List Decoding for Arbitrarily Varying Broadcast Channels With Receiver Side Information

Rafael F. Schaefer, *Member, IEEE*, and Holger Boche, *Fellow, IEEE*

Abstract—In this paper, the discrete memoryless arbitrarily varying broadcast channel (AVBC) with receiver side information is studied and its random code and deterministic code capacity regions are derived for the average error criterion. In addition, it is analyzed for deterministic list codes and it is shown that the corresponding list capacity region displays a behavior, which is similar to Ahlswede's famous dichotomy result for the single-user arbitrarily varying channel: it either equals the random code capacity region or otherwise has an empty interior. This is characterized in terms of list sizes at the receivers and an appropriate concept of symmetrizability for the AVBC with receiver side information. The scenario studied here is motivated by the broadcast phase of bidirectional relaying, where a half-duplex relay node establishes a bidirectional communication between two other nodes using a decode-and-forward protocol. The relay decodes the messages both nodes have sent in the initial multiple access phase and broadcasts a re-encoded composition of them in the succeeding broadcast phase. Then, the broadcast phase corresponds to the AVBC with receiver side information, which differs from the classical broadcast channel, since both receivers can exploit their own messages from the previous phase as side information for decoding.

Index Terms—Broadcast channel with receiver side information, arbitrarily varying channel (AVC), list decoding, symmetrizability, capacity region, bidirectional relaying.

I. INTRODUCTION

RECENT research progress shows that relays can significantly increase the performance of wireless networks. Since a relay cannot transmit and receive at the same time and frequency, it needs orthogonal resources for transmission and reception which can be done more efficiently if bidirectional communication is considered [1]–[4]. In this paper, we consider *bidirectional relaying* in a three-node network, where a relay node establishes a bidirectional communication between two other nodes. This concept is also known as two-way relaying.

In this paper, we consider a two-phase decode-and-forward protocol as shown in Fig. 1. In the initial multiple access
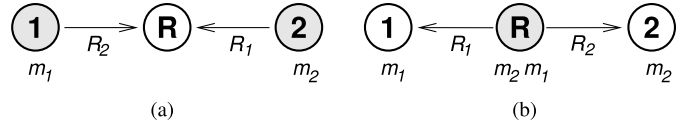
Fig. 1. Decode-and-forward bidirectional relaying. (a) In the initial multiple access (MAC) phase, nodes 1 and 2 transmit their messages m_1 and m_2 with rates R_2 and R_1 to the relay node. (b) In the succeeding broadcast (BC) phase, the relay forwards the messages to the corresponding receiving nodes.

phase, both nodes transmit their messages to the relay node. Since we assume the relay to decode both messages, we end up with the classical multiple access channel (MAC), see [5] and [6]. In the succeeding broadcast phase, it remains for the relay to broadcast an optimal re-encoded message based on the network coding idea so that both nodes are able to decode the other's message using their own message from the previous phase as side information. Note that due to the available side information at the receiving nodes this channel differs from the classical broadcast channel. To emphasize this property, this channel is called *broadcast channel (BC) with receiver side information*. In the context of bidirectional relaying, it is also known as *bidirectional broadcast channel*. Capacity achieving strategies for perfect channel state information at all nodes can be found in [7]–[10] and it is shown that the achievable rates for both users may differ although both nodes receive the same codeword of the relay node. This is an important difference to the XOR coding approach and the broadcast of a common message and is permitted by the available side information at the receiving nodes.

Channel uncertainty is an ubiquitous phenomenon in wireless systems and should be taken into account for the design of practical systems, see [11] for a short survey about reliable communication under channel uncertainty. A well accepted model for channel uncertainty is to assume that all nodes do not know the exact channel realization, but only know that this realization belongs to a pre-specified set of channels. If this realization remains fixed during the whole transmission of a codeword, this corresponds to the *compound channel* [12]–[14].

The variation of the channel from symbol to symbol in an unknown and arbitrary manner is an additional effect of channel uncertainty. This is the concept of *arbitrarily varying channels* (AVC) [15]–[17]. Interestingly, for AVCs the random code and deterministic code capacity need not be equal. In more detail, Ahlswede showed in his famous work [16] that the AVC displays a dichotomy behavior: the deterministic

code capacity either equals the random code capacity or else is zero. Unfortunately, it was not characterized in detail when the deterministic code capacity is non-zero. Finally, in [18] and [19] non-symmetrizability was established as a necessary and sufficient condition for the AVC to have a non-zero deterministic code capacity. Roughly speaking, a symmetrizable AVC leads to a zero deterministic code capacity, since such a channel can "emulate" a validate input which makes it impossible for the decoder to decide on the correct codeword. The concept of list decoding might help to overcome such impairments. The corresponding list code capacity is established and analyzed in detail in [20] and [21]. Interestingly, it is shown that the dichotomy behavior is preserved: the list code capacity either equals the random code capacity or else is zero. Interestingly, a complete characterization is established using an appropriate extension of the concept of symmetrizability. List decoding for AVCs under state constraints is analyzed in [22] where it is shown that in this case the corresponding list code capacity can be positive but strictly smaller than the random code capacity. Rateless coding approaches for AVCs are presented in [23] and [24].

A. Related Work

There are some first works on the BC with receiver side information under channel uncertainty. For example, the compound BC with receiver side information is analyzed in [25], where the corresponding capacity region is derived. The scenario where the channel states are non-causally known at the encoder is studied in [26]. There, based on the Gelfand-Pinsker coding strategy for the single-user case [27], an achievable rate region and an outer bound are established. The capacity region is only given for the special case where in addition one receiver knows the channel states. For the general case, a precise characterization of the capacity region remains open, which is in contrast to the single-user case [27], [28].

To the best of our knowledge, there are no works for the BC with receiver side information for arbitrarily varying channels, but the single-user AVC is extended to several other multi-user settings. The arbitrarily varying multiple access channel (AVMAC) is analyzed in [29]–[31]. The random code and deterministic code capacity regions are established. It is shown that the latter may have an empty interior which is completely characterized and analyzed in terms of an appropriate concept of symmetrizability [30], [31]. List decoding for the AVMAC is studied in [32], but only upper and lower bounds on the required list sizes are given. For the (general) arbitrarily varying broadcast channel (AVBC) there are only partial results known until now. An achievable deterministic code rate region for the AVBC was established in [29] but it is not further analyzed when its interior is non-empty. On the other hand, [33] analyzes an achievable deterministic code rate region of the AVBC in terms of symmetrizability but imposes the further assumption of degraded message sets. In [34] achievable rate regions of the AVBC with degraded channels and non-causal channel state information at the transmitter are studied. All these works [29], [33], [34] haven in common that only achievable rate regions are presented but no converse

results or outer bounds on the capacity regions are given. In addition, they all do not consider list decoding.

B. Contributions and Paper Organization

In this paper, we study the *arbitrarily varying broadcast channel (AVBC) with receiver side information* in detail and completely solve it. The contributions are as follows. We start in Section II with some information theoretical and combinatorial preliminaries, before we introduce the AVBC with receiver side information in detail in Section III. As a first step, in Section IV we establish a complete characterization of the random code capacity region. The proof of achievability is based on the results for the compound BC with receiver side information [25] and the robustification technique [35], [36] adapted to the corresponding communication scenario. In addition, we also present the weak converse, which then establishes the random code capacity region.

Subsequently, we consider the AVBC with receiver side information under list decoding. We show in Section V that the list capacity region of the AVBC with receiver side information either equals its random code capacity region or else has an empty interior. A similar behavior has been observed for the single-user AVC under list decoding in [21]. In [21] this was shown "indirectly" by following the approach of Csiszár and Narayan [19] without making use of the corresponding result of the random code capacity. However, we follow the approach of Ahlswede presented for the point-to-point AVC under deterministic coding [16]. There, it is proved by combining a deterministic code with a capacity-achieving random code. This has to be carefully extended as we have to combine a list code with different list sizes for both receiving nodes with a corresponding capacity-achieving random code. This has to be done in such a way that the final concatenated code is still a valid list code for the BC with receiver side information, i.e., it has to satisfy the different list sizes at both receiving nodes and the probabilities of error must behave as wanted.

It remains to characterize when the list capacity region has an empty interior. Therefore, in Sections VI and VII we use an appropriate concept of symmetrizability to establish non-symmetrizability as a necessary and sufficient condition for the list capacity region to have a non-empty interior. Furthermore, we present a weak converse that completely establishes the list capacity region. We end up with a conclusion in Section VIII.

C. Notation

Discrete random variables are denoted by capital letters and their corresponding realizations and ranges by lower case letters and script letters respectively; $\mathbb{N}$ and $\mathbb{R}_+$ denote the set of non-negative integers and non-negative real numbers; all logarithms, exponentials, and information quantities are taken to the base 2; $I(\cdot;\cdot)$, $H(\cdot)$, and $D(\cdot\|\cdot)$ are the traditional mutual information, entropy, and (Kullback-Leibler) information divergence; $W^{\otimes n}$ is the n-th memoryless extension of the stochastic matrix W, i.e., $W^{\otimes n}(y^n|x^n) := \prod_{k=1}^{n} W(y_k|x_k)$; $\mathbb{E}_X[\cdot]$ is the expectation with respect to X; $\mathcal{P}(\mathcal{X})$ is the set of all probability distributions on $\mathcal{X}$; the product distribution $P_X \otimes P_Y$ is defined by the product marginal distributions of its

components P_X and P_Y, i.e., $(P_X \otimes P_Y)(x, y) = P_X(x)P_Y(y)$ for all elements x and y from their respective ranges $\mathcal{X}$ and $\mathcal{Y}$; $X - Y - Z$ denotes a Markov chain of random variables X, Y, and Z in this order; $(\cdot)^c$ denotes the complement of a set; $\mathfrak{P}_L(\mathcal{M})$ is the set of all subsets of $\mathcal{M}$ with cardinality exactly L; $\hat{\mathfrak{P}}_L(\mathcal{M})$ is the set of all subsets of $\mathcal{M}$ with cardinality at most L; $f(n) \in o(g(n))$ if $\lim_{n\to\infty} |\frac{g(n)}{f(n)}| = 0$; $f(n) \in \mathcal{O}(g(n))$ if $\lim_{n\to\infty} |\frac{g(n)}{f(n)}| < \infty$; lhs := rhs means the value of the right hand side (rhs) is assigned to the left hand side (lhs); lhs =: rhs is defined accordingly.

II. PRELIMINARIES

In this paper we extensively make use of the concept of *types* from Csiszár and Körner [17], which is briefly reviewed in the following. Let $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$ be a sequence of length n and $N(a|x^n)$ be the number of indices i such that $x_i = a$, $i = 1, \ldots, n$. Then the *type* of a sequence $x^n \in \mathcal{X}^n$ is the distribution $P_{x^n} \in \mathcal{P}(\mathcal{X})$ defined by $P_{x^n}(a) := \frac{1}{n}N(a|x^n)$ for every $a \in \mathcal{X}$. 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 $(x_i, y_i) = (a, b)$, $i = 1, \ldots, n$.

As in [17] we represent, for notational convenience, (joint) types of sequences of length n by (joint) distributions of dummy random variables. We suppress the dependency on the length n when it is clear from the context. For instance, the random variables X and Y represent a joint type, i.e., $P_{XY} = P_{x^n, y^n}$ 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 $\tau_X = \{x^n : x^n \in \mathcal{X}^n, P_{x^n} = P_X\}$. Of course, this notation extends to joint types and sections in a self-explanatory way, e.g., $\tau_{XY} = \{(x^n, y^n) : x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n, P_{x^n, y^n} = P_{XY}\}$ or $\tau_{Y|X}(x^n) = \{y^n : (x^n, y^n) \in \tau_{XY}\}$.

Now let X, Y and X', Y' be two pairs of random variables on $\mathcal{X} \times \mathcal{Y}$ with probability mass functions $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $P_{X'Y'} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, respectively. For our proof we will need the following properties of the (Kullback-Leibler) information divergence [17, p. 7]

$$D(P_{XY} \| P_{X'Y'}) = D(P_X \| P_{X'}) + D(P_{Y|X} \| P_{Y'|X'} | P_X) \quad (1)$$

$$I(X; Y) = D(P_{XY} \| P_X \otimes P_Y) D(P_{Y|X} \| P_Y | P_X). \quad (2)$$

Next, we state as facts some few basic bounds on types which we will need for our proofs, see [17, Sec. 1.2].

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

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

Fact 2: We have

$$(n + 1)^{-|\mathcal{X}|} 2^{nH(X)} \le |\tau_X| \le 2^{nH(X)}, \quad \text{if } \tau_X \ne \emptyset$$

$$(n + 1)^{-|\mathcal{X}||\mathcal{Y}|} 2^{nH(Y|X)} \le |\tau_{Y|X}(x^n)|$$

$$\le 2^{nH(Y|X)}, \quad \text{if } \tau_{Y|X}(x^n) \ne \emptyset$$

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

$$\sum_{y^n \in \tau_{Y|X}(x^n)} W^{\otimes n}(y^n|x^n) = W^{\otimes n}(\tau_{Y|X}(x^n)|x^n)$$

$$\le 2^{-nD(P_{XY}\|P_X \otimes W)}$$

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 $P_X(x)W(y|x)$. Further, see also [21], for some $s^n \in \mathcal{S}^n$,

$$\sum_{y^n \in \tau_{Y|XS}(x^n, s^n)} W^{\otimes n}(y^n|x^n) = W^{\otimes n}(\tau_{Y|XS}(x^n, s^n)|x^n)$$

$$\le 2^{-nI(Y; S|X)}$$

which follows immediately from (1) and (2) by

$$D(P_{XSY} \| P_{XS} \otimes W) = D(P_{Y|XS} \| W | P_{XS})$$
$$= I(Y; S|X) + D(P_{Y|X} \| W | P_X)$$
$$\ge I(Y; S|X).$$

III. BROADCAST CHANNEL WITH RECEIVER SIDE INFORMATION

In the following we study the broadcast phase of the decode-and-forward bidirectional relaying protocol, see Fig. 1. Here, we assume that the relay has successfully decoded the two messages the two nodes have sent in the previous multiple access phase. Now, the relay broadcasts an optimal re-encoded message based on the network coding idea so that both nodes can decode the intended message using their own message from the previous phase as side information.

A. Arbitrarily Varying Broadcast Channel With Receiver Side Information

The transmission is affected by arbitrarily varying channels, which is modeled 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 $W^{\otimes n}(y_1^n, y_2^n|x^n, s^n) := \prod_{k=1}^n W(y_{1,k}, y_{2,k}|x_k, s_k)$.

Definition 1: The discrete memoryless *arbitrarily varying broadcast channel (AVBC)* $\mathcal{W}$ is the family

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

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

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

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

For the following analysis we need a concept of symmetrizability which distinguishes among different degrees of symmetry. In more detail, we say a channel $\widetilde{W}_i(y_i|x_1, \ldots, x_t)$

with input alphabet $\mathcal{X}^t$ and output alphabet $\mathcal{Y}_i$ is symmetric if for every permutation π on $\{1, 2, \ldots, n\}$ we have $\widetilde{W}_i(y_i|x_1, \ldots, x_t) = \widetilde{W}_i(y_i|x_{\pi(1)}, \ldots, x_{\pi(t)})$ for all $x_1, \ldots, x_t \in \mathcal{X}$ and $y_i \in \mathcal{Y}_i$. This leads to the following definition.

Definition 2: For any $t_i \geq 1$, $i = 1, 2$, an AVBC is $(\mathcal{Y}_i, t_i)$-*symmetrizable* if there is a channel $U : \mathcal{X}^{t_i} \rightarrow \mathcal{P}(\mathcal{S})$ such that

$$\widetilde{W}_i(y_i|x_0, x_1, \ldots, x_{t_i}) := \sum_{s \in \mathcal{S}} W_i(y_i|x_0, s)U(s|x_1, \ldots, x_{t_i}) \quad (4)$$

is symmetric in $x_0, x_1, \ldots, x_{t_i}$ for all $x_0, x_1, \ldots, x_{t_i} \in \mathcal{X}$ and $y_i \in \mathcal{Y}_i$. For convenience, we take all AVBCs to be $(\mathcal{Y}_i, 0)$-symmetrizable, $i = 1, 2$.

Intuitively, a $(\mathcal{Y}_i, t_i)$-symmetrizable channel can be interpreted as a channel where the state sequence can emulate t_i replicas of the channel input. Further, from the definition it is clear that if an AVBC is $(\mathcal{Y}_i, t_i)$-symmetrizable, then it is also $(\mathcal{Y}_i, t_i')$-symmetrizable for all $0 \leq t_i' \leq t_i$, $i = 1, 2$.

Definition 3: The *symmetrizability* of an AVBC is defined by the largest integers t_1 and t_2 such that the channel is $(\mathcal{Y}_1, t_1)$-symmetrizable and $(\mathcal{Y}_2, t_2)$-symmetrizable. This pair of largest integers is denoted by (T_1, T_2).

Remark 1: The concept of symmetrizability for the AVBC introduced above is a natural extension of the one proposed for the single-user AVC under list decoding in [20] and [21]. Additionally, we call a $(\mathcal{Y}_i, 1)$-symmetrizable channel in the sense of Definition 2 also a $\mathcal{Y}_i$-symmetrizable channel according to the terminology used for the single-user AVC in [18] and [19], which does not distinguish among different degrees of symmetry.

Remark 2: The symmetrizability presented above consists of two symmetrizability conditions, i.e., one for each marginal channel. This goes along with intuition that capacity results for broadcast channels only depend on the marginal channels. In addition, this is in contrast to the AVMAC, where necessarily a third, joint symmetrizability condition comes into play [30], [31].

Remark 3: Based on the observation that the symmetrizability depends only on the marginal channels, it is straightforward to construct an example for the AVBC with two different symmetrizabilities based on the example given in [21, Sec. IV].

B. List Codes

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

Definition 4: A deterministic $(n, M_{1,n}, M_{2,n}, L_1, L_2)$-*list code* $\mathcal{C}_{\mathcal{W}, \text{list}}$ of length n with list sizes (L_1, L_2) for the AVBC $\mathcal{W}$ with receiver side information consists of codewords

$$x_m^n \in \mathcal{X}^n,$$

one for each message $m = (m_1, m_2)$, and list decoders at nodes 1 and 2

$$\mathcal{L}^{(1)} : \mathcal{Y}_1^n \times \mathcal{M}_1 \rightarrow \hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2),$$
$$\mathcal{L}^{(2)} : \mathcal{Y}_2^n \times \mathcal{M}_2 \rightarrow \hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$$

where $\hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$ is the set of all subsets of $\mathcal{M}_2$ with cardinality at most L_1 and similarly $\hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$ is the set of all subsets of $\mathcal{M}_1$ with cardinality at most L_2. This means the list decoder $\mathcal{L}^{(i)}$ maps the received output y_i^n and its own message m_i into a list of at most L_i messages, $i = 1, 2$.

When x_m^n with $m = (m_1, m_2)$ has been sent, and y_1^n and y_2^n have been received at nodes 1 and 2, the list decoder at node 1 is in error if m_2 is not in $\mathcal{L}^{(1)}(y_1^n, m_1)$. Accordingly, the list decoder at node 2 is in error if m_1 is not in $\mathcal{L}^{(2)}(y_2^n, m_2)$. This allows us to define the probabilities of error for given message $m = (m_1, m_2)$ and given state sequence $s^n \in \mathcal{S}^n$ as

$$e(m, s^n|\mathcal{C}_{\mathcal{W}, \text{list}}) := \sum_{\substack{(y_1^n, y_2^n):m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1) \\ \vee m_1 \notin \mathcal{L}^{(2)}(y_2^n, m_2)}} W^{\otimes n}(y_1^n, y_2^n|x_m^n, s^n) \quad (5)$$

and the corresponding marginal probabilities of error at nodes 1 and 2 by $e_1(m, s^n|\mathcal{C}_{\mathcal{W}, \text{list}}) := \sum_{y_1^n:m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} W^{\otimes n}(y_1^n|x_m^n, s^n)$ and $e_2(m, s^n|\mathcal{C}_{\mathcal{W}, \text{list}}) := \sum_{y_2^n:m_1 \notin \mathcal{L}^{(2)}(y_2^n, m_2)} W^{\otimes n}(y_2^n|x_m^n, s^n)$, respectively. Thus, the *average probability of error* for state sequence $s^n \in \mathcal{S}^n$ is given by

$$\bar{e}(s^n|\mathcal{C}_{\mathcal{W}, \text{list}}) := \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e(m, s^n|\mathcal{C}_{\mathcal{W}, \text{list}}) \quad (6)$$

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

For given $0 < \lambda^{(n)} < 1$, $\mathcal{C}_{\mathcal{W}, \text{list}}$ is called a $(n, M_{1,n}, M_{2,n}, L_1, L_2, \lambda^{(n)})$-list code (with average probability of error $\lambda^{(n)}$) for the AVBC $\mathcal{W}$ with receiver side information if

$$\bar{e}(s^n|\mathcal{C}_{\mathcal{W}, \text{list}}) \leq \lambda^{(n)} \quad \text{for all } s^n \in \mathcal{S}^n.$$

Definition 5: A rate pair $(R_1, R_2) \in \mathbb{R}_+^2$ is said to be *list achievable* with list sizes L_1, L_2 for the AVBC $\mathcal{W}$ with receiver side information if for any $\delta > 0$ there exists an $n(\delta) \in \mathbb{N}$ and a sequence $(\mathcal{C}_{\mathcal{W}, \text{list}}^{(n)})_{n \in \mathbb{N}}$ of deterministic $(n, M_{1,n}, M_{2,n}, L_1, L_2, \lambda^{(n)})$-list codes such that for all $n \geq n(\delta)$ we have[1]

$$\frac{1}{n} \log\left(\frac{M_{2,n}}{L_1}\right) \geq R_1 - \delta \quad \text{and} \quad \frac{1}{n} \log\left(\frac{M_{1,n}}{L_2}\right) \geq R_2 - \delta$$

while

$$\max_{s^n \in \mathcal{S}^n} \bar{e}(s^n|\mathcal{C}_{\mathcal{W}, \text{list}}) \leq \lambda^{(n)}$$

with $\lambda^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. The set of all achievable rate pairs with list sizes (L_1, L_2) is the *list capacity region* of the

---

[1] The rate between the relay and receiving node i with list size L_i is denoted by R_i, $i = 1, 2$. However, the message associated to rate R_1 is denoted by $M_{2,n}$ which is due to the fact that it originates from node 2 and, thus, looks "swapped." The same applies to message $M_{1,n}$ at rate R_2.

AVBC $\mathcal{W}$ with receiver side information and is denoted by $\mathcal{R}_{\mathcal{W},\mathrm{list}}(L_1, L_2)$.

Remark 4: 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 for all possible state sequences simultaneously. This means the codes are universal with respect to the state sequence.

Remark 5: For list sizes $L_1 = L_2 = 1$ the list code $\mathcal{C}_{\mathcal{W},\mathrm{list}}$ as given in Definition 4 reduces to a usual deterministic $(n, M_{1,n}, M_{2,n})$-code $\mathcal{C}_{\mathcal{W},\mathrm{det}}$ where each decoder maps its received sequence into exactly one message. The definitions of a *deterministically achievable* rate pair and the *deterministic code capacity region* $\mathcal{R}_{\mathcal{W},\mathrm{det}}$ follow accordingly.

C. Outline

In the Section IV we prove the random code capacity region of the AVBC with receiver side information, where we allow the relay and the receivers to coordinate their choice of encoder and decoders. It will be convenient to use this result to establish the desired list capacity region, which is done in the following Sections V-VII. Therefore, in Section V we use the random code capacity result to establish the general behavior of the list capacity region similarly to Ahlswede's dichotomy result for the single-user AVC. Then, in Section VI and VII we characterize in detail when the list capacity region equals the random code capacity region and when it has an empty interior.

IV. RANDOM CODE CAPACITY REGION

In this section, we restrict the list sizes at the receiving nodes to one and derive the optimal random coding strategy for the AVBC with receiver side information. Thereby, the word "random" refers to the fact that the encoder and decoders are chosen according to a common random experiment whose outcome has to be known at all nodes in advance. This leads directly to the following definition.

Definition 6: A *random* $(n, M_{1,n}, M_{2,n}, Z, \mu)$-code $\mathcal{C}_{\mathcal{W},\mathrm{ran}}$ of length n for the AVBC $\mathcal{W}$ with receiver side information is given by a family $\mathcal{C}_{\mathcal{W},\mathrm{ran}} := \{\mathcal{C}(z) : z \in \mathcal{Z}\}$ of deterministic $(n, M_{1,n}, M_{2,n})$-codes

$$\mathcal{C}(z) := \big\{ \big(x_m^n(z), D_{m_2|m_1}^{(1)}(z), D_{m_1|m_2}^{(2)}(z) \big) : \\ m_1 \in M_{1,n}, m_2 \in M_{2,n} \big\}$$

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

Here, it will be convenient to use the notion of decoding sets to specify the decoding rule. This means that the decoding sets at nodes 1 and 2 of one deterministic code $\mathcal{C}(z)$, $z \in \mathcal{Z}$, are given by $D_{m_2|m_1}^{(1)}(z) \subseteq \mathcal{Y}_1^n$ and $D_{m_1|m_2}^{(2)}(z) \subseteq \mathcal{Y}_2^n$ for all $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$. Since $\mathcal{C}(z)$ is a deterministic code (with list sizes one), the decoding sets must be disjoint. In more detail, for given m_1 at node 1 the decoding sets must satisfy $D_{m_2|m_1}^{(1)}(z) \cap D_{m_2'|m_1}^{(1)}(z) = \emptyset$ for $m_2' \neq m_2$, and similarly for given m_2 at node 2 the decoding sets must satisfy $D_{m_1|m_2}^{(2)}(z) \cap D_{m_1'|m_2}^{(2)}(z) = \emptyset$ for $m_1' \neq m_1$.

The average probability of error of the deterministic code $\mathcal{C}(z)$, $z \in \mathcal{Z}$, for state sequence $s^n \in \mathcal{S}^n$ can be written as

$$\bar{e}(s^n | \mathcal{C}(z)) \\ := \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W^{\otimes n} \big((D_{m_2|m_1}^{(1)}(z) \times D_{m_1|m_2}^{(2)}(z))^c | x_m^n(z), s^n \big)$$

and the corresponding marginal probabilities of error at nodes 1 and 2 as $\bar{e}_1(s^n | \mathcal{C}(z)) := \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W_1^{\otimes n} ((D_{m_2|m_1}^{(1)}(z))^c | x_m^n(z), s^n)$ and $\bar{e}_2(s^n | \mathcal{C}(z)) := \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W_2^{\otimes n} ((D_{m_1|m_2}^{(2)}(z))^c | x_m^n(z), s^n)$. Then, the average probability of error of the random code $\mathcal{C}_{\mathcal{W},\mathrm{ran}}$ for state sequence $s^n \in \mathcal{S}^n$ is given by

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

and accordingly the corresponding marginal average probability of error at node i by $\bar{e}_i(s^n | \mathcal{C}_{\mathcal{W},\mathrm{ran}}) := \mathbb{E}_Z[\bar{e}_i(s^n | \mathcal{C}(Z))]$, $i = 1, 2$. For given $0 < \lambda^{(n)} < 1$, $\mathcal{C}_{\mathcal{W},\mathrm{ran}}$ is called a $(n, M_{1,n}, M_{2,n}, Z, \mu, \lambda^{(n)})$-code (with average probability of error $\lambda^{(n)}$) for the AVBC $\mathcal{W}$ with receiver side information if

$$\bar{e}(s^n | \mathcal{C}_{\mathcal{W},\mathrm{ran}}) \leq \lambda^{(n)} \qquad \text{for all } s^n \in \mathcal{S}^n.$$

Then the definitions of a *randomly achievable* rate pair and the *random code capacity region* $\mathcal{R}_{\mathcal{W},\mathrm{ran}}$ follow accordingly.

Remark 6: From the definitions of the codes it is clear that the deterministic code $\mathcal{C}_{\mathcal{W},\mathrm{det}}$ is a special or degenerated case of the random code $\mathcal{C}_{\mathcal{W},\mathrm{ran}}$. More precisely, $\mathcal{C}_{\mathcal{W},\mathrm{det}}$ can be interpreted as a random code that consists of only one deterministic code. Consequently, the deterministic code capacity region $\mathcal{R}_{\mathcal{W},\mathrm{det}}$ is contained in the random code capacity region $\mathcal{R}_{\mathcal{W},\mathrm{ran}}$, i.e., $\mathcal{R}_{\mathcal{W},\mathrm{det}} \subseteq \mathcal{R}_{\mathcal{W},\mathrm{ran}}$.

Theorem 1: The random code capacity region $\mathcal{R}_{\mathcal{W},\mathrm{ran}}$ of the AVBC $\mathcal{W}$ with receiver side information is given by

$$\mathcal{R}_{\mathcal{W},\mathrm{ran}} = \mathcal{R}_{\overline{\mathcal{W}}}$$

where

$$\mathcal{R}_{\overline{\mathcal{W}}} := \bigcup_{P_X \in \mathcal{P}(\mathcal{X})} \big\{ (R_1, R_2) \in \mathbb{R}_+^2 :$$

$$R_1 \leq \inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{1,q}) \tag{7a}$$

$$R_2 \leq \inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{2,q}) \big\} \tag{7b}$$

with $\overline{Y}_{i,q}$ is the random variable associated with the output of the averaged channel $\overline{W}_{i,q}$, see also (3).

In the following subsections we give the proof of the random code capacity region which is mainly based on Ahlswede's *robustification technique* [35], [36].

A. Compound Broadcast Channel With Receiver Side Information

The first key idea is to exploit results from the compound BC with receiver side information. Therefore, we construct a suitable compound broadcast channel by defining the convex hull of all averaged broadcast channels, see (3), as

$$\overline{\mathcal{W}} := \big\{ \overline{W}_q(y_1, y_2 | x) \big\}_{q \in \mathcal{P}(\mathcal{S})}.$$

We observe that $\overline{\mathcal{W}}$ already corresponds to a compound broadcast channel where each probability distribution $q \in \mathcal{P}(\mathcal{S})$ parametrizes one element of the compound channel $\overline{\mathcal{W}}$. The capacity region of the compound BC with receiver side information is known and can be found in [25] which is restated in the following theorem.

Theorem 2 ([25]): The deterministic code capacity region $\mathcal{R}_{\overline{\mathcal{W}},\det}$ of the compound BC $\overline{\mathcal{W}}$ with receiver side information is given by

$$\mathcal{R}_{\overline{\mathcal{W}},\det} = \mathcal{R}_{\overline{\mathcal{W}}},$$

see also (7). ∎

The achievability of the rates specified by $\mathcal{R}_{\overline{\mathcal{W}}}$, see (7a) and (7b), is proved by showing the existence of a deterministic $(n, M_{1,n}, M_{2,n})$-code $\mathcal{C}_{\overline{\mathcal{W}},\det}$ for the compound BC $\overline{\mathcal{W}}$ with receiver side information with arbitrarily small average probability of error. In more detail, in [25] it is shown that the average probability of error of $\mathcal{C}_{\overline{\mathcal{W}},\det}$ can be bounded from above by

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W_q^{\otimes n}\big((D_{m_2|m_1}^{(1)} \times D_{m_1|m_2}^{(2)})^c | x_m^n\big) \leq \lambda_{\overline{\mathcal{W}}}^{(n)}$$

$$\text{for all } q \in \mathcal{P}(\mathcal{S})$$

with $\lambda_{\overline{\mathcal{W}}}^{(n)} = \lambda_{\overline{\mathcal{W}},1}^{(n)} + \lambda_{\overline{\mathcal{W}},2}^{(n)}$ where $\lambda_{\overline{\mathcal{W}},i}^{(n)}$ is an upper bound on the marginal average probability of error at node i, $i = 1, 2$. More precisely, for n large enough $\lambda_{\overline{\mathcal{W}},i}^{(n)}$ is given by

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

where δ, τ, and c are positive constants, see also [25]. Note that $\lambda_{\overline{\mathcal{W}},i}^{(n)}$ decreases exponentially fast for increasing block length n.

Together with the definition of the averaged broadcast channel (3) this immediately implies that for $\mathcal{C}_{\overline{\mathcal{W}},\det}$ the probability of a successful transmission over the compound BC $\overline{\mathcal{W}}$ with receiver side information is bounded from below by

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W_q^{\otimes n}\big(D_{m_2|m_1}^{(1)} \times D_{m_1|m_2}^{(2)} | x_m^n\big) > 1 - \lambda_{\overline{\mathcal{W}}}^{(n)}$$

or equivalently by

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{s^n \in \mathcal{S}^n} W^{\otimes n}\big(D_{m_2|m_1}^{(1)} \times D_{m_1|m_2}^{(2)} | x_m^n, s^n\big) q^{\otimes n}(s^n)$$

$$> 1 - \lambda_{\overline{\mathcal{W}}}^{(n)} \quad (8)$$

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

B. Robustification

Next, we follow [35], [36] and use the deterministic code $\mathcal{C}_{\overline{\mathcal{W}},\det}$ for the compound BC $\overline{\mathcal{W}}$ with receiver side information to construct a random code $\mathcal{C}_{\mathcal{W},\text{ran}}$ for the AVBC $\mathcal{W}$ with receiver side information.

Let Π_n be the group of permutations acting on $\{1, 2, \ldots, n\}$. For given sequence $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$ and permutation $\pi \in \Pi_n : \mathcal{S}^n \to \mathcal{S}^n$ we denote the permuted sequence

$(s_{\pi(1)}, \ldots, s_{\pi(n)}) \in \mathcal{S}^n$ by $\pi(s^n)$. Further, we denote the inverse permutation by π^{-1} so that $\pi^{-1}(\pi(s^n)) = s^n$.

Theorem 3 (Robustification Technique [36]): Let $f : \mathcal{S}^n \to [0, 1]$ be a function such that for some $\alpha \in (0, 1)$ the inequality

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

is satisfied. Then the inequality

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

is also satisfied. ∎

Next, we set

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

$$(10)$$

With π being the identity map, (10) becomes $f(s^n) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} W^{\otimes n}(D_{m_2|m_1}^{(1)} \times D_{m_1|m_2}^{(2)} | x_m^n, s^n)$. Now, with this and $\alpha = \lambda_{\overline{\mathcal{W}}}^{(n)}$, the condition (9) of Theorem 3 is fulfilled by (8). Thus, the robustification technique immediately yields a random $(n, M_{1,n}, M_{2,n}, \Pi_n, \mu)$-code $\mathcal{C}_{\mathcal{W},\text{ran}}$ for the AVBC $\mathcal{W}$ with receiver side information given by the family

$$\mathcal{C}_{\mathcal{W},\text{ran}} = \big\{(\pi^{-1}(x_m^n), \pi^{-1}(D_{m_2|m_1}^{(1)}), \pi^{-1}(D_{m_1|m_2}^{(2)})) :$$
$$m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2, \pi \in \Pi_n, \mu\big\} \quad (11)$$

with μ the uniform distribution on Π_n and

$$\pi^{-1}(D_{m_2|m_1}^{(1)}) = \bigcup_{y_1^n \in D_{m_2|m_1}^{(1)}} \pi^{-1}(y_1^n),$$

$$\pi^{-1}(D_{m_1|m_2}^{(2)}) = \bigcup_{y_2^n \in D_{m_1|m_2}^{(2)}} \pi^{-1}(y_2^n).$$

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

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

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

$$(12)$$

The way how we constructed the random code $\mathcal{C}_{\mathcal{W},\text{ran}}$ from the deterministic code $\mathcal{C}_{\overline{\mathcal{W}},\det}$ has the following consequence. All rate pairs achievable for compound BC $\overline{\mathcal{W}}$ with receiver side information using the deterministic code $\mathcal{C}_{\overline{\mathcal{W}},\det}$ are also achievable for the AVBC $\mathcal{W}$ with receiver side information using the random code $\mathcal{C}_{\mathcal{W},\text{ran}}$. Consequently, the random code $\mathcal{C}_{\mathcal{W},\text{ran}}$ actually achieves all rate pairs satisfying (7a) and (7b) as stated in Theorem 1, which proves the achievability.

C. Converse

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

As a first step it is easy to show that the average probability of error of the random code $\mathcal{C}_{\mathcal{W},\text{ran}}$ for the AVBC $\mathcal{W}$ with receiver side information equals the average probability of error of the random code for the AVBC with receiver side information obtained by taking the convex closure, i.e. considering channels $\overline{W}_{q^n}(y_1^n, y_2^n|x^n) := \sum_{s^n \in \mathcal{S}^n} W^{\otimes n}(y_1^n, y_2^n|x^n, s^n)q^n(s^n)$ for all possible $q^n = \prod_{k=1}^{n} q_k$. Hence, the corresponding achievable rates are the same, see also [17, Lemma 12.3] and [17, Corollary 12.3] for the corresponding single-user result. Now, it is clear that restricting q^n to be constant for the whole block length, i.e., $q^n = \prod_{k=1}^{n} q$, can only increase the achievable rates. But this corresponds to the compound BC $\overline{\mathcal{W}}$ with receiver side information. Hence, it is clear that for AVBC $\mathcal{W}$ with receiver side information we cannot achieve higher rates as for the constructed compound BC $\overline{\mathcal{W}}$ with receiver side information with random codes. The deterministic rates of the compound channel are given in Theorem 2. As in [37] for the single-user compound channel, it can be easily shown that for the compound BC $\overline{\mathcal{W}}$ with receiver side information the achievable rates for deterministic and random codes are equal. Since the constructed random code $\mathcal{C}_{\mathcal{W},\text{ran}}$ for the AVBC $\mathcal{W}$ with receiver side information already achieves these rates, the converse is proved.

This finishes the proof of Theorem 1 and therewith establishes the random code capacity region $\mathcal{R}_{\mathcal{W},\text{ran}}$ of the AVBC $\mathcal{W}$ with receiver side information. ∎

V. LIST CAPACITY REGION

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, see (11), which has to be known to all nodes in advance. If this kind of resource is not available, one is interested in deterministic (list) strategies.

Theorem 4: The list capacity region $\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)$ with list sizes (L_1, L_2) of the AVBC $\mathcal{W}$ with receiver side information with symmetrizability (T_1, T_2) is

$$\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2) = \mathcal{R}_{\overline{\mathcal{W}}} \quad \text{if } L_1 > T_1 \text{ and } L_2 > T_2.$$

If $L_i \leq T_i$, then we have $R_i = 0$, $i = 1, 2$. This means we have interior$(\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)) = \emptyset$ if and only if $L_1 \leq T_1$ or $L_2 \leq T_2$.

The result shows that every AVBC with receiver side information has a characteristic pair of minimum list sizes (T_1+1, T_2+1) that enables a successful bidirectional exchange of both individual messages in the bidirectional relay network at all rate pairs $(R_1, R_2) \in \mathcal{R}_{\overline{\mathcal{W}}}$. On the other hand, if $L_i \leq T_i$, then we have for the corresponding rate $R_i = 0$, $i = 1, 2$. From this we can deduce the two cases where one list size is greater than the symmetrizability and the other one smaller or equal.

Corollary 1: If $L_1 > T_1$ and $L_2 \leq T_2$, then the list capacity region $\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)$ is given by

$$R_1 \leq \max_{P_X \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{S})} I(X; \overline{Y}_{1,q}), \qquad R_2 = 0.$$

If $L_1 \leq T_1$ and $L_2 > T_2$, then the list capacity region $\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)$ is given by

$$R_1 = 0, \qquad R_2 \leq \max_{P_X \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{S})} I(X; \overline{Y}_{2,q}).$$

In addition, we immediately obtain the deterministic code capacity region $\mathcal{R}_{\mathcal{W},\text{det}}$ if we restrict the list sizes at the receiving nodes to one.

Corollary 2: For a non-$\mathcal{Y}_1$-symmetrizable and non-$\mathcal{Y}_2$-symmetrizable AVBC $\mathcal{W}$ with receiver side information the deterministic code capacity region $\mathcal{R}_{\mathcal{W},\text{det}}$ is given by

$$\mathcal{R}_{\mathcal{W},\text{det}} = \mathcal{R}_{\overline{\mathcal{W}}}.$$

We have interior$(\mathcal{R}_{\mathcal{W},\text{det}}) = \emptyset$ if and only if the AVBC $\mathcal{W}$ with receiver side information is $\mathcal{Y}_1$-symmetrizable or $\mathcal{Y}_2$-symmetrizable.

In the following we give a characterization of the list capacity region which is mainly based on Ahlswede's *elimination technique* [16]. However, the approach in [16] has to be carefully extended to work for list codes for the BC with receiver side information, where each receiving node has a list decoder of different list size.

A. Random Code Reduction

The first step of the elimination technique [16] is the *random code reduction*. Here, we construct a new random code by selecting a relatively small number of deterministic codes from the original random code using the following lemma suitable for the BC with receiver side information.

Lemma 1 (Random Code Reduction): As given in (11) let $\mathcal{C}_{\mathcal{W},\text{ran}}$ be a random code for the AVBC $\mathcal{W}$ with receiver side information and let $\lambda_{\mathcal{W},\text{ran}}^{(n)}$ be an upper bound on the average probability of error of this code as specified in (12). For any ϵ and K^2 that satisfy

$$\epsilon > 2\lambda_{\mathcal{W},\text{ran}}^{(n)} \qquad \text{and} \qquad K^2 > \frac{2}{\epsilon} \log(|\mathcal{S}|^n) \qquad (13)$$

there exist K^2 deterministic codes $\mathcal{C}_i$, $i = 1, \ldots, K^2$ such that

$$\frac{1}{K^2} \sum_{i=1}^{K^2} \bar{e}(s^n|\mathcal{C}_i) < \epsilon \qquad \text{for all } s^n \in \mathcal{S}^n. \qquad (14)$$

Proof: A random code reduction for the single-user AVC was first proposed in [16]. Our proof for the AVBC with receiver side information is inspired by [17, Lemma 12.8] where a similar result for the single-user AVC in terms of maximal probability of error is established.

First, from the random code $\mathcal{C}_{\mathcal{W},\text{ran}}$ we select K^2 independent permutations $\pi_i \in \Pi_n$, $i = 1, \ldots, K^2$ according to the uniform distribution μ, see (11). Each such permutation π_i specifies one deterministic code and is denoted by $\mathcal{C}_i$ in the following. Then, for given state sequence $s^n \in \mathcal{S}^n$, we have

$$\mathbb{P}\left\{\frac{1}{K^2} \sum_i \bar{e}(s^n|\mathcal{C}_i) \geq \epsilon\right\} = \mathbb{P}\left\{2^{\sum_i \bar{e}(s^n|\mathcal{C}_i)} \geq 2^{K^2\epsilon}\right\}$$

$$\leq 2^{-K^2\epsilon} \mathbb{E}\left[2^{\sum_i \bar{e}(s^n|\mathcal{C}_i)}\right] \qquad (15)$$

where the last step follows from Markov's inequality. Since the random variables $\mathcal{C}_i$, $i = 1, \ldots, K^2$ are independent and identically distributed, we get for the expectation

$$\mathbb{E}\Big[2^{\sum_i \bar{e}(s^n|\mathcal{C}_i)}\Big] = \mathbb{E}\Big[2^{\bar{e}(s^n|\mathcal{C}_1)}\Big]^{K^2}$$
$$\leq \big(1 + \mathbb{E}[\bar{e}(s^n|\mathcal{C}_1)]\big)^{K^2}$$
$$\leq \big(1 + \lambda_{\mathcal{W},\mathrm{ran}}^{(n)}\big)^{K^2}$$

where we further used the inequality $2^x \leq 1 + x$, $0 \leq x \leq 1$. This and (15) yield

$$\mathbb{P}\Big\{\frac{1}{K^2}\sum_i \bar{e}(s^n|\mathcal{C}_i) \geq \epsilon\Big\} \leq 2^{-K^2\epsilon}\big(1 + \lambda_{\mathcal{W},\mathrm{ran}}^{(n)}\big)^{K^2}$$
$$\leq 2^{-K^2(\epsilon - \lambda_{\mathcal{W},\mathrm{ran}}^{(n)})}.$$

Finally, with (13) this implies

$$\mathbb{P}\Big\{\frac{1}{K^2}\sum_i \bar{e}(s^n|\mathcal{C}_i) \geq \epsilon \quad \text{for all } s^n \in \mathcal{S}^n\Big\}$$
$$\leq |\mathcal{S}|^n 2^{-K^2(\epsilon - \lambda_{\mathcal{W},\mathrm{ran}}^{(n)})}. \quad (16)$$

This means that there exist a collection of deterministic codes $\mathcal{C}_i$, $i = 1, \ldots, K^2$ which satisfy (14) proving the lemma. ∎

The random code reduction shows that for any random code there exists another "reduced" random code which is uniformly distributed over K^2 deterministic codes with an average probability of error less than ϵ under the assumption that (13) holds.

A direct consequence of Lemma 1 is that for any random code $\mathcal{C}_{\mathcal{W},\mathrm{ran}}$ which achieves the random code capacity of the AVBC $\mathcal{W}$ with receiver side information, there exists another "reduced" random code $\tilde{\mathcal{C}}_{\mathcal{W},\mathrm{ran}}$ which does likewise. Furthermore, from [16] we know that it is sufficient to select no more than $K^2 = n^2$ deterministic codes to obtain $\tilde{\mathcal{C}}_{\mathcal{W},\mathrm{ran}}$ with the desired properties. This can easily be seen in (16) where the choice $K^2 = n^2$ leads to a code whose probability of error exceeds ϵ with a super exponentially small probability since $|\mathcal{S}|^n$ grows exponentially in n only.

In more detail, for any $\epsilon > 0$ and sufficiently large n there exist n^2 deterministic codes

$$\mathcal{C}_i := \big\{\big(\pi_i^{-1}(x_m^n), \pi_i^{-1}(D_{m_2|m_1}^{(1)}), \pi_i^{-1}(D_{m_1|m_2}^{(2)})\big) : $$
$$m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2\big\} \in \mathcal{C}_{\mathcal{W},\mathrm{ran}},$$

$i = 1, \ldots, n^2$, such that

$$\bar{e}(s^n|\tilde{\mathcal{C}}_{\mathcal{W},\mathrm{ran}}) = \frac{1}{n^2}\sum_i \bar{e}(s^n|\mathcal{C}_i) < \epsilon =: \tilde{\lambda}_{\mathcal{W},\mathrm{ran}}^{(n)}$$
$$\text{for all } s^n \in \mathcal{S}^n. \quad (17)$$

The "reduced" random code $\tilde{\mathcal{C}}_{\mathcal{W},\mathrm{ran}}$ with "exponentially few" elements is given by

$$\tilde{\mathcal{C}}_{\mathcal{W},\mathrm{ran}} := \{\mathcal{C}_i : i = 1, \ldots, n^2\}$$

where the index i is drawn according to the uniform distribution on $\{1, \ldots, n^2\}$. Clearly, the "reduced" random code $\tilde{\mathcal{C}}_{\mathcal{W},\mathrm{ran}}$ also achieves the random code capacity of the AVBC $\mathcal{W}$ with receiver side information.
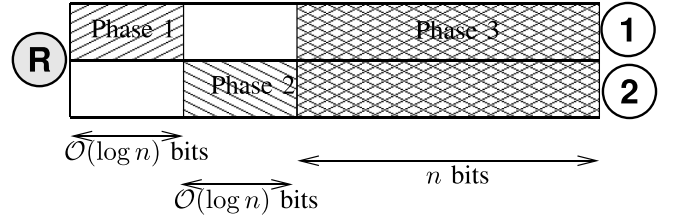

Fig. 2. Three-phase protocol that achieves the list capacity region $\mathcal{R}_{\mathcal{W},\mathrm{list}}(L_1, L_2)$ of the AVBC $\mathcal{W}$ with receiver side information. In the first two phases, the relay uses list codes for the classical point-to-point AVC to inform both receiving nodes which random code is used in the following third phase.

B. Elimination of Randomness

One way to ensure that the decoders are chosen according to the same random permutation as the encoder, is to inform the receivers which one is used by the encoder. Consequently, first the transmitter has to communicate the chosen permutation to the receivers and then transmits the message according to the randomly selected code. If the number of all possible codes could be kept small enough, the transmission of those additional information would not cause an essential loss in rate. Inspired by this idea we establish the following behavior of the list capacity region which is similar to Ahlswede's dichotomy result for the single-user AVC. In more detail, we show in the following that the list capacity region $\mathcal{R}_{\mathcal{W},\mathrm{list}}(L_1, L_2)$ for the AVBC $\mathcal{W}$ with receiver side information displays the following behavior:

$$\mathcal{R}_{\mathcal{W},\mathrm{list}}(L_1, L_2) = \mathcal{R}_{\mathcal{W},\mathrm{ran}} \quad \text{if interior}(\mathcal{R}_{\mathcal{W},\mathrm{list}}(L_1, L_2)) \neq \emptyset.$$
$$(18)$$

Up to now we have constructed a random code with "exponentially few" elements that achieves the random code capacity of the AVBC $\mathcal{W}$ with receiver side information. The next step of the elimination technique [16] is the *elimination of randomness*. This means that we convert the "reduced" random code into a list code by adding short prefixes to the original codewords to inform the decoders which of the n^2 deterministic codes is actually used [16], [17]. Clearly, this is only possible, if the list capacity region $\mathcal{R}_{\mathcal{W},\mathrm{list}}(L_1, L_2)$ fulfills interior$(\mathcal{R}_{\mathcal{W},\mathrm{list}}(L_1, L_2)) \neq \emptyset$, which means that transmission at positive rates is possible to both receivers.

We propose a three-phase protocol as depicted in Fig. 2. Prior to any transmission, the relay chooses an index $i \in \{1, \ldots, n^2\}$ uniformly at random for the random code. In the first two phases of the protocol, the relay uses then list codes with list sizes L_1 and L_2 respectively to inform both receiving nodes which $i \in \{1, \ldots, n^2\}$ was chosen. This indicates which out of the n^2 deterministic codes will be used in the third phase. After the two phases, the relay and both receiving nodes have established common randomness. Thus, in the third phase, a random code can be used where encoder and decoders are chosen according to $i \in \{1, \ldots, n^2\}$. We have to ensure that the resources spent in the first two phases to inform the receiving nodes are negligible so that there is no loss in overall rate.

Since interior($\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)$) $\neq \emptyset$, for the first to phases there exists for sufficiently large n sequences of ordinary list codes $\mathcal{C}_{\text{pre}}^1$ and $\mathcal{C}_{\text{pre}}^2$ with sequences

$$x_i^{k_n} \in \mathcal{X}^{k_n}, \qquad x_i^{l_n} \in \mathcal{X}^{l_n} \tag{19}$$

of length k_n and l_n for all $i = 1, \ldots, n^2$, respectively, and ordinary list decoders at nodes 1 and 2

$$\mathcal{L}_1^{(1)} : \mathcal{Y}_1^{k_n} \to \hat{\mathfrak{P}}_{L_1}(\{1, \ldots, n^2\})$$
$$\mathcal{L}_2^{(2)} : \mathcal{Y}_2^{l_n} \to \hat{\mathfrak{P}}_{L_2}(\{1, \ldots, n^2\})$$

and average probabilities of error satisfying

$$\max_{s^{k_n} \in \mathcal{S}^{k_n}} \frac{1}{n^2} \sum_{i \in \{1,\ldots,n^2\}} \sum_{y_1^{k_n}:i \notin \mathcal{L}_1^{(1)}(y_1^{k_n})} W_1^{k_n}(y_1^{k_n}|x_i^{k_n}, s^{k_n}) \leq \lambda_{\text{pre}}^{(k_n)} \tag{20a}$$

$$\max_{s^{l_n} \in \mathcal{S}^{l_n}} \frac{1}{n^2} \sum_{i \in \{1,\ldots,n^2\}} \sum_{y_2^{l_n}:i \notin \mathcal{L}_2^{(2)}(y_2^{l_n})} W_2^{l_n}(y_2^{l_n}|x_i^{l_n}, s^{l_n}) \leq \lambda_{\text{pre}}^{(l_n)}. \tag{20b}$$

Then to transmit the message $m = (m_1, m_2) \in \mathcal{M}$ having $i \in \{1, \ldots, n^2\}$ available, the relay transmit the concatenated codeword

$$x_m^{k_n+l_n+n}(i) = \left(x_i^{k_n}, x_i^{l_n}, \pi_i^{-1}(x_m^n)\right) \tag{21}$$

where $x_i^{k_n}$ and $x_i^{l_n}$ are codewords of the corresponding single-user list codes, see (19) to inform the receiving nodes about $i \in \{1, \ldots, n^2\}$. As the amount is polynomial in n, we have k_n, l_n of order $\mathcal{O}(\log n)$ so that $\frac{k_n}{n}, \frac{l_n}{n} \to 0$ as $n \to \infty$.

The signals $y_1^{k_n+l_n+n}$ and $y_2^{k_n+l_n+n}$ are received by nodes 1 and 2 which use list decoders of sizes L_1 and L_2 respectively to obtain a list of possible realizations $i \in \{1, \ldots, n^2\}$ from the first and second phase respectively. Based on these lists, the receiving nodes create list decoders for the message $m = (m_1, m_2) \in \mathcal{M}$ transmitted in the third phase, i.e.,

$$\mathcal{L}_3^{(1)}(y_1^n) = \{m_2 : \exists\, i \in \mathcal{L}_1^{(1)}(y_1^{k_n}) \text{ and } y_1^n \in \pi_i^{-1}(D_{m_2|m_1}^{(1)})\}$$
$$\mathcal{L}_3^{(2)}(y_2^n) = \{m_1 : \exists\, i \in \mathcal{L}_2^{(2)}(y_2^{l_n}) \text{ and } y_2^n \in \pi_i^{-1}(D_{m_1|m_2}^{(2)})\}.$$

Thus, for any $y_1^{k_n+l_n+n}$ let $i_1, i_2, \ldots, i_{L_1'}$ with $L_1' \leq L_1$ the list $\mathcal{L}_1^{(1)}(y_1^{k_n}) = \{i_1, i_2, \ldots, i_{L_1'}\}$, the constructed final list decoder

$$\mathcal{L}^{(1)}(y_1^{k_n+l_n+n}) = \left(\mathcal{L}_1^{(1)}(y_1^{k_n}), \mathcal{L}_3^{(1)}(y_1^n)\right)$$
$$= \{(i, m_2) : i \in \{i_1, \ldots, i_{L_1'}\} \text{ and }$$
$$y_1^n \in \pi_i^{-1}(D_{m_2|m_1}^{(1)})\}$$

defines a valid list decoder for receiving node 1 with list size not greater than L_1 as required in Definition 4. The list decoder $\mathcal{L}^{(2)}$ at receiving node 2 with list size not greater than L_2 is defined accordingly.

Next, we want to show that the error probabilities of the final concatenated list code $\mathcal{C}_{\mathcal{W},\text{list}}$ are small as well. We present the analysis for receiving node 1, the other follows accordingly.

We define

$$\bar{e}(s^{k_n+l_n+n}|\mathcal{C}_{\mathcal{W},\text{list}}) = \frac{1}{n^2} \frac{1}{|\mathcal{M}|} \sum_{i \in \{1,\ldots,n^2\}} \sum_{m \in \mathcal{M}}$$
$$\times \sum_{\substack{y_1^{k_n+l_n+n}:\\(i,m_2)\notin\mathcal{L}^{(1)}(y_1^{k_n+l_n+n})}} W_1^{k_n+l_n+n}\left(y_1^{k_n+l_n+n}|x_m^{k_n+l_n+n}(i), s^{k_n+l_n+n}\right) \tag{22}$$

and have to show that

$$\max_{s^{k_n+l_n+n} \in \mathcal{S}^{k_n+l_n+n}} \bar{e}(s^{k_n+l_n+n}|\mathcal{C}_{\mathcal{W},\text{list}}) \leq \lambda^{(n)}, \tag{23}$$

see Definition 5.

Now, if $(i, m_2) \notin \mathcal{L}^{(1)}(y_1^{k_n+l_n+n})$, then either a) $i \notin \mathcal{L}_1^{(1)}(y_1^{k_n})$ or b) $i \in \mathcal{L}_1^{(1)}(y_1^{k_n})$ and we have $y_1^n \notin \pi_j^{-1}(D_{m_2|m_1}^{(1)})$ for all $j \in \mathcal{L}_1^{(1)}(y_1^{k_n})$. Accordingly, we define the error events

$$\mathcal{E}_1(i) := \{y_1^{k_n+l_n+n} : i \notin \mathcal{L}_1^{(1)}(y_1^{k_n})\}$$
$$\mathcal{E}_2(i,m_2) := \{y_1^{k_n+l_n+n} : i \in \mathcal{L}_1^{(1)}(y_1^{k_n}) \text{ and } \forall\, j \in \mathcal{L}_1^{(1)}(y_1^{k_n})$$
$$\text{we have } y_1^n \notin \pi_j^{-1}(D_{m_2|m_1}^{(1)})\}$$

so that

$$\{y_1^{k_n+l_n+n} : (i, m_2) \notin \mathcal{L}^{(1)}(y_1^{k_n+l_n+n})\} \subset \mathcal{E}_1(i) \cup \mathcal{E}_2(i, m_2).$$

With this, the average probability of error in (22) can be bounded from above by

$$\bar{e}(s^{k_n+l_n+n}|\mathcal{C}_{\mathcal{W},\text{list}})$$
$$\leq \frac{1}{n^2} \frac{1}{|\mathcal{M}|} \sum_{i \in \{1,\ldots,n^2\}} \sum_{m \in \mathcal{M}}$$
$$\times \Bigg(\sum_{y_1^{k_n+l_n+n} \in \mathcal{E}_1(i)} W_1^{k_n+l_n+n}\left(y_1^{k_n+l_n+n}|x_m^{k_n+l_n+n}(i), s^{k_n+l_n+n}\right)$$
$$+ \sum_{y_1^{k_n+l_n+n} \in \mathcal{E}_2(i,m_2)} W_1^{k_n+l_n+n}\left(y_1^{k_n+l_n+n}|x_m^{k_n+l_n+n}(i), s^{k_n+l_n+n}\right) \Bigg)$$

where we bound both terms individually. For the first term we observe that

$$\sum_{s^{k_n+l_n+n} \in \mathcal{E}_1(i)} W_1^{k_n+l_n+n}\left(y_1^{k_n+l_n+n}|x_m^{k_n+l_n+n}(i), s^{k_n+l_n+n}\right)$$
$$= \sum_{y_1^{k_n}:i \notin \mathcal{L}_1^{(1)}(y_1^{k_n})} W_1^{k_n}\left(y_1^{k_n}|x_i^{k_n}, s^{k_n}\right)$$

where the equality follows from the concatenated structure of the codewords (21) and the fact that the error event $\mathcal{E}_1(i)$ only depends on the first part. Thus, by (20a) we end up with

$$\frac{1}{n^2} \frac{1}{|\mathcal{M}|} \sum_{i \in \{1,\ldots,n^2\}} \sum_{m \in \mathcal{M}} \sum_{i \notin \mathcal{L}_1^{(1)}(y_1^{k_n})} W_1^{k_n}\left(y_1^{k_n}|x_i^{k_n}, s^{k_n}\right)$$
$$= \frac{1}{n^2} \sum_{i \in \{1,\ldots,n^2\}} W_1^{k_n}\left(y_1^{k_n}|x_i^{k_n}, s^{k_n}\right) \leq \lambda_{\text{pre}}^{(k_n)}.$$

For the second event we observe that if $y_1^{k_n+l_n+n} \in \mathcal{E}_2(i, m_2)$, then we have $i \in \mathcal{L}_1^{(1)}(y_1^{k_n})$ and $y_1^n \notin \pi_i^{-1}(D_{m_2|m_1}^{(1)})$ so that

$$\mathcal{E}_2(i, m_2) \subset \{y_1^{k_n+l_n+n} : i \in \mathcal{L}_1^{(1)}(y_1^{k_n}) \text{ and } y_1^n \notin \pi_i^{-1}(D_{m_2|m_1}^{(1)})\}.$$

With this we obtain for the second term

$$\sum_{y_1^{k_n+l_n+n} \in \mathcal{E}_2(i,m_2)} W_1^{k_n+l_n+n}\big(y_1^{k_n+l_n+n}|x_m^{k_n+l_n+n}(i), s^{k_n+l_n+n}\big)$$

$$\leq \sum_{\substack{y_1^{k_n+l_n+n}:i\in\mathcal{L}_1^{(1)}(y_1^{k_n}) \\ \text{and } y_1^n \notin \pi_i^{-1}(D_{m_2|m_1}^{(1)})}} W_1^{k_n+l_n+n}\big(y_1^{k_n+l_n+n}|x_m^{k_n+l_n+n}(i), s^{k_n+l_n+n}\big)$$

$$\leq \sum_{y_1^n:y_1^n \notin \pi_i^{-1}(D_{m_2|m_1}^{(1)})} W_1^n\big(y_1^n|\pi_i^{-1}(x_m^n), s^n\big)$$

$$= W_1^n\big((\pi_i^{-1}(D_{m_2|m_1}^{(1)}))^c|\pi_i^{-1}(x_m^n), s^n\big)$$

where the last inequality follows from the concatenated structure of the codewords (21) and the last equality from the definition of the decoding sets of random codes. Thus,

$$\frac{1}{n^2}\frac{1}{|\mathcal{M}|} \sum_{i\in\{1,\dots,n^2\}} \sum_{m\in\mathcal{M}} W_1^n\big((\pi_i^{-1}(D_{m_2|m_1}^{(1)}))^c|\pi_i^{-1}(x_m^n), s^n\big)$$

$$\leq \tilde{\lambda}_{\mathcal{W},\text{ran}}^{(n)}$$

since it is a "good" random code according to Definition 6. Since $\tilde{\lambda}_{\mathcal{W},\text{ran}}^{(n)}$ and $\lambda_{\text{pre}}^{(k_n)}$ become arbitrary small, the final error (23) is small as well establishing the desired behavior of the error probability.

Moreover, the rate of the final list code $\mathcal{C}_{\mathcal{W},\text{list}}$ from the relay node to node 1 is given by

$$\frac{1}{k_n+l_n+n} \log\left(\frac{2nM_{2,n}}{L_1 L_2}\right)$$

$$= \frac{1}{k_n+l_n+n}\left(\log M_{2,n} + \log\left(\frac{n}{L_1}\right) + \log\left(\frac{n}{L_2}\right)\right)$$

$$= \frac{1}{\frac{k_n+l_n}{n}+1}\frac{1}{n}\log M_{2,n}$$

$$\quad + \frac{1}{1+\frac{n}{k_n+l_n}}\frac{1}{k_n+l_n}\left(\log\left(\frac{n}{L_1}\right) + \log\left(\frac{n}{L_2}\right)\right)$$

$$\xrightarrow{n\to\infty} R_1.$$

since $\frac{1}{\frac{k_n+l_n}{n}+1} \to 1$ and $\frac{1}{1+\frac{n}{k_n+l_n}} \to 0$ as $n \to \infty$ and $\frac{1}{n}\log M_{2,n} = R_1$. Similarly, we get $\frac{1}{k_n+l_n+n}\log(\frac{2nM_{1,n}}{L_1 L_2}) \to R_2$ for the rate from the relay node to node 2. This shows that the overall rate of the final list code is only negligibly affected by the addition of the prefixes.

Consequently, all rate pairs achievable with the random code $\mathcal{C}_{\mathcal{W},\text{ran}}$ are also achievable with the list code $\mathcal{C}_{\mathcal{W},\text{list}}$ with arbitrarily small average probability of error if interior($\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)$) $\neq \emptyset$ as stated in (18).

Remark 7: Due to the product structure of $\mathcal{C}_{\mathcal{W},\text{list}}$ this code is a special case of a list code for the AVBC $\mathcal{W}$ with receiver side information, see Definition 4, and consequently, $\mathcal{C}_{\mathcal{W},\text{list}}$ might not achieve the maximal achievable rates. But the converse in Section VII-D shows that $\mathcal{C}_{\mathcal{W},\text{list}}$ actually achieves all possible rate pairs so that this product structure is already optimal.

VI. SYMMETRIZABILITY

Already Blackwell, Breiman, and Thomasian observed that under certain conditions the deterministic code capacity of the single-user AVC is zero [15]. Based on an idea of Ericson [18], Csiszár and Narayan showed that non-symmetrizability is a necessary condition for the single-user AVC to have a non-zero capacity [19]. Independently, Blinovsky, Narayan, and Pinsker [20] and Hughes [21] extended this idea to the case of list decoding.

Here, we want to establish similar results for the AVBC with receiver side information. For this purpose we use the concept of symmetrizability as introduced in Section III-A and define the maximum single-user rates as

$$R_{i,\max} := \max_{P_X \in \mathcal{P}(\mathcal{X})} \inf_{q\in\mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{i,q}), \qquad i = 1, 2.$$

The following theorem relates the symmetrizability and the maximum single-user rates.

Theorem 5: If $R_{i,\max} = 0$, then the AVBC $\mathcal{W}$ with receiver side information is $(\mathcal{Y}_i, t_i)$-symmetrizable for all $t_i \geq 1$, $i = 1, 2$. If $R_{i,\max} > 0$, then any $(\mathcal{Y}_i, t_i)$-symmetrizable AVBC $\mathcal{W}$ with receiver side information satisfies

$$t_i \leq \frac{\log(\min\{|\mathcal{Y}_i|, |\mathcal{S}|\})}{R_{i,\max}} \tag{24}$$

Proof: The proof can be found in Appendix A. ∎

It follows that for any AVBC with receiver side information, whose random code capacity region has a non-empty interior, the symmetrizability is always finite. With the result in Theorem 4 about the list capacity region, this implies that the list sizes at the receiving nodes are also finite. The next lemma presents a lower bound on the average probability of error in a similar way as in [21] for the single-user case.

Lemma 2: Let (T_1, T_2) be the symmetrizability of an AVBC $\mathcal{W}$ with receiver side information. Then any list code $\mathcal{C}_{\mathcal{W},\text{list}}$ of block length n with $M_{1,n}M_{2,n}$ messages and $L_1 \leq T_1$ satisfies

$$\max_{s^n\in\mathcal{S}^n} \bar{e}_1(s^n|\mathcal{C}_{\mathcal{W},\text{list}}) \geq \left(1 - \frac{L_1}{K_1+1}\right)\left(\frac{M_{2,n}-K_1}{M_{2,n}}\right)$$

where $K_1 = \min\{M_{2,n}-1, T_1\}$. Similarly, any list code $\mathcal{C}_{\mathcal{W},\text{list}}$ of block length n with $M_{1,n}M_{2,n}$ messages and $L_2 \leq T_2$ satisfies

$$\max_{s^n\in\mathcal{S}^n} \bar{e}_2(s^n|\mathcal{C}_{\mathcal{W},\text{list}}) \geq \left(1 - \frac{L_2}{K_2+1}\right)\left(\frac{M_{1,n}-K_2}{M_{1,n}}\right)$$

where $K_2 = \min\{M_{1,n}-1, T_2\}$.

Proof: The proof can be found in Appendix B. ∎

The lemma indicates when the interior of the list capacity region of the AVBC $\mathcal{W}$ with receiver side information will be empty. In more detail, if $L_i \leq T_i$, $i = 1, 2$, then $\max_{s^n\in\mathcal{S}^n} \bar{e}_i(s^n|\mathcal{C}_{\mathcal{W},\text{list}}) > 0$, $i = 1, 2$, which results in interior($\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)$) = $\emptyset$. Consequently, $L_i > T_i$, $i = 1, 2$ is a necessary condition for $\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2) = \mathcal{R}_{\mathcal{W},\text{ran}}$. In other words, for fixed list sizes (L_1, L_2), non-$(\mathcal{Y}_1, L_1)$-symmetrizability and non-$(\mathcal{Y}_2, L_2)$-symmetrizability is necessary for $\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2) = \mathcal{R}_{\mathcal{W},\text{ran}}$.

VII. ACHIEVING POSITIVE RATES

In this section, we present a coding strategy that achieves the desired rates as specified in Theorem 4 if $L_i > T_i$, $i = 1, 2$. Moreover, this immediately shows that $L_i > T_i$, $i = 1, 2$, is also a sufficient condition for interior($\mathcal{R}_{\mathcal{W},\text{list}}(L_1, L_2)$) $\neq \emptyset$.

The coding strategy in the following is based on [21] where a similar strategy is presented for the single-user case.

A. Coding Strategy

To achieve positive rates we need a suitable set of codewords $x^n_{m_1,m_2}$, $m_1 = 1,\ldots,M_{1,n}$, $m_2 = 1,\ldots,M_{2,n}$, with properties as stated in the following lemma.

Lemma 3: For any $L_1 \geq 1, L_2 \geq 1, \epsilon > 0$, $n \geq \max\{n_0(\epsilon,L_1), n_0(\epsilon,L_2)\}$, $M_{1,n} \geq L_2 2^{n\epsilon}$, $M_{2,n} \geq L_1 2^{n\epsilon}$, and given type P_X, there exist codewords $x^n_{m_1,m_2} \in \mathcal{X}^n$, $m_1 = 1,\ldots,M_{1,n}$, $m_2 = 1,\ldots,M_{2,n}$, each of type P_X, such that for every $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n$, and every joint type $P_{XX^{L_1}S}$ with $X^{L_1} = X_1, X_2, \ldots, X_{L_1}$ we have for each $m_1 \in \mathcal{M}_1$

$$\left|\{m_2' : (x^n, x^n_{m_1,m_2'}, s^n) \in \tau_{XX_k S}\}\right|$$
$$\leq 2^{n(|R_1 - I(X_k;XS)|^+ + \epsilon)} \quad (25a)$$

$$\frac{1}{M_{2,n}}\left|\{m_2 : (x^n_{m_1,m_2}, s^n) \in \tau_{XS}\}\right|$$
$$\leq 2^{-n\frac{\epsilon}{2}} \quad \text{if } I(X;S) \geq \epsilon \quad (25b)$$

$$\frac{1}{M_{2,n}}\left|\{m_2 : (x^n_{m_1,m_2}, x^n_{m_1,m_2'}, s^n) \in \tau_{XX_k S}\right.$$
$$\left.\text{for some } m_2' \neq m_2\}\right| \leq 2^{-n\frac{\epsilon}{2}}$$
$$\text{if } I(X;X_k S) - |R_1 - I(X_k;S)|^+ \geq \epsilon \quad (25c)$$

for $k = 1,\ldots,L_1$. Moreover, if $R_1 < \min_k I(X_k;S)$, then $x^n_{m_1,m_2}$, $m_1 = 1,\ldots,M_{1,n}$, $m_2 = 1,\ldots,M_{2,n}$, further satisfy

$$\left|\{\mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2) : (x^n, x^n_{m_1,\mathcal{J}}, s^n) \in \tau_{XX^{L_1}S}\}\right| \leq 2^{n\epsilon} \quad (25d)$$

$$\frac{1}{M_{2,n}}\left|\{m_2 : (x^n_{m_1,m_2}, x^n_{m_1,\mathcal{J}}, s^n) \in \tau_{XX^{L_1}S}\right.$$
$$\left.\text{for some } \mathcal{J} \in \mathfrak{P}_{L_1}(\mathcal{M}_2\setminus\{m_2\})\}\right| \leq 2^{-n\frac{\epsilon}{2}}$$
$$\text{if } I(X;X^{L_1}S) \geq \epsilon \quad (25e)$$

with $\mathcal{J} = \{j_1,\ldots,j_{L_1}\} \in \mathfrak{P}_{L_1}(\mathcal{M}_2)$ and $x^n_{m_1,\mathcal{J}}$ denotes the ordered L_1-tuple $(x^n_{m_1,j_1}, x^n_{m_1,j_2}, \ldots x^n_{m_1,j_{L_1}})$ where the indices are ordered as $j_1 < j_2 < \cdots < j_{L_1}$. Similarly, for every $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n$, and every joint type $P_{XX^{L_2}S}$ we have for each $m_2 \in \mathcal{M}_2$

$$\left|\{m_1' : (x^n, x^n_{m_1',m_2}, s^n) \in \tau_{XX_k S}\}\right|$$
$$\leq 2^{n(|R_2 - I(X_k;XS)|^+ + \epsilon)} \quad (25f)$$

$$\frac{1}{M_{1,n}}\left|\{m_1 : (x^n_{m_1,m_2}, s^n) \in \tau_{XS}\}\right|$$
$$\leq 2^{-n\frac{\epsilon}{2}} \quad \text{if } I(X;S) \geq \epsilon \quad (25g)$$

$$\frac{1}{M_{1,n}}\left|\{m_1 : (x^n_{m_1,m_2}, x^n_{m_1',m_2}, s^n) \in \tau_{XX_k S}\right.$$
$$\left.\text{for some } m_1' \neq m_1\}\right| \leq 2^{-n\frac{\epsilon}{2}}$$
$$\text{if } I(X;X_k S) - |R_2 - I(X_k;S)|^+ \geq \epsilon \quad (25h)$$

for $k = 1,\ldots,L_2$. Moreover, if $R_2 < \min_k I(X_k;S)$, then $x^n_{m_1,m_2}$, $m_1 = 1,\ldots,M_{1,n}$, $m_2 = 1,\ldots,M_{2,n}$, further

satisfy

$$\left|\{\mathcal{J}' \in \mathfrak{P}_{L_2}(\mathcal{M}_1) : (x^n, x^n_{\mathcal{J}',m_2}, s^n) \in \tau_{XX^{L_2}S}\}\right| \leq 2^{n\epsilon} \quad (25i)$$

$$\frac{1}{M_{1,n}}\left|\{m_1 : (x^n_{m_1,m_2}, x^n_{\mathcal{J}',m_2}, s^n) \in \tau_{XX^{L_2}S}\right.$$
$$\left.\text{for some } \mathcal{J}' \in \mathfrak{P}_{L_2}(\mathcal{M}_1\setminus\{m_1\})\}\right| \leq 2^{-n\frac{\epsilon}{2}}$$
$$\text{if } I(X;X^{L_2}S) \geq \epsilon \quad (25j)$$

with $\mathcal{J}' = \{j_1',\ldots,j_{L_2}'\} \in \mathfrak{P}_{L_2}(\mathcal{M}_1)$ and $x^n_{\mathcal{J}',m_2}$ denotes the ordered L_2-tuple $(x^n_{j_1',m_2}, x^n_{j_2',m_2}, \ldots, x^n_{j_{L_2}',m_2})$ where the indices are ordered as $j_1' < j_2' < \cdots < j_{L_2}'$.

Sketch of Proof: We have to show that $M_{1,n}M_{2,n}$ with $M_{1,n} = 2^{nR_2}$ and $M_{2,n} = 2^{nR_1}$ randomly selected codewords will possess, with probability arbitrarily close to one, the properties (25a)-(25j). Basically, the desired properties follow from [21]. For each $m_1 \in \mathcal{M}_1$ the proof technique of [21, Lemma 1] establishes the properties (25a)-(25e) and, similarly, again for each $m_2 \in \mathcal{M}_2$ the proof technique of [21, Lemma 1] establishes the properties (25f)-(25j). The details are omitted for brevity. ∎

The proof of the lemma shows that these good codewords are obtained by randomly selecting codewords from the set of sequences of a fixed type. The collection of selected codewords will possess the desired properties with probability arbitrarily close to 1.

B. Decoding Strategy

A crucial part is to define suitable decoding rules at the receiving nodes 1 and 2. For the single-user AVC with list size one Csiszár and Narayan use in [19] a generalized divergence typicality decoder based on an idea of Dobrushin and Stambler [38] which decides on the basis of a joint typicality test together with a threshold test using empirical mutual information quantities. Blinovsky et. al. [20] and Hughes [21] use a generalization of the above mentioned decoder that is modified to apply to greater list sizes. We follow their approach and define for this purpose a family of joint distributions P_{XSY_i} of random variables X, S, and Y_i with values in $\mathcal{X}$, $\mathcal{S}$, and $\mathcal{Y}_i$, respectively, by

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

with $\eta_i \geq 0$ and where $P_X \otimes P_S \otimes W_i$ denotes a joint distribution on $\mathcal{X} \times \mathcal{S} \times \mathcal{Y}_i$ with probability mass function $P_X(x)P_S(s)W_i(y_i|x,s)$. In particular, we have $P_{XSY_i} \in \mathcal{D}_0$ if and only if

$$P_{XSY_i}(x,s,y_i) = P_X(x)P_S(s)W_i(y_i|x,s).$$

With this we are able to define the decoding rule at node 1 for list size L_1 as follows.

Definition 7: For given codewords $x^n_{m_1,m_2} \in \tau_X$, $m_1 = 1,\ldots,M_{1,n}, 1$, $m_2 = 1,\ldots,M_{2,n}$, and (small) $\eta_1 > 0$ the decoding rule $\mathcal{L}^{(1)} : \mathcal{Y}^n_1 \times \mathcal{M}_1 \to \hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$ at node 1 is defined as follows: we have $m_2 \in \mathcal{L}^{(1)}(y^n_1, m_1)$ if and only if

i) there exists an $s^n \in \mathcal{S}^n$ such that

$$P_{x^n_{m_1,m_2}, s^n, y^n_1} \in \mathcal{D}_{\eta_1}$$

ii) for each choice of L_1 other distinct codewords $x_{m_1,j_1}^n, \ldots, x_{m_1,j_{L_1}}^n$, where each satisfies

$$P_{x_{m_1,j_i}^n, s_i^n, y_1^n} \in \mathcal{D}_{\eta_1} \qquad 1 \leq i \leq L_1$$

for some $s_i^n \in \mathcal{S}^n$, we have

$$I(XY_1; X^{L_1}|S) \leq \eta_1$$

where $X^{L_1} = X_1, X_2, \ldots, X_{L_1}$ and $P_{XX^{L_1}SY_1}$ is the joint type of $(x_{m_1,m_2}^n, x_{m_1,j_1}^n, \ldots, x_{m_1,j_{L_1}}^n, s^n, y_1^n)$.

The decoding rule $\mathcal{L}^{(2)} : \mathcal{Y}_2^n \times \mathcal{M}_2 \to \hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$ at node 2 with list size L_2 is defined accordingly with (small) constant $\eta_2 > 0$. To establish the list capacity region for $L_i > T_i$, $i = 1, 2$ (see Theorem 4), we have to ensure that the decoding rule as specified in Definition 7 is well defined. This means that the decoding rule satisfy the given constraints on the list sizes, i.e., $|\mathcal{L}^{(1)}(y_1^n, m_1)| \leq L_1$ for all $m_1 \in \mathcal{M}_1$ and $|\mathcal{L}^{(2)}(y_2^n, m_2)| \leq L_2$ for all $m_2 \in \mathcal{M}_2$. We show that the decoding rule already satisfies $|\mathcal{L}^{(i)}(y_i^n, m_i)| \leq T_i + 1$ for all $y_i^n \in \mathcal{Y}_i^n$ and $m_i \in \mathcal{M}_i$, $i = 1, 2$, which is clearly sufficient. Here is where the symmetrizability conditions come in.

Lemma 4: Let $\beta > 0$, then for a sufficiently small η_i, $i = 1, 2$, no ensemble $X^{T_i+2}, S^{T_i+2}, Y_i$ can simultaneously satisfy

$$\min_x P_X(x) \geq \beta$$

and

$$P_{X_k} = P, \qquad P_{X_k S_k Y_i} \in \mathcal{D}_{\eta_i}$$
$$I(X_k Y_i; X_k^{T_i+2}|S_k) \leq \eta_i \qquad 1 \leq k \leq T_i + 2 \qquad (26)$$

with $X_k^{T_i+2} = X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_{T_i+2}$.

Proof: The proof can be found in [21, Lemma 2]. ∎

C. Positive Rates

So far we defined coding and decoding rules. Next, we show that this strategy is sufficient to achieve the desired rates if the list sizes are great enough, which means $L_i > T_i$, $i = 1, 2$. Clearly, it suffices to show this for $L_k = T_k + 1$, since any rate pair achievable with these list sizes is also achievable with greater list sizes.

Lemma 5: Let $L_i = T_i + 1$, $i = 1, 2$, and $\beta > 0$, $\delta > 0$. For any type P_X satisfying $\min_x P_X(x) \geq \beta$, there exists a list code $\mathcal{C}_{\mathcal{W},\text{list}}$ of block length $n \geq n_2$ with list sizes (L_1, L_2) and codewords $x_{m_1,m_2}^n \in \tau_X$, $m_1 = 1, \ldots, M_{1,n}$, $m_2 = 1, \ldots, M_{2,n}$, such that

$$R_1 > \inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{1,q}) - \delta, \qquad R_2 > \inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{2,q}) - \delta,$$

while

$$\max_{s^n \in \mathcal{S}^n} \bar{e}_i(s^n|\mathcal{C}_{\mathcal{W},\text{list}}) < 2^{-n\gamma_i}, \qquad i = 1, 2, \qquad (27)$$

where n_2 and $\gamma_i > 0$ depend only on β, δ, and the AVBC $\mathcal{W}$ with receiver side information.

Proof: The proof follows [21, Lemma 3] where a similar result is shown for the single-user AVC. Let $x_{m_1,m_2}^n \in \tau_X$, $m_1 = 1, \ldots, M_{1,n}$, $m_2 = 1, \ldots, M_{2,n}$, be codewords with properties as specified in Lemma 3 and $R_1 = \frac{1}{n}\log(\frac{M_{2,n}}{L_1})$ and

$R_2 = \frac{1}{n}\log(\frac{M_{1,n}}{L_2})$. Let the list decoders $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ be as given in Definition 7. By Lemma 4 we can choose η_1 and η_2 small enough to ensure that $|\mathcal{L}^{(i)}(y_i^n, m_i)| \leq T_i + 1$ for all $y_i^n \in \mathcal{Y}_i^n$ and $m_i \in \mathcal{M}_i$, $i = 1, 2$. Now, we have to show that (27) is satisfied, i.e., the decoding error at the receiving nodes vanishes. With the properties of the codewords, see Lemma 3, and the decoding rule, see Definition 7 and Lemma 4, this can be shown analogously to [21, Lemma 3], where a similar result is shown for the single-user case. We omit the details for brevity. ∎

D. Converse

To complete the proof of Theorem 4 it remains to show that the presented strategy actually achieves all possible rate pairs so that no other rate pairs are achievable. For $L_i \leq T_i$ the converse part is already established by Lemma 2, since it shows that for $L_i \leq T_i$ no positive rates are achievable. Consequently, it remains to consider the case $L_i > T_i$, $i = 1, 2$. To avoid trivialities we further assume $L_1 \leq M_{2,n}$ and $L_2 \leq M_{1,n}$. We need a version of Fano's lemma for list decoding.

Lemma 6: Let U be a random variable with values in $\mathcal{M} := \{1, \ldots, M\}$ and V a random variable with values in $\hat{\mathfrak{P}}_L(\mathcal{M})$, i.e., the set of all subsets of $\mathcal{M}$ that contains at most L elements. Then

$$H(U|V) \leq h(\mathbb{P}\{U \notin V\}) + \log L + \mathbb{P}\{U \notin V\}\log\left(\frac{M}{L} - 1\right)$$

with $h(\cdot)$ the binary entropy.

Proof: The proof can be found in [39]. ∎

To prove the converse we have to show that any given sequence of $(n, M_{1,n}, M_{2,n}, L_1, L_2, \lambda^{(n)})$-list codes with list sizes $L_i > T_i$, $i = 1, 2$, and $\lambda^{(n)} \to 0$ must satisfy

$$R_2 = \frac{1}{n}\log\left(\frac{M_{1,n}}{L_2}\right) \leq \inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{2,q}) + o(1) \quad (28a)$$

$$R_1 = \frac{1}{n}\log\left(\frac{M_{2,n}}{L_1}\right) \leq \inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{1,q}) + o(1) \quad (28b)$$

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

As a first step it is easy to show that any list code that is a good code for an AVBC with receiver side information is also a good code for an appropriately constructed compound BC with receiver side information. In more detail, let $\mathcal{C}_{\mathcal{W},\text{list}}$ be a $(n, M_{1,n}, M_{2,n}, L_1, L_2, \lambda^{(n)})$-list code for an AVBC $\mathcal{W}$ with receiver side information with average probability of error at node i, $i = 1, 2$,

$$\bar{e}_i(s^n|\mathcal{C}_{\mathcal{W},\text{list}}) \leq \lambda^{(n)} \qquad \text{for all } s^n \in \mathcal{S}^n. \qquad (29)$$

Since (29) holds for all $s^n \in \mathcal{S}^n$, it immediately follows that the same is also true for any affine combination, i.e.,

$$\sum_{s^n \in \mathcal{S}^n} \bar{e}_i(s^n|\mathcal{C}_{\mathcal{W},\text{list}})q^{\otimes n}(s^n) \leq \lambda^{(n)} \qquad \text{for all } q \in \mathcal{P}(\mathcal{S}). \quad (30)$$

With the definition of the probability of error, see (5) and (6), for receiving node 1 Equation (30) reads as

$$\sum_{s^n \in \mathcal{S}^n} \frac{1}{|\mathcal{M}|} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} W_1^{\otimes n}(y_1^n|x_m^n, s^n)q^{\otimes n}(s^n) \leq \lambda^{(n)}$$

$$\text{for all } q \in \mathcal{P}(\mathcal{S})$$

or equivalently with the definition of an averaged broadcast channel, see (3), as

$$\frac{1}{|\mathcal{M}|} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} \overline{W}_{1,q}^{\otimes n}(y_1^n | x_m^n) \leq \lambda^{(n)}$$
$$\text{for all } q \in \mathcal{P}(\mathcal{S}). \quad (31)$$

The same arguments leads for receiving node 2 to $\frac{1}{|\mathcal{M}|} \sum_{y_2^n : m_2 \notin \mathcal{L}^{(2)}(y_2^n, m_2)} \overline{W}_{2,q}^{\otimes n}(y_2^n | x_m^n) \leq \lambda^{(n)}$ for all $q \in \mathcal{P}(\mathcal{S})$. Since (31) holds for all $q \in \mathcal{P}(\mathcal{S})$, the list code $\mathcal{C}_{\mathcal{W}, \text{list}}$ is also a good code for the compound BC $\overline{\mathcal{W}} = \{\overline{W}_q(y_1, y_2 | x)\}_{q \in \mathcal{P}(\mathcal{S})}$. Consequently for the AVBC $\mathcal{W}$ with receiver side information we cannot achieve higher rates as for the constructed compound BC $\overline{\mathcal{W}}$ with receiver side information. Therefore, to establish the converse result for the AVBC $\mathcal{W}$, it remains to show that the rates for the compound BBC $\overline{\mathcal{W}}$ are already bounded from above by (28).

Furthermore, it is sufficient to show that for a specific $q \in \mathcal{P}(\mathcal{S})$ the rates are bounded by

$$R_1 \leq I(X, \overline{Y}_{1,q}) + o(1) \quad \text{and} \quad R_2 \leq I(X, \overline{Y}_{2,q}) + o(1). \quad (32)$$

Since for the compound BC $\overline{\mathcal{W}}$ with receiver side information the rates have to satisfy (32) for all possible $q \in \mathcal{P}(\mathcal{S})$, the rates are immediately bounded by the corresponding infima $\inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{1,q}) + o(1)$ and $\inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{2,q}) + o(1)$.

Now we are in the position to prove (32). Let $\mathcal{C}_{\mathcal{W}, \text{list}}$ be any $(n, M_{1,n}, M_{2,n}, L_1, L_2, \lambda^{(n)})$-list code with

$$\frac{1}{|\mathcal{M}_1|} \sum_{m_1 \in \mathcal{M}_1} \frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)}$$
$$\times \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1, m_2}^n) \leq \lambda_1^{(n)},$$
$$\frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \frac{1}{|\mathcal{M}_1|} \sum_{m_1 \in \mathcal{M}_1} \sum_{y_2^n : m_1 \notin \mathcal{L}^{(2)}(y_2^n, m_2)}$$
$$\times \overline{W}_{2,q}^{\otimes n}(y_2^n | x_{m_1, m_2}^n) \leq \lambda_2^{(n)}.$$

Let us consider random variables $U_i, X^n, Y_i^n, i = 1, 2$, with values in $\mathcal{M}_i, \mathcal{X}^n, \mathcal{Y}_i^n, i = 1, 2$, respectively, and with

$$\mathbb{P}\{U_1 = m_1, U_2 = m_2, X^n = x^n, Y_1^n = y_1^n, Y_2^n = y_2^n\}$$
$$= \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} p(x_{m_1, m_2}^n | m_1, m_2) \overline{W}_q^{\otimes n}(y_1^n, y_2^n | x_{m_1, m_2}^n)$$
$$= \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} p(x_{m_1, m_2}^n | m_1, m_2)$$
$$\times \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1, m_2}^n) \overline{W}_{2,q}^{\otimes n}(y_2^n | x_{m_1, m_2}^n),$$

where $p(x_{m_1, m_2}^n | m_1, m_2) = 1$ if x_{m_1, m_2}^n is the codeword corresponding to $m = (m_1, m_2)$ or is equal to 0 else. Further, let V_1 and V_2 be random variables with values in $\hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$ and $\hat{\mathfrak{P}}_{L_2}(\mathcal{M}_1)$, respectively, and set $V_1 := \mathcal{L}^{(1)}(Y_1^n, U_1)$ and $V_2 := \mathcal{L}^{(2)}(Y_2^n, U_2)$.

By definition

$$\mathbb{P}\{U_2 \notin V_1\} = \sum_{m_2 \in \mathcal{M}_2} \mathbb{P}\{U_2 = m_2, m_2 \notin \mathcal{L}^{(1)}(Y_1^n, m_1)\}$$
$$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{x_{m_1, m_2}^n \in \mathcal{X}^n} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)}$$
$$\times p(x_{m_1, m_2}^n | m_1, m_2) \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1, m_2}^n)$$
$$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)} \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1, m_2}^n)$$
$$= \frac{1}{|\mathcal{M}_1|} \sum_{m_1 \in \mathcal{M}_1} \frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1)}$$
$$\times \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1, m_2}^n)$$
$$\leq \lambda_1^{(n)}.$$

By averaging separately over all $m_1 \in \mathcal{M}_1$ and all $m_2 \in \mathcal{M}_2$ we see that there exists for each block length n a fixed $m_1^* \in \mathcal{M}_1$ such that

$$\frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} \sum_{y_1^n : m_2 \notin \mathcal{L}^{(1)}(y_1^n, m_1^*)} \overline{W}_{1,q}^{\otimes n}(y_1^n | x_{m_1^*, m_2}^n) \leq \lambda_1^{(n)}$$

is fulfilled. Then $U_2 - X^n - Y_1^n - V_1(m_1^*, q)$ forms a Markov chain and we are in the same position as in the single-user case with a single-user list code. We want to emphasize that the random variable $V_1(m_1^*, q)$ clearly depends on the specific $m_1^* \in \mathcal{M}_1$ and $q \in \mathcal{P}(\mathcal{S})$ but for the sake of brevity we write V_1 for $V_1(m_1^*, q)$ in the following. We get

$$\log M_{2,n} = H(U_2) = H(U_2 | V_1) + I(U_2; V_1)$$
$$\leq H(U_2 | V_1) + I(X^n; Y_1^n)$$
$$\leq h\big(\mathbb{P}\{U_2 \notin V_1\}\big) + \log L_1$$
$$\quad + \mathbb{P}\{U_2 \notin V_1\} \log\left(\frac{M_{2,n}}{L_1}\right) + I(X^n; Y_1^n)$$
$$\leq h\big(\mathbb{P}\{U_2 \notin V_1\}\big) + \log L_1$$
$$\quad + \mathbb{P}\{U_2 \notin V_1\} \log\left(\frac{M_{2,n}}{L_1}\right) + \sum_{k=1}^{n} I(X_k; Y_{1,k}) \quad (33)$$

where the equality and inequalities follow from the definition of mutual information, the data processing inequality, Lemma 6 with $\mathcal{L}^{(1)} : \mathcal{Y}_1^n \times \mathcal{M}_1 \to \hat{\mathfrak{P}}_{L_1}(\mathcal{M}_2)$, and the memoryless property of the channel. We can rewrite the mutual information term on the right hand side of (33), see Section II, as

$$\log M_{2,n} \leq h\big(\mathbb{P}\{U_2 \notin V_1\}\big) + \log L_1$$
$$\quad + \mathbb{P}\{U_2 \notin V_1\} \log\left(\frac{M_{2,n}}{L_1}\right) + \sum_{k=1}^{n} I(X_k, \overline{Y}_{1,q})$$
$$\leq h\big(\mathbb{P}\{U_2 \notin V_1\}\big) + \log L_1$$
$$\quad + \mathbb{P}\{U_2 \notin V_1\} \log\left(\frac{M_{2,n}}{L_1}\right) + n I(X, \overline{Y}_{1,q}) \quad (34)$$

where the last inequality follows from the concavity of mutual information and $P_X := \frac{1}{n} \sum_{k=1}^{n} P_{X_k}$. We note that due to the continuity of the mutual information and the compactness of the set of probability distributions, the dependency of P_X on the block length n vanishes asymptotically.

Rearranging the terms in (33) and dividing by n leads to

$$\left(1 - \mathbb{P}\{U_2 \notin V_1\}\right)\frac{1}{n}\log\left(\frac{M_{2,n}}{L_1}\right) \leq \frac{1}{n}h\left(\mathbb{P}\{U_2 \notin V_1\}\right) + I(X, \overline{Y}_{1,q})$$

or

$$\left(1 - \lambda_1^{(n)}\right)\frac{1}{n}\log\left(\frac{M_{2,n}}{L_1}\right) \leq \frac{1}{n}h\left(\lambda_1^{(n)}\right) + I(X, \overline{Y}_{1,q}). \quad (35)$$

The same arguments lead for receiving node 2 to $(1 - \lambda_2^{(n)})$ $\frac{1}{n}\log(\frac{M_{1,n}}{L_2}) \leq \frac{1}{n}h(\lambda_2^{(n)}) + I(X, \overline{Y}_{2,q})$. Finally, (32) and therewith the converse follow readily from this and (35). ■

VIII. CONCLUSION

In this paper we studied the arbitrarily varying broadcast channel (AVBC) with receiver side information and completely solved it by characterizing the capacity regions for random, deterministic, and list codes. Similar to Ahlswede's result for the single-user AVC, it is shown that the AVBC with receiver side information under list decoding displays a dichotomy behavior: the list capacity region either equals the random code capacity region or else has an empty-interior. A generalized notion of symmetrizability, which distinguishes among different degrees of symmetry, allowed us to completely characterize when the capacity has a non-empty interior. In particular, if the list size of a receiver is greater than the corresponding symmetrizability, transmission at random code rates is possible. Moreover, the symmetrizability is shown to be always finite so that sufficient list sizes are also always finite.

A complete understanding of (multi-user) scenarios under arbitrarily varying channels is the indispensable basis for further investigations. As an example, the concept of arbitrarily varying channels is a suitable model to analyze secure communication in scenarios, where non-legitimate wiretappers are active in the sense that they can influence or control the channel of the legitimate users. Such adversaries may have different possible strategies; they can try to eavesdrop the confidential communication or can try to jam the channel of the legitimate users. Especially in the latter case, a complete characterization of the capacity of an AVC is needed to decide under which conditions the adversary can successfully jam the channel and how the legitimate users can overcome this.

APPENDIX

A. Proof of Theorem 5

We follow [21, Th. 1] where a similar result for the single-user AVC is proved. We start with the trivial case where $R_{i,\max} = 0$, $i = 1, 2$, which implies that $\inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{i,q}) = 0$ for every $P_X \in \mathcal{P}(\mathcal{X})$. If $P_X(x) > 0$ for all $x \in \mathcal{X}$, this implies the existence of a distribution $P_{XSY_i} = P_X \otimes P_S \otimes W_i$ such that the input X and the output Y_i are independent, which means

$$\sum_s W_i(y_i|x, s)P_S(s) = P_{Y_i}(y_i).$$

Now, for any $t_i \geq 1$ we set

$$U(s|x_1, \ldots, x_{t_i}) := P_S(s)$$

for all $s, x_1, \ldots, x_{t_i}$ to obtain a channel which is symmetric in $x, x_1, \ldots, x_{t_i}$, see also (4). This implies immediately that the AVBC $\mathcal{W}$ with receiver side information is $(\mathcal{Y}_i, t_i)$-symmetrizable for all $t_i \geq 1$, $i = 1, 2$.

Next, we assume that $R_{i,\max} > 0$ and that the AVBC $\mathcal{W}$ with receiver side information is $(\mathcal{Y}_i, t_i)$-symmetrizable and further $\inf_{q \in \mathcal{P}(\mathcal{S})} I(P_X, \overline{W}_{i,q}) > 0$, $i = 1, 2$. Consequently, there is a channel $U : \mathcal{X}^{t_i} \to \mathcal{P}(\mathcal{S})$ such that

$$\widetilde{W}_i(y_i|x_1, \ldots, x_{t_i+1}) := \sum_s W_i(y_i|x_1, s)U(s|x_2, \ldots, x_{t_i+1}) \quad (36)$$

is symmetric in $x_1, \ldots, x_{t_i+1}$. Let $X^{t_i+1} = X_1, \ldots, X_{t_i+1}$ be a sequence of independent random variables each with distribution P_X. Further, denote the output of the auxiliary channel U by S' corresponding to the input $X_2, \ldots, X_{t_i+1}$, and the output of the channel W_i by Y_i' for the inputs X_1 and S'. As in [21] for the single-user AVC we observe that $X^{t_i+1} - X_1S' - Y_i'$ forms a Markov chain, so that the Data Processing Inequality [17, Lemma 3.11] gives

$$I(X_1S'; Y_i') \geq I(X^{t_i+1}; Y_i')$$
$$\geq \sum_{k=1}^{t_i+1} I(X_k; Y_i')$$
$$= (t_i + 1)I(X_1; Y_i')$$

where the second inequality follows from the independence of $X_1, \ldots, X_{t_i+1}$ and the non-negativity of the (conditional) mutual information and the last equality from $P_{X_kY_i'} = P_{X_1Y_i'}$, which is a consequence of the symmetry of $\widetilde{W}_i$, see (36). If we subtract $I(X_1; Y_i')$ from both sides, having $I(X_1; Y_i') \geq \inf_{q \in \mathcal{P}(\mathcal{S})} I(X, \overline{Y}_{i,q}) > 0$ in mind, we get

$$t_i \leq \frac{I(S'; Y_i'|X_1)}{I(X_1; Y_i')}$$
$$\leq \max_{\substack{P_{XSY_i}:P_{XSY_i}=P_X\otimes P_S\otimes W_i \\ \text{for some } S}} \frac{I(S; Y_i|X)}{I(X; Y_i)}.$$

Clearly, this holds for all P_X so that we finally obtain

$$t_i \leq \min_{P_X} \max_{\substack{P_{XSY_i}:P_{XSY_i}=P_X\otimes P_S\otimes W_i \\ \text{for some } S}} \frac{I(S; Y_i|X)}{I(X; Y_i)}$$
$$\leq \frac{\log(\min\{|\mathcal{S}||\mathcal{Y}_i|\})}{R_{i,\max}}$$

which proves (24). ■

B. Proof of Lemma 2

The lemma follows immediately from Hughes [21, Lemma 4], who proves a similar result for the single-user AVC. But for completeness we present the proof in the following. We carry out the analysis for receiving node 1, then the analysis for node 2 follows accordingly.

First, we observe that it suffices to consider $K_1 \geq 1$, since otherwise there is nothing to prove. We consider any list code $\mathcal{C}_{\mathcal{W},\text{list}}$ with codewords $x_{m_1,m_2}^n = (x_{m_1,m_2,1}, \ldots, x_{m_1,m_2,n}) \in \mathcal{X}^n$, $m_1 = 1, \ldots, M_{1,n}$, $m_2 = 1, \ldots, M_{2,n}$, and a list decoder at node 1 with list size $L_1 \leq T_1$. Since $K_1 \leq T_1$, the AVBC $\mathcal{W}$ with receiver side information is $(\mathcal{Y}_1, K_1)$-symmetrizable so that there exists a channel $U : \mathcal{X}^{K_1} \to \mathcal{P}(\mathcal{S})$ such that

$$\sum_{s \in \mathcal{S}} W_1(y_1|x_1, s) U(s|x_2, \ldots, x_{K_1+1})$$

is symmetric in $x_1, \ldots, x_{K_1+1}$, see also (4).

Then for each $m_1 \in \mathcal{M}_1$ the following holds. For any set $\mathcal{J} = \{j_1, \ldots, j_{K_1}\} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)$ of K_1 messages, let $S_{m_1,\mathcal{J}}^n = (S_{m_1,\mathcal{J},1}, \ldots, S_{m_1,\mathcal{J},n}) \in \mathcal{S}^n$ be a random state sequence with

$$\mathbb{P}\{S_{m_1,\mathcal{J},k} = s\} = U(s|x_{m_1,j_1,k}, \ldots, x_{m_1,j_{K_1},k})$$

for all $k = 1, 2, \ldots, n$. For any $i \in \mathcal{M}_2$ and any $\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)$ as defined above it follows that

$$\mathbb{E}[W_1^n(y_1^n|x_{m_1,i}^n, S_{m_1,\mathcal{J}}^n)] = \prod_{k=1}^n \mathbb{E}[W_1(y_{1,k}|x_{m_1,i,k}, S_{m_1,\mathcal{J},k})]$$

$$= \prod_{k=1}^n \sum_{s \in \mathcal{S}} W_1(y_{1,k}|x_{m_1,i,k}, s) \mathbb{P}\{S_{m_1,\mathcal{J},k} = s\}$$

$$= \prod_{k=1}^n \sum_{s \in \mathcal{S}} W_1(y_{1,k}|x_{m_1,i,k}, s) U(s|x_{m_1,j_1,k}, \ldots, x_{m_1,j_{K_1},k})$$

$$= \sum_{s^n \in \mathcal{S}^n} W_1^n(y_1^n|x_{m_1,i}^n, s^n) U^n(s^n|x_{m_1,j_1}^n, \ldots, x_{m_1,j_{K_1}}^n)$$

is symmetric in $i, j_1, \ldots, j_{K_1}$. Consequently, for any $\mathcal{J}' \in \mathfrak{P}_{K_1+1}(\mathcal{M}_2)$ and any fixed $i_0 \in \mathcal{J}'$, we have

$$\mathbb{E}[W_1^n(y_1^n|x_{m_1,i}^n, S_{m_1,\mathcal{J}'\setminus\{i\}}^n)] = \mathbb{E}[W_1^n(y_1^n|x_{m_1,i_0}^n, S_{m_1,\mathcal{J}'\setminus\{i_0\}}^n)]$$

for all $i \in \mathcal{J}'$. Since the list size of the list decoder at node 1 is L_1, the received y_1^n can be decoded in at most L_1 ways so that it follows for the probability of error that

$$\sum_{i \in \mathcal{J}'} \mathbb{E}[e_1((m_1, i), S_{m_1,\mathcal{J}'\setminus\{i\}}^n | \mathcal{C}_{\mathcal{W},\text{list}})]$$

$$= \sum_{i \in \mathcal{J}'} \left(1 - \sum_{y_1^n : i \in \mathcal{L}^{(1)}(y_1^n, m_1)} \mathbb{E}[W_1^n(y_1^n|x_{m_1,i}^n, S_{m_1,\mathcal{J}'\setminus\{i\}}^n)] \right)$$

$$= K_1 + 1 - \sum_{i \in \mathcal{J}'} \sum_{y_1^n : i \in \mathcal{L}^{(1)}(y_1^n, m_1)} \mathbb{E}[W_1^n(y_1^n|x_{m_1,i_0}^n, S_{m_1,\mathcal{J}'\setminus\{i_0\}}^n)]$$

$$\geq K_1 + 1 - L_1 \sum_{y_1^n \in \mathcal{Y}_1^n} \mathbb{E}[W_1^n(y_1^n|x_{m_1,i_0}^n, S_{m_1,\mathcal{J}'\setminus\{i_0\}}^n)]$$

$$= K_1 + 1 - L_1,$$

with $\mathcal{L}^{(1)}(y_1^n, m_1)$ the list decoder for received output y_1^n and own message m_1. For a fixed $m_1 \in \{1, \ldots, M_{1,n}\}$ this

leads to

$$\frac{1}{|\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \sum_{\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)} \mathbb{E}[\bar{e}_1(S_{m_1,\mathcal{J}}^n | \mathcal{C}_{\mathcal{W},\text{list}})]$$

$$= \frac{1}{M_{1,n} M_{2,n} |\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \sum_{\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)} \sum_{m_1'=1}^{M_{1,n}} \sum_{m_2'=1}^{M_{2,n}}$$

$$\times \mathbb{E}[e_1((m_1', m_2'), S_{m_1,\mathcal{J}}^n | \mathcal{C}_{\mathcal{W},\text{list}})]$$

$$\geq \frac{1}{M_{1,n} M_{2,n} \binom{M_{2,n}}{K_1}} \sum_{m_1'=1}^{M_{1,n}} \sum_{\mathcal{J}' \in \mathfrak{P}_{K_1+1}(\mathcal{M}_2)} \sum_{m_2' \in \mathcal{J}'}$$

$$\times \mathbb{E}[e_1((m_1', m_2'), S_{m_1,\mathcal{J}'\setminus\{m_2'\}}^n | \mathcal{C}_{\mathcal{W},\text{list}})]$$

$$\geq \frac{1}{M_{1,n}} \sum_{m_1'=1}^{M_{1,n}} \frac{\binom{M_{2,n}}{K_1+1}(K_1 + 1 - L_1)}{M_{2,n}\binom{M_{2,n}}{K_1}}$$

$$= \left(1 - \frac{L_1}{K_1 + 1}\right) \left(\frac{M_{2,n} - K_1}{M_{2,n}}\right).$$

Thus, we obtain for the average probability of error

$$\frac{1}{M_{1,n} |\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \sum_{m_1=1}^{M_{1,n}} \sum_{\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)} \mathbb{E}[\bar{e}_1(S_{m_1,\mathcal{J}}^n | \mathcal{C}_{\mathcal{W},\text{list}})]$$

$$= \frac{1}{M_{1,n}} \sum_{m_1=1}^{M_{1,n}} \left(\frac{1}{|\mathfrak{P}_{K_1}(\mathcal{M}_2)|} \sum_{\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)} \mathbb{E}[\bar{e}_1(S_{m_1,\mathcal{J}}^n | \mathcal{C}_{\mathcal{W},\text{list}})] \right)$$

$$\geq \frac{1}{M_{1,n}} \sum_{m_1=1}^{M_{1,n}} \left(1 - \frac{L_1}{K_1 + 1}\right) \left(\frac{M_{2,n} - K_1}{M_{2,n}}\right)$$

$$= \left(1 - \frac{L_1}{K_1 + 1}\right) \left(\frac{M_{2,n} - K_1}{M_{2,n}}\right),$$

which implies the existence of at least one $m_1 \in \mathcal{M}_2$ and $\mathcal{J} \in \mathfrak{P}_{K_1}(\mathcal{M}_2)$ with

$$\mathbb{E}[\bar{e}_1(S_{m_1,\mathcal{J}}^n | \mathcal{C}_{\mathcal{W},\text{list}})] \geq \left(1 - \frac{L_1}{K_1 + 1}\right) \left(\frac{M_{2,n} - K_1}{M_{2,n}}\right).$$

Consequently, there is a realization s^n of $S_{m_1,\mathcal{J}}^n$ with $\bar{e}_1(s^n|\mathcal{C}_{\mathcal{W},\text{list}}) \geq \left(1 - \frac{L_1}{K_1+1}\right) \left(\frac{M_{2,n}-K_1}{M_{2,n}}\right)$ which finally implies

$$\max_{s^n \in \mathcal{S}^n} \bar{e}_1(s^n|\mathcal{C}_{\mathcal{W},\text{list}}) \geq \left(1 - \frac{L_1}{K_1 + 1}\right) \left(\frac{M_{2,n} - K_1}{M_{2,n}}\right)$$

so that the first part of the lemma is proved. Clearly, the analysis for node 2 follows accordingly using the same argumentation. ∎

REFERENCES

[1] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

[2] P. Larsson, N. Johansson, and K.-E. Sunell, "Coded bi-directional relaying," in *Proc. 5th Scandinavian Workshop Ad Hoc Netw.*, Stockholm, Sweden, May 2005, pp. 851–855.

[3] Y. Wu, P. Chou, and S.-Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," in *Proc. Conf. Inf. Sci. Syst.*, Baltimore, MD, USA, Mar. 2005, pp. 1–6.

[4] R. Knopp, "Two-way radio networks with a star topology," in *Proc. Int. Zurich Seminar Commun.*, Zürich, Switzerland, Feb. 2006, pp. 154–157.

[5] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Tsahkadsor, Armenian, Sep. 1971, pp. 23–52.

[6] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elect. Eng., Univ. Hawaii, Honolulu, HI, USA, 1972.

[7] T. J. Oechtering, C. Schnurr, I. Bjelaković, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 454–458, Jan. 2008.

[8] S. J. Kim, P. Mitran, and V. Tarokh, "Performance bounds for bidirectional coded cooperation protocols," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5235–5241, Nov. 2008.

[9] L.-L. Xie, "Network coding and random binning for multi-user channels," in *Proc. Can. Workshop Inf. Theory*, Jun. 2007, pp. 85–88.

[10] G. Kramer and S. Shamai (Shitz), "Capacity for classes of broadcast channels with receiver side information," in *Proc. IEEE Inf. Theory Workshop*, Tahoe City, CA, USA, Sep. 2007, pp. 313–318.

[11] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[12] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[13] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Anal.*, vol. 4, no. 1, pp. 371–386, 1960.

[14] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. New York, NY, USA: Springer-Verlag, 1978.

[15] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, no. 3, pp. 558–567, 1960.

[16] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 44, no. 2, pp. 159–175, 1978.

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

[18] T. Ericson, "Exponential error bounds for random codes in the arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 1, pp. 42–48, Jan. 1985.

[19] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, Mar. 1988.

[20] V. Blinovsky, P. Narayan, and M. Pinsker, "Capacity of the arbitrarily varying channel under list decoding," *Probl. Pered. Inform.*, vol. 31, no. 2, pp. 99–113, 1995.

[21] B. L. Hughes, "The smallest list for the arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 803–815, May 1997.

[22] A. D. Sarwate and M. Gastpar, "List-decoding for the arbitrarily varying channel under state constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1372–1384, Mar. 2012.

[23] A. D. Sarwate and M. Gastpar, "Rateless codes for AVC models," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3105–3114, Jul. 2010.

[24] S. C. Draper, F. R. Kschischang, and B. Frey, "Rateless coding for arbitrary channel mixtures with decoder channel state information," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4119–4133, Sep. 2009.

[25] R. F. Wyrembelski, I. Bjelaković, T. J. Oechtering, and H. Boche, "Optimal coding strategies for bidirectional broadcast channels under channel uncertainty," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2984–2994, Oct. 2010.

[26] T. J. Oechtering and M. Skoglund, "Bidirectional broadcast channel with random states noncausally known at the encoder," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 64–75, Jan. 2013.

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

[28] M. H. M. Costa, "Writing on dirty paper (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[29] J.-H. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Trans. Inf. Theory*, vol. 27, no. 2, pp. 212–226, Mar. 1981.

[30] J. A. Gubner, "On the deterministic-code capacity of the multiple-access arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 262–275, Mar. 1990.

[31] R. Ahlswede and N. Cai, "Arbitrarily varying multiple-access channels. I. Ericson's symmetrizability is adequate, Gubner's conjecture is true," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 742–749, Mar. 1999.

[32] S. Nitinawarat, "On the deterministic code capacity region of an arbitrarily varying multiple-access channel under list decoding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2683–2693, May 2013.

[33] E. Hof and S. I. Bross, "On the deterministic-code capacity of the two-user discrete memoryless arbitrarily varying general broadcast channel with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5023–5044, Nov. 2006.

[34] A. Winshtok and Y. Steinberg, "The arbitrarily varying degraded broadcast channel with states known at the encoder," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, USA, Jul. 2006, pp. 2156–2160.

[35] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user source coding—II," *J. Combinat. Inform. Syst. Sci.*, vol. 5, no. 3, pp. 220–268, 1980.

[36] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inf. Theory*, vol. 32, no. 5, pp. 621–629, Sep. 1986.

[37] R. Ahlswede and J. Wolfowitz, "The structure of capacity functions for compound channels," in *Proc. Int. Symp. Probab. Inf. Theory*, Apr. 1969, pp. 12–54.

[38] R. L. Dobrushin and S. Z. Stambler, "Coding theorems for classes of arbitrarily varying discrete memoryless channels," *Probl. Pered. Inform.*, vol. 11, no. 2, pp. 3–22, 1975.

[39] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 34, no. 2, pp. 157–177, 1976.

Rafael F. Schaefer (S'08–M'12) received his Dipl.-Ing. degree in electrical engineering and computer science in 2007 from the Technische Universität Berlin, Germany, and his Dr.-Ing. degree in electrical engineering in 2012 from the Technische Universität München, Germany. He worked as a research and teaching assistant at the Heinrich-Hertz-Lehrstuhl für Mobilkommunikation at the Technische Universität Berlin from 2007 to 2010 and at the Lehrstuhl für Theoretische Informationstechnik at the Technische Universität München from 2010 to 2013. He is currently a Postdoctoral Research Fellow with the Department of Electrical Engineering at Princeton University. In 2013, he received the VDE Johann-Philipp-Reis Award. He is one of the Exemplary Reviewers 2013 for the IEEE COMMUNICATION LETTERS.

Holger Boche (M'04–SM'07–F'11) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the Technische Universität Dresden, Dresden, Germany, in 1990 and 1994, respectively. He graduated in mathematics from the Technische Universität Dresden in 1992. From 1994 to 1997, he did Postgraduate studies in mathematics at the Friedrich-Schiller Universität Jena, Jena, Germany. He received his Dr. rer. nat. degree in pure mathematics from the Technische Universität Berlin, Berlin, Germany, in 1998. In 1997, he joined the Heinrich-Hertz-Institut (HHI) für Nachrichtentechnik Berlin, Berlin, Germany. Starting in 2002, he was a Full Professor for mobile communication networks with the Institute for Communications Systems, Technische Universität Berlin. In 2003, he became Director of the Fraunhofer German-Sino Lab for Mobile Communications, Berlin, Germany, and in 2004 he became the Director of the Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany. Since October 2010 he has been with the Institute of Theoretical Information Technology and Full Professor at the Technische Universität München, Munich, Germany. He was a Visiting Professor with the ETH Zurich, Zurich, Switzerland, during the 2004 and 2006 Winter terms, and with KTH Stockholm, Stockholm, Sweden, during the 2005 Summer term. Prof. Boche is a Member of IEEE Signal Processing Society SPCOM and SPTM Technical Committee. He was elected a Member of the German Academy of Sciences (Leopoldina) in 2008 and of the Berlin Brandenburg Academy of Sciences and Humanities in 2009. He received the Research Award "Technische Kommunikation" from the Alcatel SEL Foundation in October 2003, the "Innovation Award" from the Vodafone Foundation in June 2006, and the Gottfried Wilhelm Leibniz Prize from the Deutsche Forschungsgemeinschaft (German Research Foundation) in 2008. He was corecipient of the 2006 IEEE Signal Processing Society Best Paper Award and recipient of the 2007 IEEE Signal Processing Society Best Paper Award.