Optimal Coding Strategies for Bidirectional Broadcast Channels Under Channel Uncertainty

Rafael F. Wyrembelski, *Student Member, IEEE,* Igor Bjelaković, Tobias J. Oechtering, *Member, IEEE,* and Holger Boche, *Senior Member, IEEE*

***Abstract*—Bidirectional relaying is a promising approach to improve the performance in wireless networks such as sensor, ad-hoc, and even cellular systems. Bidirectional relaying applies to three-node networks, where a relay establishes a bidirectional communication between two other nodes using a decode-and-forward protocol. First, the two nodes transmit their messages to the relay which decodes them. Then, the relay broadcasts a re-encoded message in such a way that both nodes can decode their intended message using their own message as side information. We consider uncertainty in the channel state information (CSI) and assume that all nodes only know that the channel over which the transmission takes place is from a pre-specified set of channels. In this work, we concentrate on the second phase, which is called the *compound bidirectional broadcast channel*. We present a robust coding strategy which enables reliable communication under channel uncertainty and show that this strategy actually achieves the compound capacity. Further, we analyze scenarios where either the receivers or the transmitter have perfect CSI. We show that CSI at the receivers does not affect the maximal achievable rates, while CSI at the transmitter improves the capacity region. A numerical example and a game-theoretic interpretation complete this work.**

***Index Terms*—Relay, channel uncertainty, information rates, robustness, wireless networks.**

I. Introduction

FUTURE wireless communication systems will provide services that make great demands on throughput and coverage. For cellular systems this is a challenging task especially at the cell edges. Relays provide a promising approach to meet performance targets and are currently being intensively discussed by the Third Generation Partnership Program's Long-Term Evolution Advanced (3GPP LTE-Advanced) group. In this work, we consider the *bidirectional relay channel*, which is specified by a three-node network where a half-duplex relay node establishes a bidirectional communication between the two other nodes.

Paper approved by A. Yener, the Editor for Information Theory and Physical Layer Security of the IEEE Communications Society. Manuscript received March 23, 2009; revised October 26, 2009 and March 31, 2010.

This work was partly presented at IEEE-ICC, Dresden, Germany, June 2009.

The work of R. F. Wyrembelski and I. Bjelaković was supported by the German Research Foundation (DFG) under Grants BO 1734/25-1 and BO 1734/20-1. The work of T. J. Oechtering was supported by the European Commission through the FP7 project, FeedNetBack. The work of H. Boche was supported by the German Ministry of Education and Research (BMBF) under Grant 01BU0631.

R. F. Wyrembelski, I. Bjelaković, and H. Boche are with the Heinrich-Hertz-Lehrstuhl für Informationstheorie und theoretische Informationstechnik, Technische Universität Berlin, Berlin, Germany (e-mail: {rafael.wyrembelski, igor.bjelakovic, holger.boche}@mk.tu-berlin.de).

T. J. Oechtering is with the Communication Theory Lab, School of Electrical Engineering and the ACCESS Linnaeus Center, Royal Institute of Technology, Stockholm, Sweden (e-mail: Tobias.Oechtering@ee.kth.se).

Digital Object Identifier 10.1109/TCOMM.2010.082010.090170

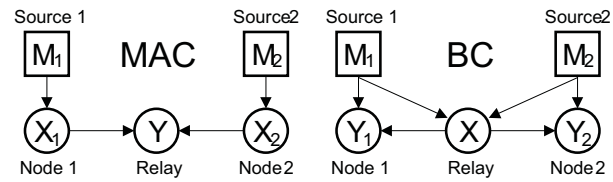


Fig. 1. Multiple access (MAC) and bidirectional broadcast (BBC) phase of the bidirectional relay channel.

There exist several strategies which are usually classified by the processing at the relay node, namely the amplify-and-forward strategy [1], [2], the compress-and-forward strategy [3], [4], and the decode-and-forward strategy [1], [5]. Here, we consider a spectrally efficient two-phase decode-and-forward protocol where cooperation between the encoders of nodes 1 and 2 is not allowed, cf. Fig. 1.

In the initial multiple access phase both nodes transmit their messages to the relay node. Since we assume the relay to decode the messages, we end up with the classical multiple access channel (MAC). Consequently, if the rates are chosen within the corresponding capacity region, it is reasonable to assume that the relay can successfully decode both messages so that it has perfect knowledge about the two messages the two nodes want to exchange.

Since each node knows its own transmitted message, in the succeeding broadcast phase it only remains for the relay to broadcast a re-encoded composition which allows each node to decode the message it is intended to receive using the message that it transmitted in 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, e.g., see [6], [7] for the multiple-input multiple-output broadcast channel. To emphasize the difference, the channel considered here is called *bidirectional broadcast channel* (BBC). We derived the optimal coding strategy for discrete memoryless channels with finite alphabets for perfect channel state information (CSI) at all nodes in [5]. There it is shown that the achievable rates for both users may differ even though both nodes receive the same codeword from the relay. The side information available at the receiving nodes creates this important difference between the optimal coding approach and the XOR coding approach [8] and the broadcast of a common message.

This shows that for the decode-and-forward strategy the two phases may be separated. The consequence is that we

0090-6778/10$25.00 © 2010 IEEE

have to distribute the available spectral resources between the two phases, which can be easily done in the time or frequency domain. As a result, we have to scale the achievable rate regions accordingly. The achievable rate region for the bidirectional relay channel is given by the intersection of the corresponding capacity regions for the MAC and the BBC phase. How to optimally divide the available resources among both phases is beyond the scope of this work. For a detailed discussion we refer to [9].

To date, bidirectional relaying has been analyzed under the assumption of perfect channel state information (CSI) at all nodes. However, due to the nature of the wireless channel, uncertainty in the channel state information is a ubiquitous phenomenon in practical systems. The traditional and most popular approach to mitigate the channel uncertainty is based on channel estimation. But there are other approaches to treat the problem of reliable communication under channel uncertainty. A survey can be found in [10]. For example, a robust power allocation for multi-antenna transmission using a game-theoretic approach is given in [11]. In [12] the performance of zero-forcing precoding and detection techniques under channel uncertainties is evaluated. A robust transceiver design based on convex optimization methods for the downlink in multiuser systems is presented in [13]. Clearly, these are only specific approaches and it seems natural to treat this problem from a more general point of view to gain insights to the best possible approach under channel uncertainty.

A well accepted model for channel uncertainty is to assume that the exact channel realization is not known to the nodes; rather, it is assumed that it is only known that a realization belongs to a pre-specified set of channels. If this channel remains fixed during the whole transmission of a codeword, this corresponds to the concept of the *compound channel* [14]–[16]. It seems worthwhile to study this model of channel uncertainty from an optimal coding perspective to gain an understanding of how robust coding strategies should be designed. This is especially important to know for wireless systems, which make strict demands on the quality of service. As an example, one can think of wireless control applications where certain rates have to be guaranteed regardless of the current channel realization. For such applications most performance measures, for example the ergodic capacity, are not appropriate, since they characterize rates which are only achievable on average. Rather, a performance measure is needed which characterizes the guaranteed rates. The concept of the compound channel is an attractive model which allows one to treat such problems from a general point of view and to obtain bounds on the maximal achievable rates. Moreover, this concept allows one to assess the resulting gain based on an improvement in the channel state information. This makes it possible to consider the trade-off between the contribution of CSI and the effort which would be needed to improve it, e.g., by using better or longer training sequences.

The analysis of bidirectional relaying for compound channels is not only relevant in itself since it yields results for bidirectional relaying in common communication scenarios as flat fading channels, but also since these results constitute the basis for further analysis of more complex uncertainty models as arbitrarily varying channels where the channel may vary during the transmission from symbol to symbol in an unknown and arbitrary manner as for example in fast fading channels. Bidirectional relaying is extended to this model of uncertainty in [17]. Moreover, it constitutes the basis for the analysis of multi-user settings in uncoordinated wireless networks where the receiving nodes are confronted with unknown varying interference [18].

As observed, the two phases of the decode-and-forward protocol decouple so that it suffices to analyze both phases under channel uncertainty separately. Since the compound MAC is well understood [19], [20], we concentrate on the BBC phase in this work. In Section II we briefly motivate and introduce the compound channel as a model for channel uncertainty and state some preliminaries in Section III. In Section IV we present a universal coding strategy, which allows one to communicate reliably in the case of channel uncertainty. In Sections V and VI, we discuss the cases of partial CSI, where either the receivers or the transmitter have perfect information about the channel state. It is shown that CSI at the receivers (CSIR) does not effect the maximal achievable rates, while CSI at the transmitter (CSIT) can be exploited to improve the capacity region. Then we present a numerical example which illustrates the gain in the capacity region based on available CSIT and give a short game-theoretic interpretation in Section VII. Finally, we end with a conclusion in Section VIII.

Notation

Discrete random variables are denoted by capital letters and their corresponding realizations and ranges by lower case letters and calligraphic letters respectively; $\mathbb{N}$ and $\mathbb{R}_+$ denote the set of natural and non-negative real numbers; $\mathbb{E}_X[\cdot]$ is the expectation with respect to X; $\mathcal{P}(\cdot)$ denotes the set of all probability distributions; $(\cdot)^c$ is the complement of a set.

The mutual information [21, p. 21] between the input random variable X and the output random variable Y is denoted by $I(X;Y)$. To emphasize the dependency of the mutual information on the input distribution p and the channel W we also write $I(X;Y) = I(p,W)$ interchangeably.

II. Bidirectional Broadcast Channel Under Channel Uncertainty

As a motivation we briefly recall the case where all nodes exactly know the channel over which the transmission takes place. For perfect CSI the capacity region and optimal coding strategy for the BBC with discrete memoryless channels and finite alphabets are known [5]. The capacity region can be characterized in terms of mutual information and is the set of all rate pairs $(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}_+^2$ which satisfy

$$R_{\mathrm{R1}} \leq I(X;Y_1|U) \qquad \text{and} \qquad R_{\mathrm{R2}} \leq I(X;Y_2|U), \tag{1}$$

where X denotes the input random variable, Y_k the output random variable at node k, $k = 1, 2$, and U an auxiliary random variable which describes a possible time-sharing operation. This result can be extended to Gaussian channels with a mean input power constraint for the case of multiple transmit and receive antennas as done in [22] and further analyzed in [23].

Remark 1: The necessity of the time-sharing variable U in (1) indicates that there are some rate pairs which are

only achievable by time-sharing between two other rate pairs. Especially in the case where the set of input distributions is not convex, the use of the time-sharing variable is beneficial and improves the achievable rates. Knowledge of such properties of the capacity region is indispensable for further cross-layer designs.

Perfect CSI, especially on the transmitter side, is a challenging task in wireless systems, so in general we are confronted with channel uncertainty, which should be taken into account if one designs wireless systems. We assume that the transmission takes place over a channel which is unknown to the transmitter and the receivers. It is only known to the nodes that this channel is from a pre-specified set of channels $\mathcal{S}$ and remains fixed during the whole transmission of a codeword. No restrictions are imposed on the set $\mathcal{S}$. Since $\mathcal{S}$ can be arbitrary, it includes, in particular, the case of infinitely many channels. Accordingly, this constitutes an appropriate model for common communication scenarios, for example flat fading channels.

Let $\mathcal{X}$ and $\mathcal{Y}_k$, $k = 1, 2$, be finite input and output sets. Then for a fixed $s \in \mathcal{S}$ and for input and output sequences $x^n \in \mathcal{X}^n$ and $y_k^n \in \mathcal{Y}_k^n$, $k = 1, 2$, of length n, the discrete memoryless broadcast channel is given by $W_s^{\otimes n}(y_1^n, y_2^n|x^n) := \prod_{i=1}^n W_s(y_{1,i}, y_{2,i}|x_i)$.

Definition 1: The discrete memoryless *compound broadcast channel* is the family $\{W_s^{\otimes n} : \mathcal{X}^n \to \mathcal{P}(\mathcal{Y}_1^n \times \mathcal{Y}_2^n)\}_{n\in\mathbb{N}, s\in\mathcal{S}}$. We simply denote the compound broadcast channel by $\mathcal{S}$.

Since we do not allow any cooperation between the receiving nodes, it is sufficient to consider the marginal transition probabilities $W_{k,s}^{\otimes n}(y_k^n|x^n) = \prod_{i=1}^n W_{k,s}(y_{k,i}|x_i)$, $k = 1, 2$, only. Thereby, $W_{k,s}$ denotes the channel between the relay and node k for channel realization $s \in \mathcal{S}$.

III. Preliminaries

In this work we use the concept of types and typical sequences from Csiszár and Körner [21], which will be a crucial technique to prove our results. We say a sequence $x^n \in \mathcal{X}^n$ is p-typical with constant δ or simply *typical* if $|\frac{1}{n}N(a|x^n) - p(a)| \leq \delta$ for all $a \in \mathcal{X}$ and, in addition, $N(a|x^n) = 0$ if $p(a) = 0$. Here, $N(a|x^n)$ denotes the number of indices i such that $x_i = a$, $i = 1, ..., n$. The set of typical sequences is denoted by $\mathcal{T}_{p,\delta}^n$. For $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ a sequence $y^n \in \mathcal{Y}^n$ is called W-*typical* under the condition $x^n \in \mathcal{X}^n$ if $|\frac{1}{n}N(a, b|x^n, y^n) - \frac{1}{n}N(a|x^n)W(b|a)| \leq \delta$ for all $b \in \mathcal{Y}$ and, in addition, $N(a, b|x^n, y^n) = 0$ if $W(b|a) = 0$. The set of such sequences is denoted by $\mathcal{T}_{W,\delta}^n(x^n)$. Further, we denote the probability of all typical sequences by $p^{\otimes n}(\mathcal{T}_{p,\delta}^n) = \sum_{x^n\in\mathcal{T}_{p,\delta}^n} p^{\otimes n}(x^n)$. The notation extends to W-typical sequences in a self-explanatory way, i.e., $W^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n)|x^n) = \sum_{y^n\in\mathcal{T}_{W,\delta}^n(x^n)} W^{\otimes n}(y^n|x^n)$. Moreover, we need the following lemmas which give us exponential rates of convergence for typical sequences.

Lemma 1: For every $\delta > 0$ and every $p \in \mathcal{P}(\mathcal{A})$ the following

$$p^{\otimes n}(\mathcal{T}_{p,\delta}^n) \geq 1 - (n+1)^{|\mathcal{A}|}2^{-nc\delta^2} \tag{2}$$

with $c = \frac{1}{2\ln 2}$ always holds for all $n \in \mathbb{N}$.

Proof: The proof can be found in [24, Lemma III.1.3] and is therefore omitted. ■

Lemma 2: For every $\delta > 0$ and every $x^n \in \mathcal{A}^n$, $W : \mathcal{A} \to \mathcal{P}(\mathcal{B})$ the following

$$W^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n)|x^n) \geq 1 - (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-nc\delta^2} \tag{3}$$

with $c = \frac{1}{2\ln 2}$ always holds for all $n \in \mathbb{N}$.

Proof: The proof follows [24, Lemma III.1.3] and can be found in Appendix A. ■

The next lemma relates typical sequences generated by different distributions to the same output distribution and plays a crucial role especially in the proof of the optimal coding strategy for the case of CSIT in Section VI.

Lemma 3: Let $p, \tilde{p} \in \mathcal{P}(\mathcal{A})$ be input distributions, $W, \tilde{W} : \mathcal{A} \to \mathcal{P}(\mathcal{B})$, and $q, \tilde{q} \in \mathcal{P}(\mathcal{B})$ the corresponding output distributions. Further, let $\delta \in (0, \frac{1}{4|\mathcal{A}||\mathcal{B}|})$. Then for every $n \in \mathbb{N}$ and all $\tilde{x}^n \in \mathcal{T}_{\tilde{p},\delta}^n$ and $x^n \in \mathcal{T}_{p,\delta}^n$ it holds

$$q^{\otimes n}(\mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)) \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-n(I(\tilde{p},\tilde{W})-\varphi(\delta)-\psi(\delta))}, \tag{4a}$$

$$q^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n)) \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-n(I(p,W)-\varphi(\delta)-\psi(\delta))} \tag{4b}$$

with universal $\varphi(\delta), \psi(\delta) > 0$ and $\lim_{\delta\searrow 0}\varphi(\delta) = \lim_{\delta\searrow 0}\psi(\delta) = 0$.

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

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

Definition 2: A $(M_1^{(n)}, M_2^{(n)}, n)$-*code* for the compound BBC consists of one universal encoder at the relay node $x^n : \mathcal{M} \to \mathcal{X}^n$ and universal decoders at nodes 1 and 2 given by the decoding sets $D_{m_2|m_1}^{(1)} \subseteq \mathcal{Y}_1^n$ and $D_{m_1|m_2}^{(2)} \subseteq \mathcal{Y}_2^n$ for all $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$. For given $m_1 \in \mathcal{M}_1$ at node 1 the decoding sets have to be disjoint, i.e., $D_{m_2|m_1}^{(1)} \cap D_{\hat{m}_2|m_1}^{(1)} = \emptyset$ for $\hat{m}_2 \neq m_2$, and similarly for given $m_2 \in \mathcal{M}_2$ at node 2 the decoding sets have to be disjoint, i.e., $D_{m_1|m_2}^{(2)} \cap D_{\hat{m}_1|m_2}^{(2)} = \emptyset$ for $\hat{m}_1 \neq m_1$.

This allows us to define the average probabilities of error for channel realization $s \in \mathcal{S}$ as

$$\mu_{1,s}^{(n)} := \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W_{1,s}^{\otimes n}\big((D_{m_2|m_1}^{(1)})^c|x^n(m)\big), \tag{5a}$$

$$\mu_{2,s}^{(n)} := \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W_{2,s}^{\otimes n}\big((D_{m_1|m_2}^{(2)})^c|x^n(m)\big) \tag{5b}$$

at nodes 1 and 2, respectively. Further, we define $\mu_k^{(n)} = \sup_{s\in\mathcal{S}}\mu_{k,s}^{(n)}$, $k = 1, 2$.

Definition 3: A rate pair $(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}_+^2$ is said to be *achievable* for the compound BBC if there is a sequence of $(M_1^{(n)}, M_2^{(n)}, n)$-codes such that $\liminf_{n\to\infty}\frac{1}{n}\log M_1^{(n)} \geq R_{\mathrm{R2}}$ and $\liminf_{n\to\infty}\frac{1}{n}\log M_2^{(n)} \geq R_{\mathrm{R1}}$ while $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. The set of all achievable rate pairs is the *capacity region* of the compound BBC and is denoted by $\mathcal{C}_{\mathrm{BBC}}$.

Remark 2: The definitions require that we have to find codes such that $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$ for all channels in the set $\mathcal{S}$ simultaneously. This means the codes are universal with respect to the channel realization.

IV. Universal Strategy

Now we are in a position to present the universal strategy for the second phase of the decode-and-forward protocol which overcomes the channel uncertainty at the transmitter and the receivers. But, first, we prove an outer bound of the capacity region which gives us an intuition to what is at best possible for the compound BBC.

Lemma 4: Any given sequence of $(M_1^{(n)}, M_2^{(n)}, n)$-codes with $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ must satisfy

$$\frac{1}{n}\log M_1^{(n)} \leq \inf_{s\in\mathcal{S}} I(p, W_{2,s}|U) + o(n^0), \tag{6a}$$

$$\frac{1}{n}\log M_2^{(n)} \leq \inf_{s\in\mathcal{S}} I(p, W_{1,s}|U) + o(n^0) \tag{6b}$$

for joint probability distributions $\{q(u)p(x|u)W_s(y_1, y_2|x)\}_{s\in\mathcal{S}}$.

Proof: From [5] we know that for a specific channel realization $s \in \mathcal{S}$ the rates are bounded from above by $\frac{1}{n}H(M_2) \leq I(X;Y_1|U) + \epsilon_1^{(n)} = I(p, W_{1,s}|U) + \epsilon_1^{(n)}$ and $\frac{1}{n}H(M_1) \leq I(X;Y_2|U) + \epsilon_2^{(n)} = I(p, W_{2,s}|U) + \epsilon_2^{(n)}$ where $\epsilon_1^{(n)}, \epsilon_2^{(n)} \to 0$ as $n \to \infty$. Since the rates have to be achievable for all $s \in \mathcal{S}$ simultaneously, it follows immediately that for the compound BBC with channel uncertainty at the transmitter and the receivers the rates are bounded from above by the infimum of the mutual information terms as stated in (6). This proves the lemma. ■

Next, we consider the case where the compound channel has finitely many elements and derive the corresponding capacity region in Section IV-A. Using this result we are able to solve the general case of an arbitrary, not necessarily finite, index set in Section IV-B, which is the more relevant case since it covers communication scenarios as for example flat fading channels.

A. Finite Compound Channel

In this section we restrict $\mathcal{S}$ to be finite and present a universal strategy which actually achieves the rates stated in (6) under this condition. Therefore we need the following lemma which shows the existence of such a strategy whose probability of error is arbitrarily small.

Lemma 5: Let the index set $\mathcal{S} = \{1, ..., S\}$ denote a finite compound BBC. For any block length $n \in \mathbb{N}$, input distribution p, and $R_{\mathrm{R}1} \leq \min_{i=1,\dots,S} I(p, W_{1,i}) - \frac{\tau}{2}$, $R_{\mathrm{R}2} \leq \min_{i=1,\dots,S} I(p, W_{2,i}) - \frac{\tau}{2}$, $\tau > 0$, there is a $(M_1^{(n)}, M_2^{(n)}, n)$-code where the probability of error $\mu_k^{(n)}$ at node k, $k = 1, 2$, averaged over all codebooks is bounded from above by[1]

$$\mathbb{E}_{X^n}[\mu_k^{(n)}] \leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_k|}2^{-nc\delta^2} + S^2\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_k|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}}2^{-n\frac{\tau}{4}} \tag{7}$$

with $c = \frac{1}{2\ln 2}$.

Proof: For given distribution $p^{\otimes n}(x^n)$ and $\delta > 0$ let $\mathcal{T}_{p,\delta}^n$ denote the set of typical sequences on $\mathcal{X}^n$. We restrict the possible inputs to this subset and define the new input distribution

$$p^n(x^n) := \begin{cases} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}_{p,\delta}^n)} & \text{if } x^n \in \mathcal{T}_{p,\delta}^n, \\ 0 & \text{else.} \end{cases} \tag{8}$$

Let $\tau > 0$ and set the rates $R_{\mathrm{R}1} := \min_{i=1,\dots,S} I(p, W_{1,i}) - \frac{\tau}{2}$ and $R_{\mathrm{R}2} := \min_{i=1,\dots,S} I(p, W_{2,i}) - \frac{\tau}{2}$. Then we generate $M_1^{(n)}M_2^{(n)}$ independent codewords $X^n(m)$, one for each $m = (m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, of length n with $M_1^{(n)} := \lfloor 2^{nR_{\mathrm{R}2}} \rfloor$ and $M_2^{(n)} := \lfloor 2^{nR_{\mathrm{R}1}} \rfloor$ according to p^n. This implies that all generated random codewords $X^n(m) \in \mathcal{T}_{p,\delta}^n$ almost surely.

Next, we specify the decoding sets of nodes 1 and 2 in detail. They are given by (9a) and (9b), at the top of the next page, where $X^n := \{X^n(m_1, m_2)\}_{1\leq m_1 \leq M_1^{(n)}, 1 \leq m_2 \leq M_2^{(n)}}$ so that the decoding sets depend on all generated codewords. The definition of the decoding sets is motivated as follows. The first part in (9a) ensures that for given $m = (m_1, m_2)$ the decoding sets for node 1 are mostly defined by all output sequences that are $W_{1,s}$-typical under the input $X^n(m_1, m_2)$ for all channel realizations $s \in \mathcal{S}$. The second part excludes all such output sequences that are $W_{1,s}$-typical to another input $X^n(m_1, j)$ with $j \neq m_2$ so that the decoding sets are unambiguously defined. Clearly, the decoding sets in (9b) are motivated accordingly.

When $x^n(m)$ 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 decoder at node 1 is in error if either y_1^n is not in $\bigcup_{i=1}^S \mathcal{T}_{W_{1,i},\delta}^n(x^n(m))$ or if y^n is in $\bigcup_{i=1}^S \mathcal{T}_{W_{1,i},\delta}^n(x^n(m_1, \hat{m}_2))$ with $\hat{m}_2 \neq m_2$, cf. (9a). The error events at node 2 are defined in an analogous way.

In the following we present the analysis of the probability of error for node 1, the analysis for node 2 follows accordingly using the same arguments. For a given channel realization $s \in \mathcal{S}$ the union bound yields for average probability of error $\mu_{1,s}^{(n)} \leq P_{e,1,s}^{(1)} + P_{e,2,s}^{(1)}$ with (10a) and (10b) (see top of next page).

Next, we average over all codebooks and show that $\mathbb{E}_{X^n}[\mu_{1,s}^{(n)}] \leq \mathbb{E}_{X^n}[P_{e,1,s}^{(1)} + P_{e,2,s}^{(1)}]$ can be bounded uniformly in s from above by a term which decreases exponentially fast for increasing block length n. For fixed $s \in \mathcal{S}$ we get for (10a), see (11) on the next page, with $c = \frac{1}{2\ln 2}$ where the first inequality follows from the monotonicity of the probability and the last one from Lemma 2, cf. (3). For the second error event (10b) we have (12) on the next page, where the last equality follows from the fact that X^n is an iid sequence. Next, we compute the expectations. For the inner expectation we get (13), see next page, where $c = \frac{1}{2\ln 2}$ and q_s denotes the output distribution generated by p and $W_{1,s}$. The second equality follows from (8) and the last inequality follows from Lemma 1, cf. (2). Since $X^n(m_1, j) \in \mathcal{T}_{p,\delta}^n$ almost surely, we can apply Lemma 3, more precisely (4b), and obtain for the second expectation

$$\mathbb{E}_{X^n(m_1,j)}\left[q_s^{\otimes n}\left(\mathcal{T}_{W_{1,i},\delta}^n\big(X^n(m_1,j)\big)\right)\right] \leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-n(I(p,W_{1,i})-\varphi(\delta)-\psi(\delta))} \tag{14}$$

for $\delta \in (0, \frac{1}{4|\mathcal{X}||\mathcal{Y}_1|})$. From (12)–(14) we get (15), see next page. Since $M_2^{(n)} = \lfloor 2^{nR_{\mathrm{R}1}} \rfloor$ and $R_{\mathrm{R}1} =$

[1]Without loss of generality we can assume that $R_{\mathrm{R}k} > 0$, $k = 1, 2$, since the rate $R_{\mathrm{R}k} = 0$ is always achievable.

$$D^{(1)}_{m_2|m_1}(X^n) := \Big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m)\big)\Big) \cap \Big(\bigcup_{\substack{j=1\\ j\neq m_2}}^{M_2^{(n)}} \bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)\Big)^c \tag{9a}$$

$$D^{(2)}_{m_1|m_2}(X^n) := \Big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{2,i},\delta}\big(X^n(m)\big)\Big) \cap \Big(\bigcup_{\substack{j=1\\ j\neq m_1}}^{M_1^{(n)}} \bigcup_{i=1}^{S} \mathcal{T}^n_{W_{2,i},\delta}\big(X^n(j,m_2)\big)\Big)^c \tag{9b}$$

$$P^{(1)}_{e,1,s} := \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W^{\otimes n}_{1,s}\Big(\big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m)\big)\big)^c \big| X^n(m)\Big) \tag{10a}$$

$$P^{(1)}_{e,2,s} := \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} W^{\otimes n}_{1,s}\Big(\bigcup_{\substack{j=1\\ j\neq m_2}}^{M_2^{(n)}} \bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big) \big| X^n(m_1,m_2)\Big) \tag{10b}$$

$$\begin{aligned}
\mathbb{E}_{X^n}[P^{(1)}_{e,1,s}] &= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathbb{E}_{X^n}\Big[W^{\otimes n}_{1,s}\Big(\big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m)\big)\big)^c \big| X^n(m)\Big)\Big]\\
&= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathbb{E}_{X^n}\Big[W^{\otimes n}_{1,s}\Big(\bigcap_{i=1}^{S} {\mathcal{T}^n_{W_{1,i},\delta}}^c\big(X^n(m)\big) \big| X^n(m)\Big)\Big]\\
&\le \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathbb{E}_{X^n}\Big[W^{\otimes n}_{1,s}\Big({\mathcal{T}^n_{W_{1,s},\delta}}^c\big(X^n(m)\big) \big| X^n(m)\Big)\Big]\\
&\le (n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-nc\delta^2}
\end{aligned} \tag{11}$$

$$\begin{aligned}
\mathbb{E}_{X^n}[P^{(1)}_{e,2,s}] &= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathbb{E}_{X^n}\Big[W^{\otimes n}_{1,s}\Big(\bigcup_{\substack{j=1\\ j\neq m_2}}^{M_2^{(n)}} \bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big) \big| X^n(m_1,m_2)\Big)\Big]\\
&\le \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \sum_{\substack{j=1\\ j\neq m_2}}^{|\mathcal{M}_2|} \sum_{i=1}^{S} \mathbb{E}_{X^n}\Big[W^{\otimes n}_{1,s}\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)|X^n(m_1,m_2)\big)\Big)\Big]\\
&= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \sum_{\substack{j=1\\ j\neq m_2}}^{|\mathcal{M}_2|} \sum_{i=1}^{S} \mathbb{E}_{X^n(m_1,j)}\mathbb{E}_{X^n(m_1,m_2)}\Big[W^{\otimes n}_{1,s}\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)|X^n(m_1,m_2)\Big)\Big]
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\mathbb{E}_{X^n(m_1,m_2)}\Big[W^{\otimes n}_{1,s}\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)|X^n(m_1,m_2)\Big)\Big] = \sum_{x^n\in\mathcal{X}^n} p^n(x^n) W^{\otimes n}_{1,s}\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)|x^n\Big)\\
&= \sum_{x^n\in\mathcal{T}^n_{p,\delta}} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}^n_{p,\delta})} W^{\otimes n}_{1,s}\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)|x^n\Big) \le \sum_{x^n\in\mathcal{X}^n} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}^n_{p,\delta})} W^{\otimes n}_{1,s}\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)|x^n\Big)\\
&= \frac{1}{p^{\otimes n}(\mathcal{T}^n_{p,\delta})} q^{\otimes n}_s\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)\Big) \le \frac{1}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}} q^{\otimes n}_s\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n(m_1,j)\big)\Big)
\end{aligned} \tag{13}$$

$$\mathbb{E}_{X^n}[P^{(1)}_{e,2,s}] \le (M_2^{(n)}-1)\sum_{i=1}^{S} \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}} 2^{-n(I(p,W_{1,i})-\varphi(\delta)-\psi(\delta))} \tag{15}$$

$\min_{i=1,...,S} I(p, W_{1,i}) - \frac{\tau}{2}$ we have $\frac{\tau}{2} \leq I(p, W_{1,i}) - R_{\mathrm{R1}}$ for all $i = 1, ..., S$ so that for (15) we obtain $\mathbb{E}_{X^n}[P_{e,2,s}^{(1)}] \leq S\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}}2^{-n(\frac{\tau}{2}-\varphi(\delta)-\psi(\delta))}$. Next, we set $\delta \in (0, \frac{1}{4|\mathcal{X}||\mathcal{Y}_1|})$ small enough to ensure that $\frac{\tau}{4} \leq \frac{\tau}{2} - \varphi(\delta) - \psi(\delta)$ so that with (11) we get

$$\begin{aligned}\mathbb{E}_{X^n}[\mu_{1,s}^{(n)}] &\leq \mathbb{E}_{X^n}[P_{e,1,s}^{(1)} + P_{e,2,s}^{(1)}] \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-nc\delta^2} + S\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}}2^{-n\frac{\tau}{4}}\end{aligned} \tag{16}$$

with $c = \frac{1}{2\ln 2}$. Now, (16) allows us to bound the average probability of error $\mu_{1,s}^{(n)}$ for channel realization $s \in \mathcal{S}$. To obtain an upper bound for $\mu_1^{(n)}$ we need the average BBC $W_{1,av} := \frac{1}{S}\sum_{i=1}^{S} W_{1,i}$. From the definition it is clear that $W_{1,av} \geq \frac{1}{S}W_{1,i}$ holds for all $i = 1, ..., S$ which implies that $\mu_1^{(n)} = \max_{s\in\mathcal{S}} \mu_{1,s}^{(n)} \leq S\mu_{1,av}^{(n)}$ where $\mu_{1,av}^{(n)}$ is the average probability of error with respect to the average BBC so that $\mathbb{E}_{X^n}[\mu_{1,av}^{(n)}] = \frac{1}{S}\sum_{s\in\mathcal{S}}\mathbb{E}_{X^n}[\mu_{1,s}^{(n)}]$ due to the linearity of the expectation. Finally, we obtain for the average probability of error at node 1

$$\begin{aligned}\mathbb{E}_{X^n}[\mu_1^{(n)}] &\leq S\mathbb{E}_{X^n}[\mu_{1,av}^{(n)}] \\ &\leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}2^{-nc\delta^2} + S^2\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_1|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}}2^{-n\frac{\tau}{4}}\end{aligned} \tag{17}$$

with $c = \frac{1}{2\ln 2}$ as stated in (7). Similar reasoning leads for the probability of error at node 2 to $\mathbb{E}_{X^n}[\mu_2^{(n)}] \leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_2|}2^{-nc\delta^2} + S^2\frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_2|}}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}}2^{-n\frac{\tau}{4}}$ which proves the lemma. ■

The crucial point of this lemma is that the concept of typical sequences from Csiszár and Körner [21] allows one to establish bounds on the probability of error that decrease exponentially fast for increasing block length. This property will be important for the extension to the general case of an arbitrary set $\mathcal{S}$. However, this lemma together with Lemma 4 immediately leads to the capacity region of the finite compound BBC, which is stated in the following corollary.

Corollary 1: The capacity region $\mathcal{C}_{\mathrm{BBC}}$ of the finite compound BBC $\mathcal{S} = \{1, ..., S\}$ is the set of all rate pairs $(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}_+^2$ satisfying

$$R_{\mathrm{R1}} \leq \min_{i=1,..,S} I(p, W_{1,i}|U), \tag{18a}$$

$$R_{\mathrm{R2}} \leq \min_{i=1,..,S} I(p, W_{2,i}|U) \tag{18b}$$

for random variables $(U, X, Y_1, Y_2) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$ and joint probability distributions $\{q(u)p(x|u)W_s(y_1, y_2|x)\}_{s\in\mathcal{S}}$. Thereby, U is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of U can be bounded by $|\mathcal{U}| \leq 2$.

Proof: The universal coding strategy follows immediately from Lemma 5 which states that all rate pairs $(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}_+^2$ satisfying $R_{\mathrm{R1}} \leq \min_{i=1,..,S} I(p, W_{1,i})$, $R_{\mathrm{R2}} \leq \min_{i=1,..,S} I(p, W_{2,i})$ are achievable with $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. The desired region (18) is determined by establishing the convex hull by first introducing an auxiliary random variable U and applying standard arguments. Similarly to [5] it follows from Fenchel-Bunt's extension of Carathéodory's theorem [25] that any rate pair is achievable by time-sharing between two rate pairs, i.e., $|\mathcal{U}| = 2$ is enough.

The optimality of this strategy follows from Lemma 4. Since the strategy from Lemma 5 already achieves these rate pairs, the capacity region of the finite compound BBC is completely characterized by the corollary. ■

B. Arbitrary Compound Channel

With the previous result we are able to establish the capacity region for the compound BBC with an arbitrary set $\mathcal{S}$. Therefore we need the following two lemmas which are slightly adapted from [14] to our scenario.

Lemma 6: Let $\mathcal{X}$, $\mathcal{Y}_k$, $k = 1, 2$, be given. For every integer $L \geq 2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2$ there is a compound broadcast channel $\mathcal{S}_L$ with at most $(L+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}$ elements such that for any W_s from $\mathcal{S}$ there is a channel $\overline{W}_s$ from $\mathcal{S}_L$ such that

(a) $|W_s(y_1, y_2|x) - \overline{W}_s(y_1, y_2|x)| \leq \frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{L}$ for all x, y_1, y_2

(b) $W_s(y_1, y_2|x) \leq 2^{\frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{L}}\overline{W}_s(y_1, y_2|x)$ for all x, y_1, y_2

(c) For any $p \in \mathcal{P}(\mathcal{X})$ it holds $|I(p, W_{k,s}) - I(p, \overline{W}_{k,s})| \leq 2|\mathcal{Y}_1||\mathcal{Y}_2|(\frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{L})^{\frac{1}{2}}$, $k = 1, 2$.

Proof: The proof is almost identical to Lemma 4 in [14] and is therefore omitted. ■

This lemma shows that we can approximate any given set of channels $\mathcal{S}$ by a finite number of channels $\mathcal{S}_L$ such that any channel $s \in \mathcal{S}$ is close in several senses to one of the new constructed channels in $\mathcal{S}_L$. Further, from the next lemma we see that if there is a "good" code for a channel, then the same code can be used for all channels in a certain neighborhood of this channel.

Lemma 7: Let W_s and $\overline{W}_s$ be two channels and A a non-negative number such that $W_s(y_1, y_2|x) \leq 2^A\overline{W}_s(y_1, y_2|x)$ for all x, y_1, y_2. Then any $(M_1^{(n)}, M_2^{(n)}, n)$-code for W_s is also a $(M_1^{(n)}, M_2^{(n)}, n)$-code for $\overline{W}_s$ with $\mu_k^{(n)} \leq 2^{nA}\bar{\mu}_k^{(n)}$, $k = 1, 2$.

Proof: The proof is almost identical to Lemma 5 in [14] and is therefore omitted. ■

With these two lemmas and the result for the case of a finite set, we are able to prove our main result which is the capacity region of the compound BBC with an arbitrary set $\mathcal{S}$.

Theorem 1: The capacity region $\mathcal{C}_{\mathrm{BBC}}$ of the compound BBC $\mathcal{S}$, where $\mathcal{S}$ can be any arbitrary set, is the set of all rate pairs $(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}_+^2$ satisfying

$$R_{\mathrm{R1}} \leq \inf_{s\in\mathcal{S}} I(p, W_{1,s}|U), \quad R_{\mathrm{R2}} \leq \inf_{s\in\mathcal{S}} I(p, W_{2,s}|U) \tag{19}$$

for random variables $(U, X, Y_1, Y_2) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$ and joint probability distributions $\{q(u)p(x|u)W_s(y_1, y_2|x)\}_{s\in\mathcal{S}}$. Thereby, U is an auxiliary random variable and describes a possible time-sharing operation. The cardinality of the range of U can be bounded by $|\mathcal{U}| \leq 2$.

Proof: We start with an approximation of the arbitrary set of channels. Therefore we choose $L \geq \max\{\frac{|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\delta^2}, \frac{8|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\tau}\}$ and large enough to ensure that $\frac{\tau}{2} > 2|\mathcal{Y}_1||\mathcal{Y}_2|(\frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{L})^{\frac{1}{2}}$. For each W_s from $\mathcal{S}$ we select a $\overline{W}_s$ according to Lemma 6 and denote the set of approximated channels by $\mathcal{S}_L$. Since $\mathcal{S}_L$ has at most $(L+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}$ elements, we know from Lemma 5 that if we choose $R_{\mathrm{R1}} \leq \min_{i\in\mathcal{S}_L} I(p, \overline{W}_{1,i}) - \frac{\tau}{2}$ and $R_{\mathrm{R2}} \leq \min_{i\in\mathcal{S}_L} I(p, \overline{W}_{2,i}) - \frac{\tau}{2}$,

$\tau > 0$, then there exists a $(M_1^{(n)}, M_2^{(n)}, n)$-code with $M_1^{(n)} = \lfloor 2^{nR_{\mathrm{R2}}} \rfloor$ and $M_2^{(n)} = \lfloor 2^{nR_{\mathrm{R1}}} \rfloor$ for $\mathcal{S}_L$ with probability of error for node k, $k = 1, 2$

$$\bar{\mu}_k^{(n)} \leq S_L (n+1)^{|\mathcal{X}||\mathcal{Y}_k|} 2^{-nc\delta^2} + S_L^2 \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_k|}}{1-(n+1)^{|\mathcal{X}|} 2^{-nc\delta^2}} 2^{-n\frac{\tau}{4}} \tag{20}$$

with $S_L = (L+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}$ and $c = \frac{1}{2\ln 2}$. For each W_s from $\mathcal{S}$ there exists a $\overline{W}_s$ from $\mathcal{S}_L$ such that $W_s(y_1, y_2|x) \leq 2^{\frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{L}} \overline{W}_s(y_1, y_2|x)$ for all x, y_1, y_2 so that from Lemma 7 the code for $\mathcal{S}_L$ is also a code for $\mathcal{S}$ with

$$\begin{aligned} \mu_k^{(n)} &\leq S_L (n+1)^{|\mathcal{X}||\mathcal{Y}_k|} 2^{-n(c\delta^2 - \frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{L})} \\ &\quad + S_L^2 \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_k|}}{1-(n+1)^{|\mathcal{X}|} 2^{-nc\delta^2}} 2^{-n(\frac{\tau}{4} - \frac{2|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{L})}. \end{aligned} \tag{21}$$

Since $L \geq \max\{\frac{|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\delta^2}, \frac{8|\mathcal{Y}_1|^2|\mathcal{Y}_2|^2}{\tau}\}$, we have $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ as $n \to \infty$. This means the code constructed for the approximated channel is also a good code for the original channel. It remains to show that the code achieves rates arbitrarily close to the desired rates. From Lemma 6 we know that $|I(p, W_{k,s}) - I(p, \overline{W}_{k,s})| \leq 2|\mathcal{Y}_1||\mathcal{Y}_2|(\frac{|\mathcal{Y}_1||\mathcal{Y}_2|}{L})^{\frac{1}{2}} \leq \frac{\tau}{2}$, $k = 1, 2$ so that

$$\inf_{s \in \mathcal{S}} I(p, W_{k,s}) - \tau \leq \min_{s \in \mathcal{S}_L} I(p, \overline{W}_{k,s}) - \frac{\tau}{2} \tag{22}$$

which proves the achievability of the rates given in (19). The optimality of the strategy follows similarly to Corollary 1 from Lemma 4 which finally proves the theorem. ■

V. CSI at the Receivers

Next, we consider the scenario where the receivers have perfect CSI so that they can adapt their decoder to the specific channel realization. Consequently, we now have a whole family of decoders at nodes 1 and 2, one for each channel realization $s \in \mathcal{S}$. Note that we still only have one universal encoder at the relay due to the channel uncertainty at the transmitter. The following theorem shows that CSIR does not lead to an improved capacity region as long as the transmitter merely knows the set of channels.

Theorem 2: The capacity region $\mathcal{C}_{\mathrm{BBC}}^{\mathrm{CSIR}}$ of the compound BBC with CSIR is equal to the capacity region of the compound BBC with channel uncertainty at all nodes, i.e., $\mathcal{C}_{\mathrm{BBC}}^{\mathrm{CSIR}} = \mathcal{C}_{\mathrm{BBC}}$.

Proof: If we apply the coding strategy for channel uncertainty at all nodes, cf. Theorem 1, it is clear that we can achieve the same rate pairs if we have perfect CSI at the receivers. Consequently, it remains to show that this strategy is already optimal which means that no other rate pairs are achievable. The reasoning is as follows. In our communication scenario we have the following Markov chains $(M_1, M_2) \to X \to Y_1 \to \hat{M}_2$ and $(M_1, M_2) \to X \to Y_2 \to \hat{M}_1$ for nodes 1 and 2, respectively, where $\hat{M}_k$, $k = 1, 2$ denotes the decoded message. From the data processing inequality [21, Lemma 3.11] it follows immediately that $I((M_1, M_2); \hat{M}_1) \leq I(X; Y_2)$ and $I((M_1, M_2); \hat{M}_2) \leq I(X; Y_1)$ which shows that the decoder does not effect the achievable rate. This permits the proof of the optimality of the universal strategy similarly to the case of channel uncertainty at all nodes, cf. Theorem 1. ■

Remark 3: An intuitive explanation of why CSIR does not lead to an improved capacity region is indicated in [16]. Even if the channel used for the transmission is not known to the receivers, it can be estimated with arbitrary accuracy by the receivers. For sufficiently large block length n the part "wasted" for the estimation is a negligible part of n, and goes to zero as $n \to \infty$.

VI. CSI at the Transmitter

Here, the transmitter has perfect CSI so that it can adapt its encoder to the specific channel realization. Consequently, we now have a whole family of encoders at the relay node $x_s^n : \mathcal{M} \to \mathcal{X}^n$, one for each channel realization $s \in \mathcal{S}$. Note that we still have universal decoding sets.

The following shows that CSIT leads to an improved capacity region of the compound BBC, which is in contrast to the previous discussed case of CSIR. The introduced concept of types and typical sequences from Csiszár and Körner [21] permits a proof of the capacity region which is quite similar to the case of channel uncertainty at all nodes. Consequently, we concentrate on the crucial points where the reasoning differs from the derivation in Section IV. Similarly, we first present an outer bound of the capacity region to get an intuition what is at best possible with CSIT.

Lemma 8: Any given sequence of $(M_1^{(n)}, M_2^{(n)}, n)$-codes with $\mu_1^{(n)}, \mu_2^{(n)} \to 0$ must satisfy

$$\frac{1}{n} \log M_2^{(n)} \leq \inf_{s \in \mathcal{S}} I(p_s, W_{1,s}|U) + o(n^0), \tag{23a}$$

$$\frac{1}{n} \log M_1^{(n)} \leq \inf_{s \in \mathcal{S}} I(p_s, W_{2,s}|U) + o(n^0) \tag{23b}$$

for joint probability distributions $\{q(u)p_s(x|u)W_s(y_1, y_2|x)\}_{s \in \mathcal{S}}$.

Proof: The proof of (23) is similar to the proof of Lemma 4 and therefore omitted for brevity. ■

Next, we present a universal strategy which actually achieves the rates stated in the previous lemma. The crucial point is to establish an upper bound on the probability of error for the case of a finite set $\mathcal{S}$ similar to the one given in Lemma 5. Then the rest of the proof of the capacity region follows accordingly.

Lemma 9: Let the index set $\mathcal{S} = \{1, ..., S\}$ denote a finite compound BBC with CSIT. For any block length $n \in \mathbb{N}$, input distributions p_i, $i = 1, ..., S$, and $R_{\mathrm{R1}} \leq \min_{i=1,..,S} I(p_i, W_{1,i}) - \frac{\tau}{2}$, $R_{\mathrm{R2}} \leq \min_{i=1,..,S} I(p_i, W_{2,i}) - \frac{\tau}{2}$, $\tau > 0$, there is a $(M_1^{(n)}, M_2^{(n)}, n)$-code where the probability of error $\mu_k^{(n)}$ at node k, $k = 1, 2$, averaged over all codebooks is bounded from above by

$$\begin{aligned} \mathbb{E}_{X^n}[\mu_k^{(n)}] &\leq S(n+1)^{|\mathcal{X}||\mathcal{Y}_k|} 2^{-nc\delta^2} \\ &+ S^2 \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_k|}}{1-(n+1)^{|\mathcal{X}|} 2^{-nc\delta^2}} 2^{-n\frac{\tau}{4}} \end{aligned} \tag{24}$$

with $c = \frac{1}{2\ln 2}$.

Proof: Since the transmitter can adapt its encoder to the specific channel realization, we have a family of input distributions p_i, one for each $i \in \{1, ..., S\}$. For each distribution $p_i^{\otimes n}(x^n)$, $i = 1, ..., S$, and $\delta > 0$ let $\mathcal{T}_{p_i,\delta}^n$ denote the corresponding set of typical sequences on $\mathcal{X}^n$. We restrict

the possible inputs to these subsets and define the new input distributions

$$p_i^n(x^n) := \begin{cases} \frac{p_i^{\otimes n}(x^n)}{p_i^{\otimes n}(\mathcal{T}_{p_i,\delta}^n)} & \text{if } x^n \in \mathcal{T}_{p_i,\delta}^n, \\ 0 & \text{else.} \end{cases} \tag{25}$$

Let $\tau > 0$ and set the rates $R_{\mathrm{R1}} := \min_{i=1,...,S} I(p_i, W_{1,i}) - \frac{\tau}{2}$ and $R_{\mathrm{R2}} := \min_{i=1,...,S} I(p_i, W_{2,i}) - \frac{\tau}{2}$. For each $i \in \{1, ..., S\}$ we generate $M_1^{(n)} M_2^{(n)}$ independent codewords $X_i^n(m)$, one for each $m = (m_1, m_2)$, of length n with $M_1^{(n)} := \lfloor 2^{nR_{\mathrm{R2}}} \rfloor$ and $M_2^{(n)} := \lfloor 2^{nR_{\mathrm{R1}}} \rfloor$ according to p_i^n. This implies that all generated random codewords $X_i^n(m) \in \mathcal{T}_{p_i,\delta}^n$ almost surely, $i = 1, ..., S$.

Since for each channel realization $s \in \mathcal{S}$, the used random codewords differ, the definitions of the decoding sets of nodes 1 and 2 slightly change. See (26a) and (26b) at the top of the next page, where $X^n := \{X_i^n\}_{1 \le i \le S}$ and $X_i^n := \{X_i^n(m_1, m_2)\}_{1 \le m_1 \le M_1^{(n)}, 1 \le m_2 \le M_2^{(n)}}$ so that the decoding sets depend on all generated codewords for all channel realizations. Consequently, the corresponding error events $P_{e,1,s}^{(1)}$ and $P_{e,2,s}^{(1)}$ at node 1 are now given by (27a) and (27b), at the top of the next page.

We continue with the analysis of the probability of error for node 1, again the analysis for node 2 follows accordingly using the same arguments. As in Lemma 5 we average over all codebooks and show that $\mathbb{E}_{X^n}[\mu_{1,s}^{(n)}] \le \mathbb{E}_{X^n}[P_{e,1,s}^{(1)} + P_{e,2,s}^{(1)}]$ can be bounded uniformly in s from above by a term which decreases exponentially fast for increasing block length n. The derivation for $\mathbb{E}_{X^n}[P_{e,1,s}^{(1)}]$ proceeds exactly as in Lemma 5 and leads to same upper bound, see (28) on the next page, with $c = \frac{1}{2 \ln 2}$, cf. also (11). The first steps of the derivation for $\mathbb{E}_{X^n}[P_{e,2,s}^{(1)}]$ are similar to Lemma 5 up to (29), see next page, where q_s denotes the output distribution generated by p_s and $W_{1,s}$, cf. also (13). The crucial point is that since for each $i \in \{1, ..., S\}$ we have a different input distribution p_i, the input distribution and channel may not coincide with the output distribution. But nevertheless we can apply Lemma 3 since $X_i^n(m_1, j) \in \mathcal{T}_{p_i,\delta}^n$ almost surely and further this lemma is also applicable if the distributions do not match, cf. (4a). We obtain similarly to Lemma 5, (30) at the bottom of the next page, for $\delta \in (0, \frac{1}{4|\mathcal{X}||\mathcal{Y}_1|})$. The rest of the proof proceeds exactly as in Lemma 5 so that we end up with the following upper bound for the average probability of error at node k, $k = 1, 2$,

$$\mathbb{E}_{X^n}[\mu_k^{(n)}] \le S(n+1)^{|\mathcal{X}||\mathcal{Y}_k|} 2^{-nc\delta^2} + S^2 \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}_k|}}{1 - (n+1)^{|\mathcal{X}|} 2^{-nc\delta^2}} 2^{-n\frac{\tau}{4}} \tag{31}$$

with $c = \frac{1}{2 \ln 2}$ as stated in (24) which finally proves the lemma. ■

With the previous Lemmas 8 and 9 we are now able to establish the capacity region of the compound BBC with CSIT for the cases of a finite set and an arbitrary set. The proofs proceed exactly as in Section IV so that we omit them for brevity.

Corollary 2: The capacity region $\mathcal{C}_{\mathrm{BBC}}^{\mathrm{CSIT}}$ of the finite compound BBC $\mathcal{S} = \{1, ..., S\}$ with CSIT is the set of all rate pairs $(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}_+^2$ satisfying

$$R_{\mathrm{R1}} \le \min_{i=1,..,S} I(p_i, W_{1,i}|U), \tag{32a}$$

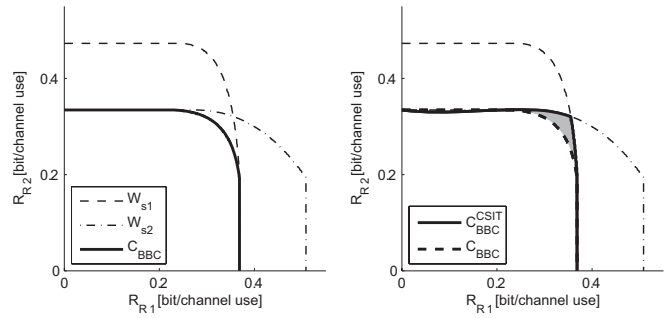
$$R_{\mathrm{R2}} \le \min_{i=1,..,S} I(p_i, W_{2,i}|U) \tag{32b}$$

for random variables $(U, X, Y_1, Y_2) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$ and joint probability distributions $\{q(u)p_s(x|u)W_s(y_1, y_2|x)\}_{s \in \mathcal{S}}$. The cardinality U can be bounded by $|\mathcal{U}| \le 2$. ■

Theorem 3: The capacity region $\mathcal{C}_{\mathrm{BBC}}^{\mathrm{CSIT}}$ of the compound BBC, where the set of channels can be arbitrary, is the set of all rate pairs $(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}_+^2$ satisfying

$$R_{\mathrm{R1}} \le \inf_{s \in \mathcal{S}} I(p_s, W_{1,s}|U), \quad R_{\mathrm{R2}} \le \inf_{s \in \mathcal{S}} I(p_s, W_{2,s}|U) \tag{33}$$

for random variables $(U, X, Y_1, Y_2) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$ and joint probability distributions $\{q(u)p_s(x|u)W_s(y_1, y_2|x)\}_{s \in \mathcal{S}}$. The cardinality U can be bounded by $|\mathcal{U}| \le 2$. ■



(a) Capacity region $\mathcal{C}_{\mathrm{BBC}}$ of the compound BBC with channel uncertainty at all nodes.

(b) Capacity region $\mathcal{C}_{\mathrm{BBC}}^{\mathrm{CSIT}}$ of the compound BBC with CSIT.

Fig. 2. Capacity regions of a particular compound BBC with $|\mathcal{X}| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and $\mathcal{S} = \{s_1, s_2\}$ with channel uncertainty at all nodes and CSIT in Fig. 2(a) and 2(b), respectively. The shaded area in Fig. 2(b) illustrates the gain in the capacity region based on the available CSIT. The rate regions for channel realizations s_1 (dashed line) and s_2 (dashed-dotted line) are included for convenience.

VII. Numerical Example and Game-Theoretic Interpretation

In this section we give a numerical example which illustrates how CSIT improves the capacity region of the compound BBC. Therefore, let $|\mathcal{X}| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and consider the particular set of channels $\mathcal{S} = \{s_1, s_2\}$ with two possible states. The marginal channels are given by the following transition probability matrices

$$W_{1,s_1} := \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0 & 0.7 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} \quad W_{2,s_1} := \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

$$W_{1,s_2} := \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0 & 1 & 0 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} \quad W_{2,s_2} := \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$$

Fig. 2 depicts the capacity regions $\mathcal{C}_{\mathrm{BBC}}$ and $\mathcal{C}_{\mathrm{BBC}}^{\mathrm{CSIT}}$ of this particular compound BBC for channel uncertainty at all nodes and CSIT, respectively. How CSIT affects the maximal achievable rates is shown in Fig. 2(b), where the shaded area illustrates the gain in the capacity region due to the available CSIT.

$$D^{(1)}_{m_2|m_1}(X^n) := \Big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n_i(m)\big)\Big) \cap \Big(\bigcup_{\substack{j=1\\ j\neq m_2}}^{M_2^{(n)}} \bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n_i(m_1,j)\big)\Big)^c \tag{26a}$$

$$D^{(2)}_{m_1|m_2}(X^n) := \Big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{2,i},\delta}\big(X^n_i(m)\big)\Big) \cap \Big(\bigcup_{\substack{j=1\\ j\neq m_1}}^{M_1^{(n)}} \bigcup_{i=1}^{S} \mathcal{T}^n_{W_{2,i},\delta}\big(X^n_i(j,m_2)\big)\Big)^c \tag{26b}$$

$$P^{(1)}_{e,1,s} := \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} W^{\otimes n}_{1,s}\Big(\big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n_i(m)\big)\big)^c \big| X^n_s(m)\Big) \tag{27a}$$

$$P^{(1)}_{e,2,s} := \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} W^{\otimes n}_{1,s}\Big(\bigcup_{\substack{j=1\\ j\neq m_2}}^{M_2^{(n)}} \bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n_i(m_1,j)\big) \big| X^n_s(m_1,m_2)\Big) \tag{27b}$$

$$\mathbb{E}_{X^n}[P^{(1)}_{e,1,s}] = \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \mathbb{E}_{X^n}\Big[W^{\otimes n}_{1,s}\Big(\big(\bigcup_{i=1}^{S} \mathcal{T}^n_{W_{1,i},\delta}\big(X^n_i(m)\big)\big)^c \big| X^n_s(m)\Big)\Big] \leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|} 2^{-nc\delta^2} \tag{28}$$

$$\mathbb{E}_{X^n}[P^{(1)}_{e,2,s}] \leq \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \sum_{\substack{j=1\\ j\neq m_2}}^{|\mathcal{M}_2|} \sum_{i=1}^{S} \frac{1}{1-(n+1)^{|\mathcal{X}|}2^{-nc\delta^2}} \mathbb{E}_{X^n_i(m_1,j)}\Big[q^{\otimes n}_s\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n_i(m_1,j)\big)\Big)\Big] \tag{29}$$

Similar to the single-user compound channel it is possible to analyze the compound BBC from a game-theoretic perspective. Therefore, we assume that the nodes and *nature* play a two-player zero-sum game [26], [27] with the mutual information I as the payoff function as depicted in Fig. 3. This is called a *game against nature* [28]. In this game, the set of channels $\mathcal{S}$ corresponds to nature's action space, and nature's aim is to establish the worst communication condition by selecting $s \in \mathcal{S}$ such that the mutual information is minimized. The set of input distributions $\mathcal{P}(\mathcal{X})$ corresponds to the action space of the player. Clearly, the nodes want to maximize the mutual information. For given $p \in \mathcal{P}(\mathcal{X})$ and $s \in \mathcal{S}$ the outcome of the game is given by the following achievable rate region

$$\mathcal{R}(p,s) = \{(R_{\mathrm{R1}}, R_{\mathrm{R2}}) \in \mathbb{R}^2_+ : R_{\mathrm{R1}} \leq I(p, W_{1,s}), \; R_{\mathrm{R2}} \leq I(p, W_{2,s})\}. \tag{34}$$

Within this game against nature framework the game can be played in two different ways. First, the player and nature moves simultaneously without knowing the other's choice. And second, nature moves first so that the player is aware of nature's choice. In the single-user scenario, these two types of the game lead to well-known maxmin and minmax formulations for the outcome of the game. Since our scenario deals with rate regions, cf. (34), we have a vector-valued problem and the max and min expressions extend to the union and intersection. Then, the outcomes of the game are

$$\mathcal{C}_{\mathrm{BBC}} = \mathrm{ConvexHull}\Big(\bigcup_{p\in\mathcal{P}(\mathcal{X})} \bigcap_{s\in\mathcal{S}} \mathcal{R}(p,s)\Big) \tag{35}$$

and

$$\mathcal{C}^{\mathrm{CSIT}}_{\mathrm{BBC}} = \mathrm{ConvexHull}\Big(\bigcap_{s\in\mathcal{S}} \bigcup_{p\in\mathcal{P}(\mathcal{X})} \mathcal{R}(p,s)\Big). \tag{36}$$

We see that (35) and (36) are equivalent to the capacity regions given in Theorem 1 and 3 and correspond therefore to the cases of channel uncertainty at all nodes and CSIT, respectively. Note that in the theorems the convex hull is established by the time-sharing variable U. Moreover, it follows immediately from (35) and (36) that $\mathcal{C}_{\mathrm{BBC}} \subseteq \mathcal{C}^{\mathrm{CSIT}}_{\mathrm{BBC}}$ which agrees with the intuition that CSIT improves the capacity region.

VIII. Conclusion

In practical wireless communication systems channel uncertainty is a ubiquitous phenomenon. The question must be asked if it is advantageous to improve the available channel state information at the nodes or if the nodes should be left

$$\mathbb{E}_{X^n_i(m_1,j)}\Big[q^{\otimes n}_s\Big(\mathcal{T}^n_{W_{1,i},\delta}\big(X^n_i(m_1,j)\big)\Big)\Big] \leq (n+1)^{|\mathcal{X}||\mathcal{Y}_1|} 2^{-n(I(p_i,W_{1,i})-\varphi(\delta)-\psi(\delta))} \tag{30}$$

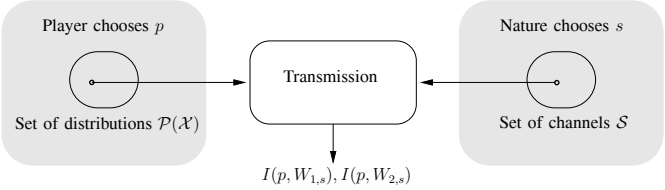


Fig. 3. Transmission as a game against nature.

with the uncertainty. The concept of the compound channel allows us to derive robust coding strategies that are appropriate for wireless applications where certain rates have to be guaranteed even in the case of channel uncertainty. Further, the analysis shows the best possible rates that are achievable under channel uncertainty. This allows us to assess if it is worthwhile to improve the channel state information at the nodes, e.g., by using longer training sequences or feedback. In particular, this is important to know for the design of wireless networks.

In this work we address the bidirectional broadcast channel and present robust coding strategies which guarantee certain rates regardless of the current channel realization. These immediately lead to a characterization of the capacity region of the compound BBC which is an useful result since it constitutes the basis for further analysis of multi-user settings under more complex models of channel uncertainty and in uncoordinated wireless networks. In particular, our results provides valuable insights since for the general broadcast channel with discrete channels and finite alphabets the capacity region for compound channels is not known and, consequently, similar results are not available. To date, only some special cases are treated as for example the case where the transmitter and the receivers have multiple antennas and the channels are degraded [7] which is a quite different setting to the one we considered here.

Furthermore, the analysis shows that CSIR does not improve the capacity region if the transmitter merely knows the set of channels, which is at first counter-intuitive. But at second glance this becomes clear if one realizes that transmitted symbols "wasted" for channel estimation are negligible for large block lengths. Further, we show that CSIT can advantageously be used to improve the capacity region since the transmitter can adapt its encoder to the specific channel realization. The game-theoretic interpretation of the compound BBC reveals interesting generalizations, which keep the characteristics of the single-user compound channel but includes now multi-user effects. This is a nice property of the compound BBC, which is not self-evident for multi-user scenarios.

APPENDIX

A. Proof of Lemma 2

We define $V_{y^n|x^n}(b|a) = \frac{p_{x^n,y^n}(a,b)}{p_{x^n}(a)}$ for empirical distributions $p_{x^n}(a) = \frac{1}{n}N(a|x^n)$ and $p_{x^n,y^n}(a,b) = \frac{1}{n}N(a,b|x^n,y^n)$ for all $a \in \mathcal{A}, b \in \mathcal{B}$. Then from [21, Lemma 2.6] we have $W^{\otimes n}(\mathcal{T}_W^n(x^n)|x^n) \leq 2^{-nD(V_{y^n|x^n}\|W|p_{x^n})}$ with

$$D(V_{y^n|x^n}\|W|p_{x^n}) = \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p_{x^n}(a)V_{y^n|x^n}(b|a)\log\frac{V_{y^n|x^n}(b|a)}{W(b|a)}$$

$$= \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p_{x^n,y^n}(a,b)\log\frac{p_{x^n,y^n}(a,b)}{W(b|a)p_{x^n}(a)} = D(p_{x^n,y^n}\|p_{x^n}W). \tag{37}$$

Observe that $\mathcal{T}_W^n(x^n)$ has at most $(n+1)^{|\mathcal{A}||\mathcal{B}|}$ V-shells, cf. also [21, Sec. 1.2], so that

$$W^{\otimes n}({\mathcal{T}_{W,\delta}^n}^c(x^n)|x^n) \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-nD_*} \tag{38}$$

with $D_* = \min\{D(p_{x^n,y^n}\|p_{x^n}W) : \exists a \in \mathcal{A}, b \in \mathcal{B} : |p_{x^n,y^n}(a,b) - p_{x^n}(a)W(b|a)| > \delta\}$. From [21] we have[2] $D(p_{x^n,y^n}\|p_{x^n}W) \geq \frac{1}{2\ln 2}(\sum_{a,b}|p_{x^n,y^n}(a,b) - p_{x^n}(a)W(b|a)|)^2$ so that $D_* \geq \frac{1}{2\ln 2}|p_{x^n,y^n}(a,b) - p_{x^n}(a)W(b|a)|^2 = \frac{\delta^2}{2\ln 2}$ for all $a \in \mathcal{A}$, $b \in \mathcal{B}$. From this we have

$$W^{\otimes n}({\mathcal{T}_{W,\delta}^n}^c(x^n)|x^n) \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-nc\delta^2} \tag{39}$$

with $c = \frac{1}{2\ln 2}$ which proves the lemma. ■

B. Proof of Lemma 3

Let $y^n \in \mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)$ where $\tilde{x}^n \in \mathcal{T}_{\tilde{p},\delta}^n$. Then we know from [21, Lemma 2.6] that $q^{\otimes n}(y^n) = 2^{-n(D(p_{y^n}\|q)+H(p_{y^n}))}$ with $p_{y^n}(b) = \frac{N(b|y^n)}{n}$. Since $D(p_{y^n}\|q) \geq 0$ we have

$$q^{\otimes n}(y^n) \leq 2^{-nH(p_{y^n})}. \tag{40}$$

Since $y^n \in \mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)$ and $\tilde{x}^n \in \mathcal{T}_{\tilde{p},\delta}^n$ it follows from [21, Lemma 2.10] that $y^n \in \mathcal{T}_{\tilde{q},2|\mathcal{A}|\delta}^n$ and therewith $\sum_{b\in\mathcal{B}}|p_{y^n}(b) - \tilde{q}(b)| \leq 2|\mathcal{A}||\mathcal{B}|\delta < 1/2$ so that [21, Lemma 2.7] implies

$$|H(p_{y^n}) - H(\tilde{q})| \leq -2|\mathcal{A}||\mathcal{B}|\delta\log\frac{2|\mathcal{A}||\mathcal{B}|\delta}{|\mathcal{B}|} =: \varphi(\delta) \tag{41}$$

with $\lim_{\delta\searrow 0}\varphi(\delta) = 0$. If we combine (40) and (41), we obtain $q^{\otimes n}(y^n) \leq 2^{-n(H(\tilde{q})-\varphi(\delta))}$ and therewith

$$q^{\otimes n}(\mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)) \leq |\mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)|2^{-n(H(\tilde{q})-\varphi(\delta))}. \tag{42}$$

Since $\tilde{x}^n \in \mathcal{T}_{\tilde{p},\delta}^n$, it follows from [21, Lemma 2.13] that $|\mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)| \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{n(H(\tilde{W}|\tilde{p})+\psi(\delta))}$ holds with universal $\psi(\delta) > 0$ and $\lim_{\delta\searrow 0}\psi(\delta) = 0$. Inserting this in (42) we obtain

$$q^{\otimes n}(\mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)) \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-n(I(\tilde{p},\tilde{W})-\varphi(\delta)-\psi(\delta))}. \tag{43}$$

The relation $q^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n)) \leq (n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-n(I(p,W)-\varphi(\delta)-\psi(\delta))}$ follows immediately for $W = \tilde{W}$ and $p = \tilde{p}$ which proves the lemma. ■

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] T. Unger and A. Klein, "Duplex schemes in multiple antenna two-hop relaying," *EURASIP J. Adv. Signal Process.*, vol. 2008, 2008.

[3] C. Schnurr, T. J. Oechtering, and S. Stanczak, "Achievable rates for the restricted half-duplex two-way relay channel," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Nov. 2007, pp. 1468–1472.

[4] D. Gündüz, E. Tuncel, and J. Nayak, "Rate regions for the separated two-way relay channel," in *Proc. Allerton Conf. Commun., Control, Computing*, Sep. 2008, pp. 1333–1340.

[2]This bound with a worse constant was first given by Pinsker [29] and is therefore also known as Pinsker's inequality.

[5] 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.

[6] H. Weingarten, Y. Steinberg, and S. S. (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[7] H. Weingarten, T. Liu, S. S. (Shitz), Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple-input multiple-output compound broadcast channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5011–5023, Nov. 2009.

[8] C.-H. Liu and F. Xue, "Network coding for two-way relaying: rate region, sum rate and opportunistic scheduling," in *Proc. IEEE Int. Conf. Commun.*, May 2008, pp. 1044–1049.

[9] T. J. Oechtering, R. F. Wyrembelski, and H. Boche, "Optimal time-division of two-phase decode-and-forward bidirectional relaying," in *Proc. Int. Symp. Inf. Theory Applications*, Dec. 2008, pp. 829–834.

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

[11] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: a game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1707–1727, July 2003.

[12] T. Weber, A. Sklavos, and M. Meurer, "Imperfect channel-state information in MIMO transmission," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 543–552, Mar. 2006.

[13] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems with channel uncertainty," in *Proc. IEEE Int. Conf. Commun.*, May 2008, pp. 3516–3520.

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

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

[16] ——, *Coding Theorems of Information Theory*, 3rd edition. Springer-Verlag, 1978.

[17] R. F. Wyrembelski, I. Bjelaković, and H. Boche, "Coding strategies for bidirectional relaying for arbitrarily varying channels," in *Proc. IEEE Global Commun. Conf.*, Dec. 2009, pp. 1–6.

[18] ——, "Bidirectional relaying in wireless networks—impact of degree of coordination," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Mar. 2010, pp. 3234–3237.

[19] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Annals Probability*, vol. 2, no. 5, pp. 805–814, Oct. 1974.

[20] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2773–2795, Nov. 1998.

[21] I. Csiszár and J. Körner, *Information Theory - Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[22] R. F. Wyrembelski, T. J. Oechtering, I. Bjelaković, C. Schnurr, and H. Boche, "Capacity of Gaussian MIMO bidirectional broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2008, pp. 584–588.

[23] T. J. Oechtering, R. F. Wyrembelski, and H. Boche, "Multiantenna bidirectional broadcast channels—optimal transmit strategies," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1948–1958, May 2009.

[24] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*. American Mathematical Society, 1996.

[25] J.-B. Hiriart-Urruty and C. Lemarèchal, *Fundamentals of Convex Analysis*. Springer-Verlag, 2001.

[26] R. J. Aumann and S. Hart, *Handbook of Game Theory with Economic Applications, Volume 2*. Elsevier Science & Technology, 1994.

[27] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd edition. Ser. Classics in Applied Mathematics. SIAM, 1998.

[28] J. Milnor, "Games against nature," RAND Corporation, pp. 49–59, 1951.

[29] M. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.

Rafael F. Wyrembelski (S'08) received the Dipl.-Ing. degree in electrical engineering and computer science from the Technische Universität Berlin, Germany, in 2007, where he is currently working towards the Ph.D. degree. Since 2007, he has worked as a research and teaching assistant at the Heinrich-Hertz-Lehrstuhl für Informationstheorie und theoretische Informationstechnik at the Technische Universität Berlin, Germany.

Igor Bjelaković received the Dipl.Phys. degree in physics and Dr.rer.nat. degree in mathematics from the Technische Universität Berlin, Berlin, Germany, in 2001 and 2004, respectively. He is currently a postdoctoral researcher at the Heinrich-Hertz-Lehrstuhl für Informationstheorie und theoretische Informationstechnik and the Institut für Mathematik at the Technische Universität Berlin.

Tobias J. Oechtering (S'01-M'08) received his Diplom-Ingenieur degree in electrical engineering and information technology in 2002 from RWTH Aachen University, Germany, and his Doktor-Ingenieur degree in electrical engineering in 2007 from the Technische Universität Berlin, Germany. Between 2002 and 2007 he worked as a research and teaching assistant at the Heinrich-Hertz-Chair for Mobile Communications of the Technische Universität Berlin, Germany. He has previously been with the Fraunhofer German-Sino Lab for Mobile Communications, Berlin, Germany, from 2007 to 2008. In November 2008, he joined the Communication Theory Lab at the Royal Institute of Technology (KTH) as a post-doctoral researcher and has been an assistant professor since July 2010. Dr. Oechtering received the "Advancement Award" from the Vodafone Foundation in June 2009.

Holger Boche (M'04-SM'07) received his Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the Technische Universität Dresden, Germany, in 1990 and 1994, respectively. In 1992, he graduated in mathematics from the Technische Universität Dresden, and in 1998 he received his Dr.rer.nat. degree in pure mathematics from the Technische Universität Berlin. From 1994 to 1997, he did postgraduate studies in mathematics at the Friedrich-Schiller Universität Jena, Germany. In 1997, he joined the Heinrich-Hertz-Institut (HHI) für Nachrichtentechnik Berlin. Since 2002, he has been the Full Professor for Mobile Communication Networks at the Technische Universität Berlin at the Institute for Communications Systems. In 2003, he became Director of the Fraunhofer German-Sino Lab for Mobile Communications, Berlin, Germany, and since 2004 he has also been the Director of the Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany. He was a visiting professor at the ETH Zurich during the winter terms of 2004 and 2006, and at KTH Stockholm during the summer term, 2005. Prof. Boche 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 co-recipient of the 2006 IEEE Signal Processing Society Best Paper Award and recipient of the 2007 IEEE Signal Processing Society Best Paper Award. He is a member of the 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.