code $\mathcal{C}_{\det}(\mathfrak{W}^n)$ for the AVBBC $\mathfrak{W}^n$ with codewords $x^n_{m_1,m_2} = (x_{m_1,m_2,1}, ..., x_{m_1,m_2,n}) \in \mathcal{X}^n$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$, and the corresponding decoding sets $\mathcal{D}^{(1)}_{m_2|m_1} \subset \mathcal{Y}^n_1$ at node 1. Next, for any channel $U_1 \in \mathcal{U}_1$ which symmetrizes the AVBBC $\mathfrak{W}^n$ in the sense of Definition 8, we define random variables $S^n_{m_1,m_2} = (S_{m_1,m_2,1}, ..., S_{m_1,m_2,n}) \in \mathcal{S}^n$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$ with statistically independent elements and

$$\mathbb{P}\{S_{m_1,m_2,k} = s\} = U_1(s|x_{m_1,m_2,k}) \tag{44}$$

Then for each $m_1 \in \mathcal{M}_1$ the following holds. For each pair $(i,j) \in \mathcal{M}_2 \times \mathcal{M}_2$ and every $y^n_1 = (y_{1,1}, ..., y_{1,n}) \in \mathcal{Y}^n_1$ we have

$$\mathbb{E}\big[W^n_1(y^n_1|x^n_{m_1,i}, S^n_{m_1,j})\big] = \prod_{k=1}^{n} \mathbb{E}\big[W_1(y_{1,k}|x_{m_1,i,k}, S_{m_1,j,k})\big]$$
$$= \prod_{k=1}^{n} \sum_{s\in\mathcal{S}} W_1(y_{1,k}|x_{m_1,i,k}, s)\mathbb{P}\{S_{m_1,j,k} = s\}$$
$$= \prod_{k=1}^{n} \sum_{s\in\mathcal{S}} W_1(y_{1,k}|x_{m_1,i,k}, s)U_1(s|x_{m_1,j,k})$$

where the equalities follow from the memoryless property of the channel, the definition of the expectation, and (44). Since the AVBBC $\mathfrak{W}^n$ is $\mathcal{Y}_1$-symmetrizable, *i.e.*, (4) holds, it follows that

$$\prod_{k=1}^{n} \sum_{s\in\mathcal{S}} W_1(y_{1,k}|x_{m_1,i,k}, s)U_1(s|x_{m_1,j,k}) = \prod_{k=1}^{n} \sum_{s\in\mathcal{S}} W_1(y_{1,k}|x_{m_1,j,k}, s)U_1(s|x_{m_1,i,k})$$
$$= \prod_{k=1}^{n} \mathbb{E}\big[W_1(y_{1,k}|x_{m_1,j,k}, S_{m_1,i,k})\big]$$
$$= \mathbb{E}\big[W^n_1(y^n_1|x^n_{m_1,j}, S^n_{m_1,i})\big]$$

so that we finally end up with

$$\mathbb{E}\big[W_1^n(y_1^n|x_{m_1,i}^n, \mathrm{S}_{m_1,j}^n)\big] = \mathbb{E}\big[W_1^n(y_1^n|x_{m_1,j}^n, \mathrm{S}_{m_1,i}^n)\big] \tag{45}$$

For the probability of error at node 1 this implies the following. For $i \neq j$ we have

$$\begin{aligned}
\mathbb{E}\big[e_1((m_1,i), &\mathrm{S}_{m_1,j}^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))\big] + \mathbb{E}\big[e_1((m_1,j), \mathrm{S}_{m_1,i}^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))\big] \\
&= \mathbb{E}\big[W_1^n((D_{i|m_1}^{(1)})^c|x_{m_1,i}^n, \mathrm{S}_{m_1,j}^n)\big] + \mathbb{E}\big[W_1^n((D_{j|m_1}^{(1)})^c|x_{m_1,j}^n, \mathrm{S}_{m_1,i}^n)\big] \\
&= \mathbb{E}\big[W_1^n((D_{i|m_1}^{(1)})^c|x_{m_1,i}^n, \mathrm{S}_{m_1,j}^n)\big] + \mathbb{E}\big[W_1^n((D_{j|m_1}^{(1)})^c|x_{m_1,i}^n, \mathrm{S}_{m_1,j}^n)\big] \\
&\geq \mathbb{E}\big[W_1^n((D_{i|m_1}^{(1)})^c|x_{m_1,i}^n, \mathrm{S}_{m_1,j}^n)\big] + \mathbb{E}\big[W_1^n(D_{i|m_1}^{(1)}|x_{m_1,i}^n, \mathrm{S}_{m_1,j}^n)\big] \\
&= \mathbb{E}\big[W_1^n((D_{i|m_1}^{(1)})^c \cup D_{i|m_1}^{(1)}|x_{m_1,i}^n, \mathrm{S}_{m_1,j}^n)\big] = 1
\end{aligned}$$

where the second equality follows from (45). For a fixed $m_1 \in \mathcal{M}_1$ this leads to

$$\begin{aligned}
\frac{1}{M_{2,n}} \sum_{j=1}^{M_{2,n}} \mathbb{E}\big[\bar{e}_1(\mathrm{S}_{m_1,j}^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))\big] &= \frac{1}{M_{2,n}} \frac{1}{M_{1,n}M_{2,n}} \sum_{j=1}^{M_{2,n}} \sum_{m_1'=1}^{M_{1,n}} \sum_{i=1}^{M_{2,n}} \mathbb{E}\big[e_1((m_1',i), \mathrm{S}_{m_1,j}^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))\big] \\
&\geq \frac{1}{M_{1,n}(M_{2,n})^2} \sum_{m_1'}^{M_{1,n}} \frac{M_{2,n}(M_{2,n}-1)}{2} \\
&= \frac{M_{1,n}M_{2,n}(M_{2,n}-1)}{2M_{1,n}(M_{2,n})^2} = \frac{M_{2,n}-1}{2M_{2,n}}
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
\frac{1}{M_{1,n}M_{2,n}} \sum_{m_1=1}^{M_{1,n}} \sum_{j=1}^{M_{2,n}} \mathbb{E}\big[\bar{e}_1(\mathrm{S}_{m_1,j}^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))\big] &= \frac{1}{M_{1,n}} \sum_{m_1=1}^{M_{1,n}} \left(\frac{1}{M_{2,n}} \sum_{j=1}^{M_{2,n}} \mathbb{E}\big[\bar{e}_1(\mathrm{S}_{m_1,j}^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))\big] \right) \\
&\geq \frac{1}{M_{1,n}} \sum_{m_1=1}^{M_{1,n}} \frac{M_{2,n}-1}{2M_{2,n}} = \frac{M_{2,n}-1}{2M_{2,n}}
\end{aligned}$$

which implies that there exists at least one $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$ such that

$$\mathbb{E}\big[\bar{e}_1(\mathrm{S}_{m_1,m_2}^n|\mathcal{C}_{\mathrm{det}}(\mathfrak{W}^n))\big] \geq \frac{M_{2,n}-1}{2M_{2,n}} \geq \frac{1}{4} \tag{46}$$

Next, we restrict to codewords of type P_{X}, *i.e.*, $x_{m_1,m_2}^n \in \mathcal{T}_{\mathrm{X}}^n$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$, with $\Lambda_1(P_{\mathrm{X}}) < \Lambda$. Further, we choose $U_1 \in \mathcal{U}_1$ such that it attains the minimum in (6). Then, with (5b) we get for the expectation

$$\begin{aligned}
\mathbb{E}[l(\mathrm{S}_{m_1,m_2}^n)] &= \frac{1}{n} \sum_{k=1}^{n} \sum_{s \in \mathcal{S}} l(s)U_1(s|x_{m_1,m_2,k}) \\
&= \sum_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} P_{\mathrm{X}}(x)U_1(s|x)l(s) \\
&= \Lambda_1(P_{\mathrm{X}})
\end{aligned}$$

and the variance

$$\mathrm{var}[l(\mathrm{S}_{m_1,m_2}^n)] \leq \frac{l_{\max}^2}{n}$$

From Chebyshev's inequality we obtain

$$
\begin{aligned}
\mathbb{P}\big\{l(\mathrm{S}^n_{m_1,m_2}) > \Lambda\big\} &= \mathbb{P}\big\{l(\mathrm{S}^n_{m_1,m_2}) - \mathbb{E}[l(\mathrm{S}^n_{m_1,m_2})] > \Lambda - \Lambda_1(P_{\mathrm{X}})\big\} \\
&\leq \frac{1}{n}\frac{l^2_{\max}}{(\Lambda - \Lambda_1(P_{\mathrm{X}}))^2}
\end{aligned}
\tag{47}
$$

Finally, since $\mathbb{E}[\bar{e}_1(\mathrm{S}^n_{m_1,m_2}|\mathcal{C}_{\det}(\mathfrak{W}^n))] \leq \max_{s^n:l(s^n)\leq\Lambda} \bar{e}_1(s^n|\mathcal{C}_{\det}(\mathfrak{W}^n)) + \mathbb{P}\{l(\mathrm{S}^n_{m_1,m_2}) > \Lambda\}$, we get from (46) and (47)

$$
\begin{aligned}
\max_{s^n:l(s^n)\leq\Lambda} \bar{e}_1(s^n|\mathcal{C}_{\det}(\mathfrak{W}^n)) &\geq \mathbb{E}[\bar{e}_1(\mathrm{S}^n_{m_1,m_2}|\mathcal{C}_{\det}(\mathfrak{W}^n))] - \mathbb{P}\{l(\mathrm{S}^n_{m_1,m_2}) > \Lambda\} \\
&\geq \frac{M_{2,n}-1}{2M_{2,n}} - \frac{1}{n}\frac{l^2_{\max}}{(\Lambda - \Lambda_1(P_{\mathrm{X}}))^2}
\end{aligned}
$$

which proves the first part of the lemma. Clearly, the second part where $\Lambda_2(P_{\mathrm{X}}) < \Lambda$ for given type P_{X} follows accordingly using the same argumentation.

A.2. Proof of Lemma 2

In the following we show that if we select randomly $M_{1,n}M_{2,n}$ codewords with $M_{1,n} = 2^{nR_2}$ and $M_{2,n} = 2^{nR_1}$, then these codewords will possess, with probability close to 1, the properties (18a)–(18f) as stated in Lemma 2. Thereby, we follow [4, Lemma 3], where a similar result is proved for the single-user case. Further, an analogous version of the lemma for the arbitrarily varying MAC can be found in [10]. But first, we restate a lemma which will be essential to prove the desired properties of the codewords.

Lemma 7. *Let* $\mathrm{Z}^n_1, ..., \mathrm{Z}^n_N$ *be arbitrary random variables, and let* $f_i(\mathrm{Z}^n_1, ..., \mathrm{Z}^n_i)$ *be arbitrary with* $0 \leq f_i \leq 1$, $i = 1, ..., N$. *Then the condition*

$$
\mathbb{E}\big[f_i(\mathrm{Z}^n_1, ..., \mathrm{Z}^n_i)|\mathrm{Z}^n_1, ..., \mathrm{Z}^n_{i-1}\big] \leq a \quad \text{almost surely}
\tag{48}
$$

$i = 1, ..., N$, *implies that*

$$
\mathbb{P}\Big\{\frac{1}{N}\sum_{i=1}^{N} f_i(\mathrm{Z}^n_1, ..., \mathrm{Z}^n_i) > t\Big\} \leq \exp\big(-N(t - a\log e)\big)
\tag{49}
$$

Proof. The proof can be found in [4] (Lemma A1) or [10]. $\qquad\qquad\square$

Now, we turn to the proof of Lemma 2. As in [4] (Lemma 3) let $\mathrm{Z}^n_{m_1,m_2}$, $m_1 = 1, ..., M_{1,n}$, $m_2 = 1, ..., M_{2,n}$ be independent random variables, each uniformly distributed on $\mathcal{T}^n_{\mathrm{X}}$. Further, we fix an $x^n \in \mathcal{T}^n_{\mathrm{X}}$, $s^n \in \mathcal{S}^n_\Lambda$, and a joint type $P_{\mathrm{XX'S}}$ with $P_{\mathrm{XS}} = P_{x^n,s^n}$ and $P_{\mathrm{X'}} = P_{\mathrm{X}}$.

First, we show that for each $m_1 \in \mathcal{M}_1$ the properties (18a)–(18c) are satisfied. Therefore, we fix an arbitrary $m_1 \in \mathcal{M}_1$ for the following analysis. We define

$$
f_{m_1,j}(\mathrm{Z}^n_{m_1,1}, ..., \mathrm{Z}^n_{m_1,j}) = \begin{cases} 1 & \text{if } \mathrm{Z}^n_{m_1,j} \in \mathcal{T}^n_{\mathrm{X'|XS}}(x^n, s^n) \\ 0 & \text{otherwise} \end{cases}
\tag{50}
$$

and apply Lemma 7. Now, the condition (48) of Lemma 7 is fulfilled with

$$
\begin{aligned}
a &= \mathbb{P}\big\{ Z_{m_1,j}^n \in \mathcal{T}_{X'|XS}^n(x^n, s^n) \big\} \\
&= \frac{|\mathcal{T}_{X'|XS}^n(x^n, s^n)|}{|\mathcal{T}_X^n|} \\
&\leq \frac{\exp\big(nH(X'|XS)\big)}{(n+1)^{-|\mathcal{X}|}\exp\big(nH(X)\big)} \\
&= (n+1)^{|\mathcal{X}|}\exp\big(-nI(X';XS)\big)
\end{aligned}
$$

where the inequality follows from Fact 2, *cf.* Section 2, and the last equality because $H(X') = H(X)$. For $R_1 = \frac{1}{n}\log M_{2,n}$ we choose

$$
t = \frac{1}{M_{2,n}}\exp\big(n(|R_1 - I(X';XS)|^+ + \epsilon)\big)
$$

so that $M_{2,n}(t - a\log e) \geq \frac{1}{2}\exp(n\epsilon)$ if $n \geq n_1(\epsilon)$, where

$$
n_1(\epsilon) = \min\Big\{ n : (n+1)^{|\mathcal{X}|}\log e < \frac{1}{2}\exp(n\epsilon) \Big\} \tag{51}
$$

Then (49) yields

$$
\mathbb{P}\Big\{ \big|\{j : Z_{m_1,j}^n \in \mathcal{T}_{X'|XS}^n(x^n, s^n)\}\big| > \exp\big(n(|R_1 - I(X';XS)|^+ + \epsilon)\big) \Big\} < \exp\big(-\frac{1}{2}\exp(n\epsilon)\big) \tag{52}
$$

The same reasoning holds if we replace $\mathcal{T}_{X'|XS}^n(x^n, s^n)$ by $\mathcal{T}_{X'|S}^n(s^n)$ in (50). Consequently, we similarly obtain

$$
\mathbb{P}\Big\{ \big|\{j : Z_{m_1,j}^n \in \mathcal{T}_{X'|S}^n(s^n)\}\big| > \exp\big(n(|R_1 - I(X';S)|^+ + \epsilon)\big) \Big\} < \exp\big(-\frac{1}{2}\exp(n\epsilon)\big) \tag{53}
$$

Moreover, if $I(X';S) > \epsilon$ (and remember that $R_1 \geq \epsilon$ since $R_1 = \frac{1}{n}\log M_{2,n}$ and $M_{2,n} \geq 2^{n\epsilon}$ as assumed) we obtain by replacing ϵ with $\epsilon/2$ for $n \geq n_1(\epsilon/2)$ from (53) that

$$
\mathbb{P}\Big\{ \frac{1}{M_{2,n}}\big|\{j : Z_{m_1,j}^n \in \mathcal{T}_{X'|S}^n(s^n)\}\big| > \exp\big(-n\frac{\epsilon}{2}\big) \Big\} < \exp\big(-\frac{1}{2}\exp(n\frac{\epsilon}{2})\big) \tag{54}
$$

Equations (52) and (54) allow us to establish the first two properties of the codewords, *i.e.*, (18a) and (18b). To obtain the third property (18c) we define the set $\mathcal{A}_{m_1,i}$ as the set of indices $j < i$ such that $z_{m_1,j}^n \in \mathcal{T}_{X'|S}^n(s^n)$. If $|\mathcal{A}_{m_1,i}| > \exp\big(n(|R_1 - I(X';S)|^+ + \epsilon/4)\big)$, we set $\mathcal{A}_{m_1,i} = \emptyset$. Let

$$
f_{m_1,i}(z_{m_1,1}^n, ..., z_{m_1,i}^n) = \begin{cases} 1 & \text{if } z_{m_1,i}^n \in \bigcup_{j\in\mathcal{A}_{m_1,i}} \mathcal{T}_{X|X'S}^n(z_{m_1,j}^n, s^n) \\ 0 & \text{otherwise.} \end{cases} \tag{55}
$$

If we replace ϵ with $\epsilon/4$, it follows from (53) that for $n \geq n_1(\epsilon/4)$

$$
\mathbb{P}\Big\{ \sum_{i=1}^{M_{2,n}} f_{m_1,i}(Z_{m_1,1}^n, ..., Z_{m_1,i}^n) \neq \big|\{i : Z_{m_1,i}^n \in \mathcal{T}_{X|X'S}^n(Z_{m_1,j}^n, s^n) \text{ for some } j < i\}\big| \Big\}
$$

$$
< \exp\big(-\frac{1}{2}\exp(n\frac{\epsilon}{4})\big) \tag{56}
$$

Then, we get from (55)

$$\mathbb{E}\big[f_{m_1,i}(Z^n_{m_1,1}, ..., Z^n_{m_1,i})|Z^n_{m_1,1}, ..., Z^n_{m_1,i-1}\big]$$

$$=\mathbb{P}\Big\{Z^n_{m_1,i} \in \bigcup_{j \in \mathcal{A}_{m_1,i}} \mathcal{T}^n_{\mathrm{X}|\mathrm{X'S}}(Z^n_{m_1,j}, s^n)\Big| Z^n_{m_1,1}, ..., Z^n_{m_1,i-1}\Big\}$$

$$\leq |\mathcal{A}_{m_1,i}| \frac{\exp\big(nH(\mathrm{X}|\mathrm{X'S})\big)}{(n+1)^{-|\mathcal{X}|}\exp\big(nH(\mathrm{X})\big)}$$

$$\leq (n+1)^{|\mathcal{X}|}\exp\big(n(|R_1 - I(\mathrm{X'};\mathrm{S})|^+ - I(\mathrm{X};\mathrm{X'S}) + \tfrac{\epsilon}{4})\big)$$

which follows from the independence of the $Z^n_{m_1,i}$ and Fact 2. Next, we assume that $I(\mathrm{X};\mathrm{X'S}) > |R_1 - I(\mathrm{X'};\mathrm{S})|^+ + \epsilon$ so that (48) of Lemma 7 is satisfied with

$$a = (n+1)^{|\mathcal{X}|}\exp\big(-n\frac{3\epsilon}{4}\big)$$

With $t = \exp\big(-n\frac{\epsilon}{2}\big)$ and for $n \geq n_1(\epsilon/4)$, *cf.* also (51), Lemma 7 yields

$$\mathbb{P}\Big\{\frac{1}{M_{2,n}}\sum_{i=1}^{M_{2,n}} f_{m_1,i}(Z^n_{m_1,1}, ..., Z^n_{m_1,i}) > \exp\big(-n\frac{\epsilon}{2}\big)\Big\} < \exp\big(-\frac{M_{2,n}}{2}\exp\big(-n\frac{\epsilon}{2}\big)\big)$$

$$< \exp\big(-\frac{1}{2}\exp\big(n\frac{\epsilon}{2}\big)\big)$$

where the last inequality follows from the assumption $M_{2,n} \geq 2^{n\epsilon}$. Combining this with (56), we get

$$\mathbb{P}\Big\{\frac{1}{M_{2,n}}\big|\{i : Z^n_{m_1,i} \in \mathcal{T}^n_{\mathrm{X}|\mathrm{X'S}}(Z^n_{m_1,j}, s^n) \text{ for some } j < i\}\big| > \exp\big(-n\frac{\epsilon}{2}\big)\Big\}$$

$$< \exp\big(-\frac{1}{2}\exp\big(n\frac{\epsilon}{2}\big)\big) + \exp\big(-\frac{1}{2}\exp\big(n\frac{\epsilon}{4}\big)\big)$$

$$< 2\exp\big(-\frac{1}{2}\exp\big(n\frac{\epsilon}{4}\big)\big)$$

If we replace "for some $j < i$" by "for some $j > i$" in (56), we obtain the same by symmetry. Consequently, we end up with

$$\mathbb{P}\Big\{\frac{1}{M_{2,n}}\big|\{i : Z^n_{m_1,i} \in \mathcal{T}^n_{\mathrm{X}|\mathrm{X'S}}(Z^n_{m_1,j}, s^n) \text{ for some } j \neq i\}\big| > \exp\big(-n\frac{\epsilon}{2}\big)\Big\}$$

$$< 4\exp\big(-\frac{1}{2}\exp\big(n\frac{\epsilon}{4}\big)\big) \qquad (57)$$

if $I(\mathrm{X};\mathrm{X'S}) > |R_1 - I(\mathrm{X'};\mathrm{S})|^+ + \epsilon$ and $n \geq n_1(\epsilon/4)$.

Now we are in the position to complete the first part of the proof. The number of all possible sequences $x^n \in \mathcal{T}^n_{\mathrm{X}}$, states $s^n \in \mathcal{S}^n_\Lambda$ and joint types $P_{\mathrm{XX'S}}$ grows exponentially with n. Since the bounds (52), (54) and (57) are doubly exponential probability bounds, the inequalities

$$\big|\{j : z^n_{m_1,j} \in \mathcal{T}^n_{\mathrm{X'}|\mathrm{XS}}(x^n, s^n)\}\big| \leq \exp\big(n(|R_1 - I(\mathrm{X'};\mathrm{XS})|^+ + \epsilon)\big)$$

$$\frac{1}{M_{2,n}}\big|\{j : z^n_{m_1,j} \in \mathcal{T}^n_{\mathrm{X'}|\mathrm{S}}(s^n)\}\big| \leq \exp\big(-n\frac{\epsilon}{2}\big) \qquad \text{if } I(\mathrm{X};\mathrm{S}) > \epsilon$$

$$\frac{1}{M_{2,n}}\big|\{i : z^n_{m_1,i} \in \mathcal{T}^n_{\mathrm{X}|\mathrm{X'S}}(z^n_{m_1,j}, s^n) \text{ for some } j \neq i\}\big| \leq \exp\big(-n\frac{\epsilon}{2}\big)$$

$$\text{if } I(\mathrm{X};\mathrm{X'S}) - |R_1 - I(\mathrm{X'};\mathrm{S})|^+ > \epsilon$$

hold simultaneously with probability arbitrarily close to 1 if n is sufficiently large and $n \geq n_0(\epsilon) = n_1(\epsilon/4)$. This establishes the properties (18a)–(18c).

It remains to show that for each fixed $m_2 \in \mathcal{M}_2$ the properties (18d)–(18f) simultaneously hold for n large enough. This can be done analogously to the first three properties and is therefore omitted for brevity.

A.3. Proof of Lemma 3

The lemma is proved by contradiction as done in [4] (Lemma 4) for the single-user AVC. For receiving node i, $i = 1, 2$, suppose that the quintuple X, X', S, S', Y_i satisfies the conditions given in (19). Since $P_{XSY_i} \in \mathcal{D}_{\eta_i}(\Lambda)$ and $I(XY_i; X'|S) \leq \eta_i$, we have

$$
\begin{aligned}
2\eta_i &\geq D(P_{XSY_i} \| P_X \otimes P_S \otimes W_i) + I(XY_i; X'|S) \\
&= \sum_{x,s,y_i} P_{XSY_i}(x, s, y_i) \log \frac{P_{XSY_i}(x, s, y_i)}{P_X(x)P_S(s)W_i(y_i|x,s)} \\
&\quad + \sum_{x,x',s,y_i} P_{XX'SY_i}(x, x', s, y_i) \log \frac{P_{X'|XSY_i}(x'|x, s, y_i)}{P_{X'|S}(x'|s)} \\
&= \sum_{x,x',s,y_i} P_{XX'SY_i}(x, x', s, y_i) \log \frac{P_{XX'SY_i}(x, x', s, y_i)}{P_X(x)P_{X'S}(x', s)W_i(y_i|x,s)} \\
&= \sum_{x,x',y_i} \sum_s P_{XX'SY_i}(x, x', s, y_i) \log \frac{P_{XX'SY_i}(x, x', s, y_i)}{P_X(x)P_{X'}(x')P_{S|X'}(s|x')W_i(y_i|x,s)} \\
&\geq \sum_{x,x',y_i} P_{XX'Y_i}(x, x', y_i) \log \frac{P_{XX'Y_i}(x, x', y_i)}{P_X(x)P_{X'}(x') \sum_s P_{S|X'}(s|x')W_i(y_i|x,s)} \\
&= D(P_{XX'Y_i} \| P_X \otimes P_{X'} \otimes V_i')
\end{aligned}
\tag{58}
$$

with $V_i'(y_i|x, x') = \sum_s P_{S|X'}(s|x')W_i(y_i|x, s)$ and the last inequality follows from the log-sum inequality.

From [3] we know that we can bound the variational distance between two probability distributions from above by the square root of their divergence times an absolute constant. (This bound with a worse constant was first given by Pinsker [70] and is therefore also known as Pinsker's inequality.) With this and (58) we get

$$
\begin{aligned}
\sum_{x,x',y_i} \big| P_{XX'Y_i}(x, x', y_i) - P_X(x)P_{X'}(x')V_i'(y_i|x, x') \big| \\
\leq c\sqrt{D(P_{XX'Y_i} \| P_X \otimes P_{X'} \otimes V_i')} \leq c\sqrt{2\eta_i}
\end{aligned}
\tag{59}
$$

with $c = \sqrt{2 \ln 2}$. Similarly, since $P_{X'S'Y_i} \in \mathcal{D}_{\eta_i}(\Lambda)$ and $I(X'Y_i; X|S') \leq \eta_i$, cf. (19), we obtain

$$
\sum_{x,x',y_i} \big| P_{XX'Y_i}(x, x', y_i) - P_{X'}(x')P_X(x)V_i(y_i|x', x) \big| \leq c\sqrt{2\eta_i}
\tag{60}
$$

with $c = \sqrt{2 \ln 2}$ and $V_i(y_i|x', x) = \sum_s P_{S'|X}(s|x)W_i(y_i|x', s)$. Next, (59) and (60) together imply

$$
\sum_{x,x',y_i} P_X(x)P_{X'}(x') \big| V_i(y_i|x', x) - V_i'(y_i|x, x') \big| \leq 2c\sqrt{2\eta_i}
$$

Since $\min_x P_X(x) \geq \beta$, it immediately follows that

$$\max_{x,x',y_i} \left| V_i(y_i|x',x) - V_i'(y_i|x,x') \right| \leq \frac{2c\sqrt{2\eta_i}}{\beta^2} \tag{61}$$

Lemma 8. *For any AVBBC $\mathfrak{W}^n$ with state constraint Λ and any input P_X with $\Lambda_i(P_X) \geq \Lambda + \alpha$, $\alpha > 0$, $i = 1,2$, for which each pair $U_1 : \mathcal{X} \to \mathcal{P}(\mathcal{S})$ and $U_2 : \mathcal{X} \to \mathcal{P}(\mathcal{S})$ satisfies*

$$\sum_{x,s} P_X(x)U_1(s|x)l(s) \leq \Lambda \tag{62a}$$

$$\sum_{x,s} P_X(x)U_2(s|x)l(s) \leq \Lambda \tag{62b}$$

there exists some $\xi > 0$ such that

$$\max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U_1(s|x') - \sum_s W_i(y_i|x',s)U_2(s|x) \right| \geq \xi \qquad i = 1,2 \tag{63}$$

Proof. The proof can be found in Appendix A.4. □

If we choose $U_1 = P_{S|X'}$ and $U_2 = P_{S'|X}$, we obtain from (63)

$$\max_{x,x',y_i} \left| V_i(y_i|x',x) - V_i'(y_i|x,x') \right| \geq \xi \tag{64}$$

Finally, (61) and (64) yield

$$\eta_i \geq \frac{\xi^2\beta^4}{8c^2}, \qquad i = 1,2$$

which contradicts the assumption that η_i can be chosen arbitrarily small proving the lemma.

A.4. Proof of Lemma 8

As in [4] (Lemma A2) we can interchange the two sums and then x and x' without changing the maximum in (63). Thus we can write for all $U_1, U_2 : \mathcal{X} \to \mathcal{P}(\mathcal{S})$

$$\max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U_1(s|x') - \sum_s W_i(y_i|x',s)U_2(s|x) \right|$$

as

$$\max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U_2(s|x') - \sum_s W_i(y_i|x',s)U_1(s|x) \right|$$

so that we get

$$
\max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U_1(s|x') - \sum_s W_i(y_i|x',s)U_2(s|x) \right|
$$

$$
= \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)\frac{U_1(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_2(s|x)}{2} \right|
$$

$$
+ \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)\frac{U_2(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_1(s|x)}{2} \right|
$$

$$
\geq \max_{x,x',y_i} \left\{ \left| \sum_s W_i(y_i|x,s)\frac{U_1(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_2(s|x)}{2} \right| \right.
$$

$$
\left. + \left| \sum_s W_i(y_i|x,s)\frac{U_2(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_1(s|x)}{2} \right| \right\}
$$

$$
\geq \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)\frac{U_1(s|x')+U_2(s|x')}{2} - \sum_s W_i(y_i|x',s)\frac{U_1(s|x)+U_2(s|x)}{2} \right|
$$

$$
= \max_{x,x',y_i} \left| \sum_s W_i(y_i|x,s)U(s|x') - \sum_s W_i(y_i|x',s)U(s|x) \right| \tag{65}
$$

with $U = \frac{1}{2}(U_1 + U_2)$. Further, since U_1 and U_2 satisfy (62) for some P_X, then also U satisfy

$$
\sum_{x,s} P_X(x)U(s|x)l(s) \leq \Lambda \tag{66}
$$

Since (65) can be considered as a continuous function of the pair (P_X, U) on the compact set of all channels $U : \mathcal{X} \to \mathcal{P}(\mathcal{S})$, it attains its minimum for some (P_X^*, U^*), where the minimization is taken over all channels U that satisfy (66). Additionally, since (P_X^*, U^*) satisfies (66), U^* cannot satisfy (4) which in turn implies that $\max_{x,x',y_i} |\sum_s W_i(y_i|x,s)U(s|x') - \sum_s W_i(y_i|x',s)U(s|x)| > 0$ completing the proof.