BIDIRECTIONAL RELAYING USING INTERFERENCE CANCELLATION

Tobias J. Oechtering and Holger Boche

Technical University of Berlin, Heinrich-Hertz-Chair for Mobile Communications,
Einsteinufer 25, 10587 Berlin, Germany

ABSTRACT

Bidirectional communication between two nodes is made possible by a half-duplex relay node in a multiple access and broadcast phase. We optimize the time-division between the two phases and give a complete characterization of the convex achievable rate region. This allows us to derive a relay selection criteria and to present a throughput optimal rate allocation policy for a cross-layer design. Furthermore, if the relay additionally multicasts a common message, we can improve the resource utilization by joint processing of two routing tasks.

1. INTRODUCTION

In wireless communication scenarios where the direct link does not have the desired quality, e.g. due to shadowing or distance, cooperative protocols which realize range extension seem to be a possible solution. Since it is practically difficult to isolate at one node a simultaneously received and transmitted signal using the same frequency sufficiently, we assume half-duplex relay nodes. Because of this, most cooperative communication protocols allocate additional resources in time or frequency, [1], [2] among others, and therefore suffer from an inherent loss in spectral efficiency.

The spectral loss can be reduced in a two phase bidirectional relay communication where one node acts as a relay and enables bidirectional communication between two other nodes. In the first phase, the two nodes transmit their messages to the relay node, which decodes the messages. In the second phase the relay broadcasts back a re-encoded composition of both messages. The knowledge of the first phase allows the receiving nodes to perform interference cancellation before decoding so that it is like an interference-free transmission effectively [3], [4]. For that reason, bidirectional relay communication circumvents the inherent spectral loss of unidirectional cooperative protocols. Conceptually, there is a close relationship to bidirectional relaying protocols based on the network coding principle where the relay node performs an XOR operation on the decoded bit streams [5] but it assumes equal channel qualities and leaves channel coding out of considerations.

In [3] and [4] bidirectional relaying with equal time division is considered. In this work, we optimize the time division and therefore achieve a larger bidirectional achievable rate region, which is studied in section 2. This allows us to derive a relay selection criteria which decides for the relay node which maximizes the weighted rate sum for any rate pair on the boundary in section 3. In a scenario with N relay nodes and iid Rayleigh fading relay selection realizes the same multi-user space diversity order $O(\log(\log(N)))$ as distributed beamforming [6]. In section 4 we add relay communication on the time-division optimized bidirectional relaying protocol. Therefore, the relay node piggybacks on the bidirectional broadcast messages an own multicast message. It shows that it is optimal for node 1 and 2 to decode the relay message first as it is for equal time-division [4], [7]. Thereby, the joint processing provides a rate trade-off which allows an optimal resource utilization, i.e we improve the efficiency by converging two routing schemes. Finally, the complete characterization of the time-division optimal bidirectional achievable rate region in section 2 allows us to present a throughput optimal rate allocation policy in section 5, which we discussed for the equal time-division case in [8]. Both are adapted from the maximum differential backlog algorithm presented in the landmark paper [9]. Further, the stability regions of policies can be proved using the well developed theory of drift analysis using a quadratic Lyapunov function on the buffer levels [10]. Since our stability analysis is adapted from the cross-layer design for a satellite broadcast scenario [11], we keep our stability discussion short. [1]

2. ACHIEVABLE RATE REGION

We consider a three node network where two phase bidirectional communication is realized between node 1 and 2 by a decode-and-forward half-duplex relay node. Thereby, let h_1 and h_2 denote the flat fading channel gain of the reciprocal channel between the relay node and node 1 and 2. We have an individual transmit power constraint P_k for each node $k \in \{1, 2, R\}$. Furthermore, we assume independent additive

[1] *Notation:* Bold and calligraphic letters denote vectors and sets respectively; $\mathrm{int}\mathcal{A}$ specifies the interior of set $\mathcal{A}$; $\mathbb{R}_+$ denotes the set of non-negative real numbers; $[a, b]$ and (a, b) specify the closed and open interval from a to b in $\mathbb{R}$; $\mathrm{lhs} := \mathrm{rhs}$ assigns to lhs the value of rhs; $[\cdot]_+ = \max(0, \cdot)$; $\mathbb{1}$ designate a vector with ones; $\mathbb{E}\{\cdot\}$ denotes the expectation value; any log is to the base two; we regard the discrete stochastic process X as an indexed collection of random variables $X(n)$, which takes realization x of state space $\mathcal{X}$.

white complex Gaussian noise $N_k \sim \mathcal{CN}(0, \sigma^2)$ at each receiver $k \in \{1, 2, R\}$. Therewith, we define the signal-to-noise ratios $\gamma_k = P_k/\sigma^2$ for each receiver $k \in \{1, 2, R\}$. Finally, we assume that all nodes are perfectly synchronized and each node knows the channel states, codebooks, and power allocations necessary for decoding.

Since the communication is performed in two phases, let $\alpha \in [0, 1]$ denote the fraction of time in the multiple access (MAC) phase and $1 - \alpha$ the fraction of time in the broadcast (BC) phase. In the following subsections, we first present the achievable rate regions for the MAC and BC phase separately. Then we look at the achievable rate region of bidirectional relaying using interference cancellation with the optimal time-division between MAC and BC phase.

2.1. Achievable Rate Region of MAC phase

In the first phase, node 1 transmits the message m_1 for node 2 with rate R_1 and node 2 transmits the message m_2 for node 1 with rate R_2 to the relay node. The encoding and decoding is exactly like the classical discrete memoryless Gaussian MAC channel. If the rate tuple $\boldsymbol{R} = [R_1, R_2] \in \mathbb{R}_+^2$ of node 1 and 2 is within the achievable rate region

$$\mathcal{R}_{\mathrm{MAC}} = \big\{\boldsymbol{R} : \; R_1 \le R_{\overrightarrow{1\mathrm{R}}},\, R_2 \le R_{\overrightarrow{2\mathrm{R}}},\, R_1 + R_2 \le R_\Sigma\big\}$$

with $R_{\overrightarrow{1\mathrm{R}}} = \log[1 + \gamma_1 |h_1|^2]$, $R_{\overrightarrow{2\mathrm{R}}} = \log[1 + \gamma_2 |h_2|^2]$, and $R_\Sigma = \log[1 + \gamma_1 |h_1|^2 + \gamma_2 |h_2|^2]$, it is assumed that the relay node decodes the messages m_1 and m_2 perfectly.

Since $\mathcal{R}_{\mathrm{MAC}}$ is a pentagon, it can be completely described by five vertices. Thereby, the vertices where the individual rate constraints intersect with the sum-rate constraint, $\nu_{1\Sigma} = [R_{\overrightarrow{1\mathrm{R}}}, R_\Sigma - R_{\overrightarrow{1\mathrm{R}}}]$, and $\nu_{\Sigma 2} = [R_\Sigma - R_{\overrightarrow{2\mathrm{R}}}, R_{\overrightarrow{2\mathrm{R}}}]$, are most interesting for the combinatoric.

2.2. Achievable Rate Region of BC phase

In the succeeding phase, the relay forwards the previously received messages m_1 to node 2 and message m_2 to node 1. Since we have Gaussian channels, the relay uses independent Gaussian codebooks to encode the messages and transmits the superposition of both codewords. Thereby, let β_1 and β_2 denote the proportion of relay transmit power spend for forwarding message m_1 and m_2 respectively. Obviously, we require $[\beta_1, \beta_2] \in \mathcal{B} = \{[\beta_1, \beta_2] \in [0, 1] \times [0, 1] : \; \beta_1 + \beta_2 \le 1\}$ to satisfy the relay transmit power constraint. Since the message m_1 and m_2 originate from node 1 and 2, each node knows one message. This a priori knowledge improves the decoding capability of the unknown message at each node. For a Gaussian channel this principle is known as interference cancellation. Therefore, each node subtracts the interference caused by the codeword of its own message and achieves an interference-free reception. For that reason, we can achieve an error-free transmission if the rate tuple $\boldsymbol{R} = [R_1, R_2] \in \mathbb{R}_+^2$ is within the achievable rate region

$$\mathcal{R}_{\mathrm{BC}} = \big\{\boldsymbol{R} : \; R_1 \le R_{\overrightarrow{\mathrm{R}2}}(\beta_1), \; R_2 \le R_{\overrightarrow{\mathrm{R}1}}(\beta_2), [\beta_1, \beta_2] \in \mathcal{B}\big\}$$

with $R_{\overrightarrow{\mathrm{R}2}}(\beta_1) = \log[1 + \gamma_\mathrm{R}\beta_1 |h_2|^2]$ and $R_{\overrightarrow{\mathrm{R}1}}(\beta_2) = \log[1 + \gamma_\mathrm{R}\beta_2 |h_1|^2]$. We emphasize that we achieve a larger rate region than the degraded broadcast channel.

2.3. Achievable Rate Region of Bidirectional Relaying

Since we use the MAC and BC phase for fraction of time only, we have to scale the achievable rate pairs according the time-division. This means that for a time-division parameter $\alpha \in [0, 1]$, we can achieve rate pairs $\boldsymbol{R}$ within the rate regions $\alpha\mathcal{R}_{\mathrm{MAC}}$ in the MAC phase and $(1 - \alpha)\mathcal{R}_{\mathrm{BC}}$ in the BC phase. Since we assume that no messages will be stored at the relay node, each message received in the MAC phase has to be forwarded in the BC phase immediately. Therefore, for a successful bidirectional relay transmission of the messages m_1 with rate R_1 and message m_2 with rate R_2 the rate pair $\boldsymbol{R} = [R_1, R_2]$ has to be within $\alpha\mathcal{R}_{\mathrm{MAC}}$ and $(1 - \alpha)\mathcal{R}_{\mathrm{BC}}$ simultaneously. This means that for given time-division parameter $\alpha \in [0, 1]$ the achievable rate region of the bidirectional relaying is given by the intersection

$$\mathcal{R}_{\mathrm{BIR}}(\alpha) = \alpha\mathcal{R}_{\mathrm{MAC}} \cap (1 - \alpha)\mathcal{R}_{\mathrm{BC}}.$$

Since this applies for any time-division parameter α the achievable rate region of the bidirectional relaying is given by the union over all possible time-division parameters,

$$\mathcal{R}_{\mathrm{BIR}} = \bigcup_{\alpha \in [0,1]} \mathcal{R}_{\mathrm{BIR}}(\alpha). \tag{1}$$

In the following, we will characterize the boundary of the bidirectional achievable rate region $\mathcal{R}_{\mathrm{BIR}}$. First, we characterize the optimal time-division and relay power distribution for a fixed operating rate pair in the MAC phase. Then, this will be used to derive an equivalent description of $\mathcal{R}_{\mathrm{BIR}}$.

Lemma 1 *For a fixed rate pair $[R_1^M, R_2^M] \in \mathcal{R}_{\mathrm{MAC}}$ the feasible set of time-division parameters where an operating at the rate pair $[R_1^M, R_2^M]$ in the MAC phase is possible is given by $\mathcal{A} = \big\{\alpha \in [0, 1] : \; \alpha R_1^M \le (1 - \alpha)R_{\overrightarrow{2\mathrm{R}}}(\beta_1)$, $\alpha R_2^M \le (1 - \alpha)R_{\overrightarrow{1\mathrm{R}}}(\beta_2)$, with $[\beta_1, \beta_2] \in \mathcal{B}\big\}$. For a time-division parameter $\alpha \in \mathcal{A}$ we achieve the bidirectional rate pair $[R_1, R_2] = \alpha[R_1^M, R_2^M]$. Then the optimal time-division parameter $\alpha^\star = \max_{\alpha \in \mathcal{A}} \alpha$ is uniquely characterized by the equations*

$$\alpha^\star R_1^M = (1 - \alpha^\star)R_{\overrightarrow{\mathrm{R}2}}(\beta_1^\star), \tag{2a}$$

$$\alpha^\star R_2^M = (1 - \alpha^\star)R_{\overrightarrow{\mathrm{R}1}}(\beta_2^\star), \tag{2b}$$

which also characterize the optimal relay power distribution $[\beta_1^\star, \beta_2^\star] \in \mathcal{B}$ with $\beta_1^\star + \beta_2^\star = 1$.

Sketch of Proof: It is clear from the construction of $\mathcal{A}$ that for any $\alpha \in \mathcal{A}$ we can achieve $[R_1, R_2] = \alpha[R_1^M, R_2^M]$. Therefore, the largest α maximizes component-wise the bidirectional rate pair. Then it can be shown by contradiction that the optimal coefficients $\alpha^\star$ and $\beta_1^\star$ and $\beta_2^\star$ are defined by the equations (2a) and (2b). □

Since the largest bidirectional rates are achieved by rate pairs on the boundary of the MAC region, we can find an equivalent characterization of $\mathcal{R}_{\mathrm{BIR}}$ by transforming the sum and individual rate constraints of the MAC region.

Theorem 1 *The bidirectional achievable rate region $\mathcal{R}_{\mathrm{BIR}}$ is given by*

$$\mathcal{R}_{\mathrm{BIR}} = \mathcal{R}_1 \cap \mathcal{R}_2 \cap \mathcal{R}_\Sigma \tag{3}$$

with rate regions

$$\begin{aligned}
\mathcal{R}_1 = \big\{[R_1,R_2] \in \mathbb{R}_+^2 :\ & \textit{there exists } \beta \in [0,1] \textit{ with} \\
& R_1 \le R_{11}(\beta) := (1-\alpha_1^\star(\beta))R_{\overrightarrow{\mathrm{R2}}}(\beta), \qquad (4a)\\
& R_2 \le R_{12}(\beta) := (1-\alpha_1^\star(\beta))R_{\overrightarrow{\mathrm{R1}}}(1-\beta)\big\},\\
\mathcal{R}_2 = \big\{[R_1,R_2] \in \mathbb{R}_+^2 :\ & \textit{there exists } \beta \in [0,1] \textit{ with} \\
& R_1 \le R_{21}(\beta) := (1-\alpha_2^\star(\beta))R_{\overrightarrow{\mathrm{R2}}}(\beta), \qquad (4b)\\
& R_2 \le R_{22}(\beta) := (1-\alpha_2^\star(\beta))R_{\overrightarrow{\mathrm{R1}}}(1-\beta)\big\},\\
\mathcal{R}_\Sigma = \big\{[R_1,R_2] \in \mathbb{R}_+^2 :\ & \textit{there exists } \beta \in [0,1] \textit{ with} \\
& R_1 \le R_{\Sigma 1}(\beta) := (1-\alpha_\Sigma^\star(\beta))R_{\overrightarrow{\mathrm{R2}}}(\beta), \qquad (4c)\\
& R_2 \le R_{\Sigma 2}(\beta) := (1-\alpha_\Sigma^\star(\beta))R_{\overrightarrow{\mathrm{R1}}}(1-\beta)\big\},
\end{aligned}$$

with optimal time-division $\alpha_1^\star(\beta) = \frac{R_{\overrightarrow{\mathrm{R2}}}(\beta)}{R_{\overrightarrow{\mathrm{1R}}}+R_{\overrightarrow{\mathrm{R2}}}(\beta)}$, $\alpha_2^\star(\beta) = \frac{R_{\overrightarrow{\mathrm{R1}}}(1-\beta)}{R_{\overrightarrow{\mathrm{2R}}}+R_{\overrightarrow{\mathrm{R1}}}(1-\beta)}$, *and* $\alpha_\Sigma^\star(\beta) = \frac{R_{\overrightarrow{\mathrm{R1}}}(1-\beta)+R_{\overrightarrow{\mathrm{R2}}}(\beta)}{R_\Sigma+R_{\overrightarrow{\mathrm{R1}}}(1-\beta)+R_{\overrightarrow{\mathrm{R2}}}(\beta)}$.

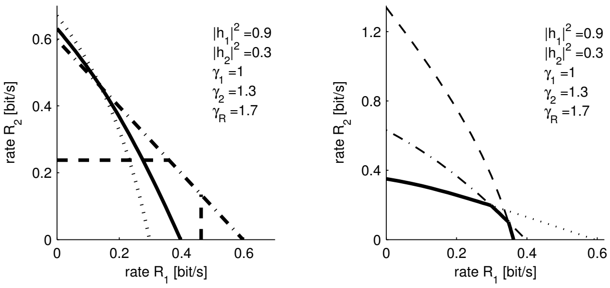
Sketch of Proof: $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_\Sigma$ follow from the transformation of the rate pairs $[R_1^M, R_2^M] \in \mathbb{R}_+^2$ satisfying the MAC individual rate constraints $R_1^M \le R_{\overrightarrow{\mathrm{1R}}}$, $R_2^M \le R_{\overrightarrow{\mathrm{2R}}}$ and sum-rate constraint $R_1^M + R_2^M \le R_\Sigma$ using Lemma 1. Thereby, we have $\beta_1^\star = \beta$ and $\beta_2^\star = 1-\beta$. □

In accordance to the theorem, the functions $\boldsymbol{R}_1(\beta) := [R_{11}(\beta), R_{12}(\beta)]$, $\boldsymbol{R}_2(\beta) := [R_{21}(\beta), R_{22}(\beta)]$, and $\boldsymbol{R}_\Sigma(\beta) := [R_{\Sigma 1}(\beta), R_{\Sigma 2}(\beta)]$ are parametrizations of the boundaries with $\beta \in [0,1]$. From the following properties of the boundaries we conclude next that $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_\Sigma$ are convex, and $\mathcal{R}_{\mathrm{BIR}}$ accordingly.

Corollary 1 *The angle of the normal vector for any boundary rate pair of $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_\Sigma$ is given by the strictly monotone decreasing functions* $\varphi_1(\beta) = \arctan\big(-\frac{\mathrm{d}R_{11}(\beta)}{\mathrm{d}\beta}/\frac{\mathrm{d}R_{12}(\beta)}{\mathrm{d}\beta}\big)$, $\varphi_2(\beta) = \arctan\big(-\frac{\mathrm{d}R_{21}(\beta)}{\mathrm{d}\beta}/\frac{\mathrm{d}R_{22}(\beta)}{\mathrm{d}\beta}\big)$, *and* $\varphi_\Sigma(\beta) = \arctan\big(-\frac{\mathrm{d}R_{\Sigma 1}(\beta)}{\mathrm{d}\beta}/\frac{\mathrm{d}R_{\Sigma 2}(\beta)}{\mathrm{d}\beta}\big)$ *with* $\beta \in [0,1]$ *respectively. It follows that $\mathcal{R}_{\mathrm{BIR}}$ is convex.*

Sketch of Proof: The monotony can be proved by inspection of the derivation of φ_1, φ_2, and φ_Σ. Then the convexity follows with the monotone behavior of R_{11}, R_{12}, R_{21}, R_{22}, $R_{\Sigma 1}(\beta)$, and $R_{\Sigma 2}$. □

Since φ_1, φ_2, and φ_Σ are continuous and strictly monotone for each function, there exists an inverse function $\varphi_1^{-1} : [\varphi_1(1), \varphi_1(0)] \to [0,1]$, $\varphi_2^{-1} : [\varphi_2(1), \varphi_2(0)] \to [0,1]$, and $\varphi_\Sigma^{-1} : [\varphi_\Sigma(1), \varphi_\Sigma(0)] \to [0,1]$ respectively. Unfortunately, φ_1^{-1}, φ_2^{-2}, and φ_Σ^{-1} have no explicit representation.



(a) Achievable rate regions $\mathcal{R}_\Sigma$ (solid line), $\mathcal{R}_{\mathrm{BC}}/2$ (dotted line), $\mathcal{R}_{\mathrm{MAC}}/2$ (dashed line), and sum-rate constraint $R_\Sigma/2$ (dashed-dotted line).

(b) Achievable rate region $\mathcal{R}_{\mathrm{BIR}}$ (solid line) after intersection of $\mathcal{R}_1$ (dashed line), $\mathcal{R}_\Sigma$ (dashed-dotted line), and $\mathcal{R}_2$ (dotted line).

Fig. 1. Optimized time-division achievable rate regions.

In Figure 1 we illustrate the achievable rate regions $\mathcal{R}_\Sigma$ and $\mathcal{R}_{\mathrm{BIR}}$ for an example scenario. For a complete characterization of the boundary of $\mathcal{R}_{\mathrm{BIR}}$ we need to understand the combinatoric of the intersection (3). Next, we see that also the combinatoric from the MAC region transforms to $\mathcal{R}_{\mathrm{BIR}}$.

Proposition 1 *For $\beta \in [0,1]$ there is exactly one intersection rate pair between*

1. *$\boldsymbol{R}_1(\beta)$ and $\boldsymbol{R}_\Sigma(\beta)$ at $\beta_{1\Sigma}$ with $\boldsymbol{R}_\Sigma(\beta_{1\Sigma}) \in \mathcal{R}_{\mathrm{BIR}}$, which is the transformed vertex $\nu_{1\Sigma} \in \mathcal{R}_{\mathrm{MAC}}$;*
2. *$\boldsymbol{R}_2(\beta)$ and $\boldsymbol{R}_\Sigma(\beta)$ at $\beta_{\Sigma 2}$ with $\boldsymbol{R}_\Sigma(\beta_{\Sigma 2}) \in \mathcal{R}_{\mathrm{BIR}}$, which is the transformed vertex $\nu_{\Sigma 2} \in \mathcal{R}_{\mathrm{MAC}}$; and*
3. *$\boldsymbol{R}_1(\beta)$ and $\boldsymbol{R}_2(\beta)$ at β_{12} with $\boldsymbol{R}_1(\beta_{12}) \notin \mathcal{R}_{\mathrm{BIR}}$.*

Furthermore, we have the maximal unidirectional rates

$$R_1^\star := \max_{\boldsymbol{R} \in \mathcal{R}_{\mathrm{BIR}}} R_1 = R_{11}(1) \quad R_2^\star := \max_{\boldsymbol{R} \in \mathcal{R}_{\mathrm{BIR}}} R_2 = R_{22}(0)$$

Sketch of Proof: Simple calculation using Lemma 1 proves the combinatoric. For the maximal unidirectional rate we additionally need the monotone behavior of the components of the parametrized boundaries. □

The next corollary about the boundary of $\mathcal{R}_{\mathrm{BIR}}$ is a direct consequence from the combinatoric.

Corollary 2 *The boundary of $\mathcal{R}_{\mathrm{BIR}}$ is characterized by the section-wise defined rate pair function $\boldsymbol{R}_{\mathrm{BIR}} : [0,1] \to \mathbb{R}_+^2$,*

$$\boldsymbol{R}_{\mathrm{BIR}} : \beta \mapsto \begin{cases} \boldsymbol{R}_2(\beta), & \textit{for } \beta_{\Sigma 2} \ge \beta \ge 0 \\ \boldsymbol{R}_\Sigma(\beta), & \textit{for } \beta_{1\Sigma} > \beta > \beta_{\Sigma 2} \\ \boldsymbol{R}_1(\beta), & \textit{for } 1 \ge \beta \ge \beta_{1\Sigma}. \end{cases} \tag{5}$$

We are now ready to characterize the rate pair where the weighted rate sum is maximized in closed form. Therefore, in the next theorem we make use of all previously introduced functions and characteristic parameters.

Theorem 2 *We define the function* $\xi : \mathbb{R}_+^2 \to [0, \pi/2]$, $\boldsymbol{q} \mapsto \arctan(q_2/q_1)$ *with* $\xi([0, q_2]) = \pi/2$, *then for a given weight vector* $\boldsymbol{q} = [q_1, q_2] \in \mathbb{R}_+^2$ *the rate pair where the weighted rate sum is maximized is given*

$$\boldsymbol{R}^\star(\boldsymbol{q}) = \underset{\boldsymbol{R}\in\mathcal{R}_{\text{BIR}}}{\arg\max}\, q_1 R_1 + q_2 R_2 = \boldsymbol{R}_{\text{BIR}}(\beta^\star(\boldsymbol{q})) \qquad (6)$$

with the optimal power distribution factor, $\beta^\star(\boldsymbol{q})$,

$$\beta^\star : \mathbb{R}_+^2 \to [0,1],$$

$$\boldsymbol{q} \mapsto \begin{cases} 1, & \textit{for } \xi(\boldsymbol{q}) < \varphi_1(1), \\ \varphi_1^{-1}(\xi(\boldsymbol{q})), & \textit{for } \varphi_1(1) \le \xi(\boldsymbol{q}) \le \varphi_1(\beta_{1\Sigma}), \\ \beta_{1\Sigma}, & \textit{for } \varphi_1(\beta_{1\Sigma}) < \xi(\boldsymbol{q}) < \varphi_\Sigma(\beta_{1\Sigma}), \\ \varphi_\Sigma^{-1}(\xi(\boldsymbol{q})), & \textit{for } \varphi_\Sigma(\beta_{1\Sigma}) \le \xi(\boldsymbol{q}) \le \varphi_\Sigma(\beta_{\Sigma 2}), \\ \beta_{\Sigma 2}, & \textit{for } \varphi_\Sigma(\beta_{\Sigma 2}) < \xi(\boldsymbol{q}) < \varphi_2(\beta_{\Sigma 2}), \\ \varphi_2^{-1}(\xi(\boldsymbol{q})), & \textit{for } \varphi_2(\beta_{\Sigma 2}) \le \xi(\boldsymbol{q}) \le \varphi_2(0), \\ 0, & \textit{for } \varphi_2(0) < \xi(\boldsymbol{q}). \end{cases}$$

Sketch of Proof: For a convex set the weighted rate sum is attained at the boundary rate pair where the direction of the normal vector is equal the direction of the weight vector. Then the result follows with the previous characterization of the boundary and angle of normal vector. □

3. RELAY SELECTION

In this section we consider a scenario where node 1 and 2 with arbitrary γ_1 and γ_2 can get support by one out of N relay nodes with individual relay transmit powers constraints (respectively $\gamma_{\text{R},n}$) and channel gains $h_{1,n}$ and $h_{2,n}$, $n \in \{1,2,\ldots,N\}$. Accordingly, let $R_{\overrightarrow{1\text{R}},n}$, $R_{\overrightarrow{2\text{R}},n}$, $R_{\overrightarrow{\text{R}1},n}$, $R_{\overrightarrow{\text{R}2},n}$, and $\mathcal{R}_{\text{BIR},n}$ denote the rates $R_{\overrightarrow{1\text{R}}}$, $R_{\overrightarrow{2\text{R}}}$, $R_{\overrightarrow{\text{R}1}}$, $R_{\overrightarrow{\text{R}2}}$, and the rate region $\mathcal{R}_{\text{BIR}}$ using the n-th relay node.

3.1. Relay Selection Criteria

Since some rate pairs can be achieved with certain channel states respectively relay nodes only, for reasonable relay selection criteria we have to look at the whole two dimensional rate region. This implies that there need not be one relay node which is the best for the whole two dimensional achievable rate region. Accordingly, we can achieve the union

$$\mathcal{R}_{\text{RS}} := \bigcup_{n=1}^{N} \mathcal{R}_{\text{BIR},n}$$

by selecting for a certain rate pair the corresponding relay node that achieves this rate pair. Since a union of convex sets need not be convex, the rate region using relay selection $\mathcal{R}_{\text{RS}}$ need not be convex. Convexity is obtained if we additionally allow *time-sharing* between the usage of the relay nodes. Therewith, we achieve the rate region

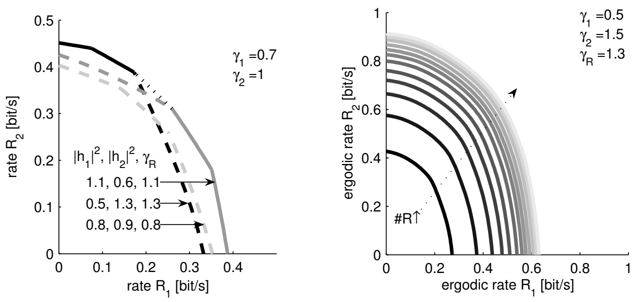
$$\mathcal{R}_{\text{RS,TS}} := \text{ConvexHull}\{\mathcal{R}_{\text{RS}}\}.$$

The boundary of $\mathcal{R}_{\text{RS,TS}}$ can be characterized by the rate pairs with the weighted rate sum maximum $\arg\max_{\boldsymbol{R}\in\mathcal{R}_{\text{RS}}} \boldsymbol{R}\boldsymbol{q}^T$ for weight vectors $\boldsymbol{q} = [q_1, q_2] \in \mathbb{R}^2$. But this is nothing else than doing relay selection for any boundary rate pair individually, what allows us to characterize the *relay selection criteria* in a single formula: For a given weight vector $\boldsymbol{q}$ let $\boldsymbol{R}_n^\star(\boldsymbol{q})$ denote the rate pair with the weighted rate sum maximum of the n-th relay node according Theorem 2. Then for this weight vector $\boldsymbol{q}$ it is optimal to select the relay node according to

$$\eta(\boldsymbol{q}) := \underset{n\in\{1,2,\ldots,N\}}{\arg\max}\, \boldsymbol{R}_n^\star(\boldsymbol{q})\boldsymbol{q}^T. \qquad (7)$$

Accordingly, we define the rate pair $\boldsymbol{R}_{\text{RS}}^\star(\boldsymbol{q}) := \boldsymbol{R}_{\eta(\boldsymbol{q})}^\star(\boldsymbol{q})$. If there are multiple solutions, we have to apply time-sharing between the relay nodes with the corresponding rate pairs to achieve all rate pairs on the boundary of $\mathcal{R}_{\text{RS,TS}}$.

Figure 2 (a) illustrates the rate region using relay selection with time-sharing for a scenario with $N = 3$ relay nodes. Thereby notice that some rate tuples can be achieved by time-sharing between two rate pairs of different relay nodes only.



(a) Achievable rate regions of three relay nodes (dashed lines), and the resulting rate region using relay selection (solid line) with time-sharing (dotted line) between the usage of relay nodes.

(b) Ergodic bidirectional achievable rate regions using relay selection with increasing number of relays $N = 1, \ldots, 12$ with iid channel coefficients $h_{k,n} \sim \mathcal{CN}(0,1)$, $\forall k, N$.

Fig. 2. Relay selection with time-sharing.

3.2. Scaling Law of Ergodic Rate Region

We now consider a scenario with N relay nodes in the presence of time-variant fading. Therefore, we assume identical and independent ergodic block-fading processes $\{h_{k,n}[m]\}_m$, $k = 1,2$, $n = 1,\ldots,N$, with a block length so that the error-free coding assumption is reasonable.

As before, the ergodic rate region is characterized by the rate pairs on its boundary, i.e. by the ergodic rate pairs with the maximal weighted rate sums

$$\overline{\boldsymbol{R}_{\text{RS}}^\star(\boldsymbol{q})} = \mathbb{E}\{\boldsymbol{R}_{\text{RS}}^\star(\boldsymbol{q})\}.$$

Again the ergodic rate region is given by the convex hull,

$$\overline{\mathcal{R}_{\text{RS,TS}}} = \text{ConvexHull}\{\overline{\boldsymbol{R}_{\text{RS}}^\star(\boldsymbol{q})} : \boldsymbol{q} = [q_1, q_2] \in \mathbb{R}^2\}.$$

In Figure 2 (b) we illustrate the enlargement of the ergodic rate region due to relay selection. Thereby, we assume equal relay transmit powers. It shows that the diversity gain decreases with increasing number of relays. In the following, we present the scaling law of this growth. For the derivation we will find bounds for the rate sum of any ergodic rate pair on the boundary using the maximal unidirectional ergodic rates. Therefore, let $R_{k,n}^\star$ denote the k-th maximal unidirectional rate using the n-th relay node, cf. Proposition 1. Then the maximal unidirectional rate using relay selection is given by

$$R^\star_{k,\mathrm{RS}} = \max_{n\in\{1,2,\dots,N\}} R^\star_{k,n}, \quad k=1,2.$$

First, we present bounds for the maximal unidirectional rates. Therefore, we define $\hat{R}^\star_{1,n} = \min\{R_{\overrightarrow{1\mathrm{R}},n}, R_{\overrightarrow{2\mathrm{R}},n}(1)\} = \log(1+\min\{\gamma_1|h_{1,n}|^2, \gamma_\mathrm{R}|h_{2,n}|^2\})$ and $\hat{R}^\star_{2,n} = \min\{R_{\overrightarrow{2\mathrm{R}},n}, R_{\overrightarrow{1\mathrm{R}},n}(0)\} = \log(1+\min\{\gamma_2|h_{2,n}|^2, \gamma_\mathrm{R}|h_{1,n}|^2\})$. From the inequalities $\frac{1}{2}\min\{x,y\} \leq \frac{xy}{x+y} \leq \min\{x,y\}$ for $x,y\geq 0$ we have $\frac{1}{2}\hat{R}^\star_{k,n} \leq R^\star_{k,n} \leq \hat{R}^\star_{k,n}$. Accordingly, we get $\frac{1}{2}\hat{R}^\star_{k,\mathrm{RS}} \leq R^\star_{k,\mathrm{RS}} \leq \hat{R}^\star_{k,\mathrm{RS}}$ with $\hat{R}^\star_{k,\mathrm{RS}} = \max_{n\in\{1,2,\dots,N\}} \hat{R}^\star_{k,n}$ for $k=1,2$.

The same inequalities hold for the maximal unidirectional ergodic rates $\frac{1}{2}\overline{\hat{R}^\star_{k,\mathrm{RS}}} \leq \overline{R^\star_{k,\mathrm{RS}}} \leq \overline{\hat{R}^\star_{k,\mathrm{RS}}}$ with $\overline{R^\star_{k,\mathrm{RS}}} = \mathbb{E}\{R^\star_{k,\mathrm{RS}}\}$ and $\overline{\hat{R}^\star_{k,\mathrm{RS}}} = \mathbb{E}\{\hat{R}^\star_{k,\mathrm{RS}}\}$ for $k=1,2$. Since $\overline{\mathcal{R}_{\mathrm{RS,TS}}}$ is convex and since the maximum of the sum is less than the sum of maximums, we have the inequalities

$$\begin{aligned}\tfrac{1}{2}\min\{\overline{\hat{R}^\star_{1,\mathrm{RS}}}, \overline{\hat{R}^\star_{2,\mathrm{RS}}}\} \leq \min\{\overline{R^\star_{1,\mathrm{RS}}}, \overline{R^\star_{2,\mathrm{RS}}}\} \leq \\ \big|\big|\overline{\boldsymbol{R}^\star(\boldsymbol{q})}\big|\big|_1 \leq \sum_{k=1}^{2}\overline{R^\star_{k,\mathrm{RS}}} \leq \sum_{k=1}^{2}\overline{\hat{R}^\star_{k,\mathrm{RS}}}\end{aligned} \tag{8}$$

for any weight vectors $\boldsymbol{q}\in\mathbb{R}^2_+\setminus\{\boldsymbol{0}\}$. Thus, we see that from a scaling law for $\overline{\hat{R}^\star_{k,\mathrm{RS}}}$ we can upper and lower bound the growth of the sum-rate of any boundary rate pair of $\overline{\mathcal{R}_{\mathrm{RS,TS}}}$. Accordingly, in the next theorem we present a tight asymptote on $\overline{\hat{R}^\star_{k,\mathrm{RS}}}$ in an iid Rayleigh fading scenario.

Theorem 3 *Let $h_{k,n}\sim\mathcal{CN}(0,\sigma^2)$ pairwise independent distributed, $\gamma_{\mathrm{R},n}=\gamma_\mathrm{R}$, and $\lambda_k = \frac{\gamma_k+\gamma_\mathrm{R}}{\sigma^2\gamma_k\gamma_\mathrm{R}}$ for all $k=1,2$ and $n=1,2,\dots,N$, then we can bound $\overline{\hat{R}^\star_{k,\mathrm{RS}}}$ as follows*

$$\begin{aligned}\log\big(1+\tfrac{1}{\lambda_k}\ln(\tfrac{N}{a})\big)\tfrac{1-\mathrm{e}^{-a}}{2} \quad \leq \quad \overline{\hat{R}^\star_{k,\mathrm{RS}}} \quad \leq \\ \log\big(1+\tfrac{1}{\lambda_k}\ln(\tfrac{N}{b})\big)\tfrac{\mathrm{e}^{-b}+b}{2} + \tfrac{b}{2}\log\big(1+\tfrac{1}{\lambda_k+\ln(\frac{N}{b})}\big)\end{aligned}$$

for any $k=1,2$ and $a,b\in(0,N)$. For the case $N\to\infty$ the asymptotic upper and lower bound meet if $a\to\infty$ and $b\to 0$. Therefore, $\frac{1}{2}\log(\ln(N))$ is a tight asymptote on $\overline{\hat{R}^\star_{k,\mathrm{RS}}}$.

Sketch of Proof: First, we see that the random variables $Z_n = \min\{a|h_{1,n}|^2, b|h_{2,n}|^2\}$ are exponential distributed with mean $(\frac{1}{a\sigma^2_{1,n}}+\frac{1}{b\sigma^2_{2,n}})^{-1}$. Then the cumulative distribution function of $Z=\max_n Z_n$ is given by the product $F_Z(z)=\prod_n F_{Z_n}(z)$, with density function $f_z(z) = N\lambda_k\,\mathrm{e}^{-\lambda_k z}(1-\mathrm{e}^{-\lambda_k z})^{N-1}$.

With substitution of $\tau = 1-\mathrm{e}^{-\lambda_k z}$ we have $\mathbb{E}\{\hat{R}^\star_{k,\mathrm{RS}}\} = \int_0^1 \frac{N}{2}\log\big(1-1/\lambda_k\ln(1-\tau)\big)\tau^{N-1}\,d\tau \geq \int_{1-\frac{a}{N}}^1\dots d\tau$ since the integrand is positive. The lower bound can be found using that the integrand is monotone increasing in τ and the inequality $\mathrm{e}^{-z}<(1+z/y)^{-y}$ for $z=\frac{wy}{w+y}$ with $w,y>0$.

Accordingly, we have $\mathbb{E}\{\hat{R}^\star_{k,\mathrm{RS}}\} \leq \int_0^x \tau^{N-1}\,d\tau\, \frac{\ln(1-1/\lambda\ln(1-x))}{2\ln(2)}N + \frac{N}{2\ln(2)}\int_x^1 \ln\big(1-1/\lambda_k\ln(1-\tau)\big)\,d\tau$ with $x=1-b/N$. We get the upper bound after solving both integrals and using the inequalities $(1-b/N)^N\leq \mathrm{e}^{-b}$ and $\mathrm{e}^\xi\,\mathrm{E}_1(\xi) < \ln(1+1/\xi)$ with $\lambda_k+\ln(N/b)=\xi$.

Since $\liminf_{N\to\infty}\frac{\mathbb{E}\{\hat{R}^\star_{k,\mathrm{RS}}\}}{\frac{1}{2}\log(\ln(N))} \geq 1-e^{-a}$ and $\limsup_{N\to\infty}\frac{\mathbb{E}\{\hat{R}^\star_{k,\mathrm{RS}}\}}{\frac{1}{2}\log(\ln(N))} \leq \mathrm{e}^{-b}+b$ we have an asymptotic lower and upper bound, which meet for $a\to\infty$ and $b\to 0$. □

It is interesting that the asymptote is independent of λ and therefore independent of the channel gain variance σ^2, power restrictions γ_R, and γ_1 or γ_2. Finally, the scaling law is a direct consequence of the theorem and the inequalities (8).

Corollary 3 *For the iid Rayleigh fading scenario with N relay nodes and $\gamma_{\mathrm{R},n}=\gamma_\mathrm{R}$ the sum of any ergodic rate pair on the boundary of the ergodic rate region $\overline{\mathcal{R}^\star_{\mathrm{RS,TS}}}$ grows with $O(\log(\log(N)))$. In more detail, we have*

$$\liminf_{N\to\infty}\frac{||\overline{\boldsymbol{R}^\star(\boldsymbol{q})}||_1}{\log(\ln(N))} \geq \frac{1}{2}, \qquad \limsup_{N\to\infty}\frac{||\overline{\boldsymbol{R}^\star(\boldsymbol{q})}||_1}{\log(\ln(N))} \leq 2.$$

4. PIGGYBACK A COMMON RELAY MESSAGE

We now consider the case where the relay node wants to transmit a common message to node one and two, this means that both nodes should decode the relay message. Therefore, in this section we consider again a scenario with one relay node only. The relay node encodes the additional relay message m_R with rate R_R using a Gaussian codebook with variance one. Let β_R denote the proportion of relay transmit power spend for the codeword of message m_R. The relay superimposes the scaled codeword on the transmit signal of the bidirectional broadcast. Now, the relay transmit power constraint requires $\beta_1+\beta_2+\beta_\mathrm{R}\leq 1$.

Since node 1 and 2 receive the codeword of their own message as interference, before decoding the unknown messages node 1 and 2 subtract the interference caused by their own message. In [4] we proved the optimal decoding order of the unknown messages in the case of equal time division. The proof carries over to the optimal time-devision case considered here.

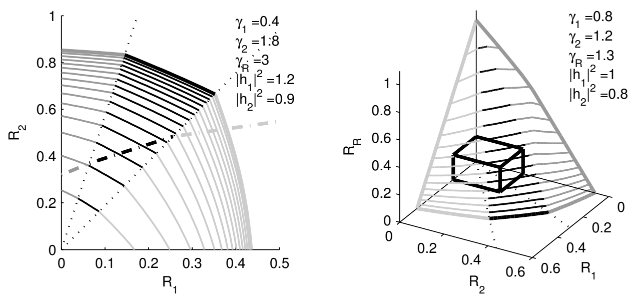
Theorem 4 *The maximal achievable additional relay rate for a desired bidirectional rate pair $[R_1,R_2]\in\mathcal{R}_\mathrm{BIR}$ is achieved if in the MAC phase node 1 and 2 transmit messages with a rate pair $[R_1^M, R_2^M]$ on the boundary of $\mathcal{R}_\mathrm{MAC}$ with $\frac{R_1}{R_2} = \frac{R_1^M}{R_2^M}$ and if in the BC phase node 1 and 2 decode the additional relay message first. Then we have the optimal time division parameter $\alpha^\star = \frac{R_1}{R_1^M} = \frac{R_2}{R_2^M}$ and the additional relay rate*

$$R_\mathrm{R}(R_1,R_2) = \min\big\{R_{\mathrm{R}@1}(R_1,R_2), R_{\mathrm{R}@2}(R_1,R_2)\big\}, \tag{9}$$

with achievable additional relay rates at node 1 and 2

$$R_{\mathrm{R}@1}(R_1,R_2) = (1-\alpha^\star)\log\big[1+\tfrac{\beta_\mathrm{R}|h_1|^2\gamma_\mathrm{R}}{1+\beta_2|h_1|^2\gamma_\mathrm{R}}\big], \tag{10a}$$

$$R_{\mathrm{R}@2}(R_1,R_2) = (1-\alpha^\star)\log\big[1+\tfrac{\beta_\mathrm{R}|h_2|^2\gamma_\mathrm{R}}{1+\beta_1|h_2|^2\gamma_\mathrm{R}}\big], \tag{10b}$$



(a) Bidirectional rate regions with uniformly increasing relay power ($\gamma_{\rm B} = 0.2, 0.4, \dots$). The dashed-dotted line denotes the bidirectional rate pairs with relay power distribution $\beta^\star$; above the line we have $R_{\rm R} = R_{\rm R@1}$ and below we have $R_{\rm R} = R_{\rm R@2}$.

(b) Comparison of the achievable rate region with comparable TDMA protocol with equal time division realizing the same routing task (cuboid). Contour lines denote achievable rate tuples for fixed $\beta_{\rm R} = \frac{1}{15}, \frac{2}{15}, \dots, 1$.

Fig. 3. Piggyback a common relay message.

with $\beta_1 = \frac{2^{\frac{R_1}{1-\alpha^\star}}-1}{|h_2|^2\gamma_{\rm R}}$, $\beta_2 = \frac{2^{\frac{R_2}{1-\alpha^\star}}-1}{|h_1|^2\gamma_{\rm R}}$, *and* $\beta_{\rm R} = 1 - \beta_1 - \beta_2$.

Sketch of Proof: The additional relay rate increases with a longer BC phase. Therefore, for a desired bidirectional relay rate pair $[R_1, R_2] \in \mathcal{R}_{\rm BIR}$ we have the longest BC phase, i.e. smallest α, if the rate pair $[R_1^M, R_2^M] \in \mathcal{R}_{\rm MAC}$ of the MAC phase is on the boundary of $\mathcal{R}_{\rm MAC}$ while it has the same ratio than the desired bidirectional rate pair, c.f. Lemma 1. With the optimal $\alpha^\star$ the proof from [4] for the equal time-division can be adapted to the optimal time-division case. □

Since it is optimal for any desired bidirectional rate tuple that the relay message is decoded first, then its interference is canceled, and in the end the bidirectional messages is decoded without any interference of the relay message, we say that the common relay message is piggybacked on the bidirectional relay communication. Effectively, this means that we have a bidirectional relay communication with reduced relay power. Accordingly, let $\gamma_{\rm B} = (1-\beta_{\rm R})\gamma_{\rm R}$ and $\gamma_{\rm P} = \beta_{\rm R}\gamma_{\rm R}$ denote the relay signal-to-noise ratios of the bidirectional relay and piggyback communication respectively. With this we normalize the optimal relay power distribution factors: For a given $\beta_{\rm R}$ we define $\beta = \frac{\beta_1}{1-\beta_{\rm R}} = 1 - \frac{\beta_2}{1-\beta_{\rm R}}$. Therewith, we can rewrite the additional relay rate of Theorem 4 as follows

$$R_{\rm R} = (1-\alpha)\log(1+\min\{\tfrac{|h_1|^2\gamma_{\rm P}}{1+(1-\beta)|h_1|^2\gamma_{\rm B}}, \tfrac{|h_2|^2\gamma_{\rm P}}{1+\beta|h_2|^2\gamma_{\rm B}}\}).$$

The terms $\frac{|h_1|^2\gamma_{\rm P}}{1+(1-\beta)|h_1|^2\gamma_{\rm B}}$ and $\frac{|h_2|^2\gamma_{\rm P}}{1+\beta|h_2|^2\gamma_{\rm B}}$ are strictly increasing and decreasing for $\beta \in [0,1]$. We have equality, and therefore maximize the min term, at the relay power distribution factor

$$\beta^\star = \tfrac{1}{2} - \tfrac{1}{2\gamma_{\rm B}}\tfrac{|h_1|^2-|h_2|^2}{|h_1|^2|h_2|^2}.$$

Accordingly, for $\beta \leq \beta^\star$ and $\beta \geq \beta^\star$ we have $R_{\rm R} = R_{\rm R@1}$ and $R_{\rm R} = R_{\rm R@2}$ for fixed α. But since the optimal time devision $\alpha^\star$ depends on β as well, the power distribution $\beta^\star$ maximizes the additional relay rate for a fixed $\beta_{\rm R}$ only if $\beta^\star \in [\beta_{\Sigma 2}, \beta_{1\Sigma}]$. This follows from the fact that $R_{\rm R@1}$ with $\alpha = \alpha_2^\star$ and $R_{\rm R@2}$ with $\alpha = \alpha_1^\star$ are strictly increasing and decreasing as well as $\alpha_\Sigma^\star$ is strictly increasing and decreasing with β for $\beta < \beta^\star$ and $\beta > \beta^\star$ respectively, i.e. $\beta^\star$ maximizes α_Σ and therefore $R_{\rm R}$ if $\beta^\star \in [\beta_{\Sigma 2}, \beta_{1\Sigma}]$ as well. If $\beta^\star \in [0, \beta_{\Sigma 2}]$ we achieve the largest additional relay rate with the power distribution that maximizes $(1-\alpha_2(\beta))\log(1+\frac{|h_2|^2\gamma_{\rm P}}{1+\beta\gamma_{\rm B}|h_2|^2})$ in the range $[\beta^\star, \beta_{\Sigma 2}]$. Accordingly, if $\beta^\star \in [\beta_{\Sigma 2}, 1]$ we achieve the largest additional relay rate with the power distribution that maximizes $(1-\alpha_1(\beta))\log(1+\frac{|h_1|^2\gamma_{\rm P}}{1+(1-\beta)\gamma_{\rm B}|h_1|^2})$ in the range $[\beta_{1\Sigma}, \beta^\star]$.

In Figure 3 we depicted the achievable rate regions for uniformly increasing $\beta_{\rm R}$. The dotted lines denote the intersections of the boundaries $\mathcal{R}_1$ and $\mathcal{R}_2$ with $\mathcal{R}_\Sigma$ and are given by the linear functions $R_2(R_1) = \frac{R_\Sigma - R_{\overrightarrow{1{\rm R}}}}{R_{\overrightarrow{1{\rm R}}}} R_1$ and $R_2(R_1) = \frac{R_{\overrightarrow{2{\rm R}}}}{R_\Sigma - R_{\overrightarrow{2{\rm R}}}} R_1$ respectively. The comparison with a simple TDMA protocol with five exclusive time slots of equal length for each transmission shows the improved resource utilization due to joint processing of two routing tasks.

5. CROSS-LAYER DESIGN

For the cross-layer design we assume a queuing model where the service rate provided from the physical layer is modeled by the previous information theoretic bidirectional rate pair. Thereby, a centralized controller chooses the service rates according a rate allocation policy.

In this section we assume a block-fading channel model, where the flat-fading channel gains are assumed to be constant during a time period T. This allows us to consider a time-slotted system model where the n-th slot denotes the time period $[(n-1)T, nT]$. Therefore, let $\boldsymbol{h}(n) = [h_1(n), h_2(n)]$ denote the channel processes with finite state space $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \times \mathcal{H}_2$ and steady-state distribution $\pi_{\boldsymbol{h}} = \pi_{h_1}\pi_{h_2}$.

We consider the scenario depicted in Figure 4, where at node 1 and 2 the packets arrive with independent negative exponential distributed inter-arrival times and average arrival rates $[\lambda_1, \lambda_2] = \boldsymbol{\lambda}$. This means that we have independent homogeneous Poisson arrival processes $[A_1(n), A_2(n)] = \boldsymbol{A}(n)$. Furthermore, we assume independent random packet length Z_i, $i = 1, 2$, with finite first and second moments at each node. Thus let $B_i(n)$, $i = 1, 2$, denote the processes of number of bits arriving in time-slot n at node 1 and 2. Then the *bit arrival rate* at node i is given by $\rho_i = \lambda_i \mathbb{E}\{Z_i\}$, $i = 1, 2$, in [bits/s]. Note that we have $\mathbb{E}\{B_i^2\} < \infty$, $i = 1, 2$, due to the previous assumptions. The packets are stored in queuing buffers with infinite size until they are served.

We observe the queue length at the end of each time-slot, therefore let $\boldsymbol{Q}(n) = [Q_1(n), Q_2(n)]$ represent the process of unfinished work in the queues after the n-th time-slot. By adjusting the optimal time-division and relay power distribution the controller decides after each time-slot for the service rates $[R_1(n+1), R_2(n+1)] \in \mathcal{R}_{\rm BIR}(\boldsymbol{h}(n))$ of the next time-slot

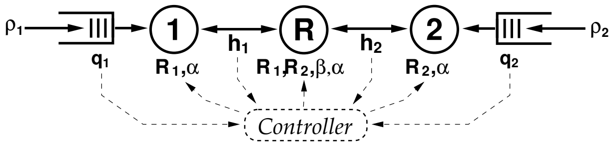


Fig. 4. Service rate allocation by a centralized controller based on the current channel and queue states in a bidirectional relaying scenario with Poisson arrival processes and queuing buffer of infinite size at two nodes.

based on the current queue state $\boldsymbol{Q}(n)$ and channel state $\boldsymbol{h}(n)$ only. Hence, the slot-to-slot dynamics of the queue backlogs are given by the equation

$$Q_i(n) = [Q_i(n-1) - R_i(n)T]_+ + B_i(n), \quad i = 1, 2.$$

Since the arrival processes are memoryless and the service rates depend on the current channel and queue state only, the process $\boldsymbol{Q}(n)$ has Markov property.

In the following, we are interested in a maximum throughput policy. The throughput is defined as the mean number of bits transmitted in a unit of time and is obviously upper bounded by the mean service rates. Therefore, let $\overline{\mathcal{R}_{\mathrm{BIR}}}$ denote the ergodic bidirectional rate region. Since $\mathcal{R}_{\mathrm{BIR}}(\boldsymbol{h})$ is convex for any channel state $\boldsymbol{h} \in \boldsymbol{\mathcal{H}}$ also the ergodic rate region $\overline{\mathcal{R}_{\mathrm{BIR}}}$ is convex. This means that the ergodic rate pair on the boundary with normal vector $\boldsymbol{q}$ is given by

$$\overline{\boldsymbol{R}^\star(\boldsymbol{q})} = \sum_{\boldsymbol{h}\in\boldsymbol{\mathcal{H}}} \pi_{\boldsymbol{h}}(\boldsymbol{h}) \operatorname*{arg\,max}_{\boldsymbol{R}\in\mathcal{R}_{\mathrm{BIR}}(\boldsymbol{h})} q_1R_1 + q_2R_2,$$

With this we get a characterization of the bidirectional ergodic rate region as follows

$$\overline{\mathcal{R}_{\mathrm{BIR}}} = \left\{\boldsymbol{R} \in \mathbb{R}_+^2 : \boldsymbol{R} \leq \overline{\boldsymbol{R}^\star(\boldsymbol{q})} \text{ with } \boldsymbol{q} \in \mathbb{R}^2\right\}. \quad (11)$$

If the arrival rate vector is outside of $\overline{\mathcal{R}_{\mathrm{BIR}}}$, i.e. average number of bits that arrive is larger than the mean service rate, then the queuing system is obviously unstable. The *stability region of a policy* is defined as the set of bit arrival rate vectors $\boldsymbol{\rho}$ such that for any vector in the interior of the stability region system stability is achieved [9]. Accordingly, a policy dominates another policy if the stability region of the one contains the other. Further, the *stability region of a system* is the set of bit arrival rate vectors $\boldsymbol{\rho}$ such that for any vector in the interior exists at least one resource allocation policy which achieves system stability. A policy that dominates any other policy is an *optimal policy*. Since the stability region of any policy is a subset of the maximum throughput region a policy which stability region is equal the maximum throughput region is optimal and is called a *maximum throughput policy* (or throughput optimal policy). In the following we present a maximum throughput rate allocation policy derived from the maximum differential backlog algorithm presented in [9] which basically tries to equalize the queue length at node 1 and 2.

Maximum Throughput Policy: The centralized network controller observes the current queue length $\boldsymbol{Q}(n) = \boldsymbol{q}$ and channel states $\boldsymbol{h}(n) = \boldsymbol{h}$ at the end of every time-slot and adjusts the optimal relay power distribution $\beta^\star$ and time-division parameter $\alpha^\star$ on the physical layer according to Theorem 2 so that in the next time-slot we achieve the rate pair which maximizes the weighted rate sum in $\mathcal{R}_{\mathrm{BIR}}(\boldsymbol{h})$ with weight vector $\boldsymbol{q} = [q_1, q_2]$,

$$\boldsymbol{R}(n+1) = \operatorname*{arg\,max}_{[R_1,R_2]\in\mathcal{R}_{\mathrm{BIR}}(\boldsymbol{h})} q_1R_1 + q_2R_2.$$

It can be shown that the stability region of the proposed policy is equal the ergodic bidirectional rate region using the well developed theory of Lyapunov drift analysis [10], [12], [11]. Therefore, one has to consider a positive quadratic Lyapunov function on the buffer levels and show that for any arrival rate vector within the ergodic rate region the Lyapunov function has a negative drift whenever the mean number of unfinished work is large. This allows one to deduce the stability-in-the-mean of a system, which is equivalent to positive recurrence and therefore the existence of a steady-state distribution for an aperiodic irreducible discrete time Markov chain. Since the proof in [11] can be easily transfered to our case, we state only the key mathematical tool here.

Theorem 5 ([11]) *Let be given the Lyapunov function* $L(\boldsymbol{q}) = \sum_{i=1}^2 q_i^2$. *If there exists a compact region* $\Lambda \subseteq \mathbb{R}^2$ *and positive values* υ *and* ζ *exist such that*

1. *whenever* $\boldsymbol{Q}(n) = \boldsymbol{q} \in \Lambda$, *there exists* $m \in \mathbb{N}, m < \infty$ *such that the probability* $\mathbb{P}(\boldsymbol{Q}(n+m) = \boldsymbol{0}) > 0$,
2. $\mathbb{E}\{L(\boldsymbol{Q}(n+1)) - L(\boldsymbol{Q}(n)) | \boldsymbol{Q}(n) = \boldsymbol{q}\} < \upsilon - \zeta \sum_{i=1}^2 q_i$

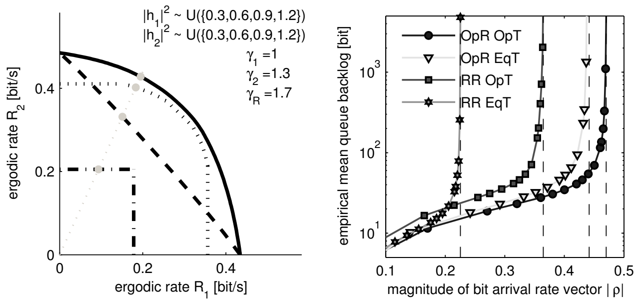
then there exists a steady-state distribution with bounded first moments $\mathbb{E}\{Q_i\} < \infty$ *such that* $\zeta \sum_{i=1}^2 \mathbb{E}\{Q_i\} < \upsilon$.

The first condition ensures that the zero state is reached infinitely often with finite mean recurrence times and therefore the Markov chain reduces to a single ergodic class. It is a necessary modification for queuing systems with uncountably infinite state space. The fundamental idea is that if the drift gets larger in magnitude as the queue lengths increase, then the first moment of the queue length is bounded (strong stability of the Markov chain).

The here proposed policy is equivalent to the dynamic power allocation policy in [11]. It is therefore possible to adapt the proof in [11] with the following constant

$$\upsilon = T^2 \max_{\boldsymbol{h}\in\boldsymbol{\mathcal{H}},\boldsymbol{R}\in\mathcal{R}_{\mathrm{BIR}}(\boldsymbol{h})} \left(R_1^2 + R_2^2\right) + \sum_{i=1}^2 \mathbb{E}\{B_i^2\}.$$

Since the arrival rate vector $\boldsymbol{\rho}$ is assumed to be strictly in the interior of the ergodic rate region, there exists a $\tilde{\zeta} > 0$ so that $\tilde{\zeta}\mathbb{1} + \boldsymbol{\rho} \in \mathrm{int}\overline{\mathcal{R}_{\mathrm{BIR}}}$ also holds. Then the proof works analog with $\zeta = 2T\tilde{\zeta}$.



(a) The solid, dashed, dotted, and dashed-dotted line denote the ergodic achievable rate region of OpR OpT, OpR EqT, RR OpT, and RR EqT respectively.

(b) The mean arrival rate vector $\boldsymbol{\rho}$ is increased along the radial line in Figure (a). After 10^6 time-slots the queue length strongly grows if $\boldsymbol{\rho}$ approaches the boundary of the stability region.

Fig. 5. Cross-layer design.

Finally, let $\mathbb{E}\{D_i\}$ denote the average bit delay at node i. With Little's Theorem we have $\mathbb{E}\{Q_i\} = \rho_i \mathbb{E}\{D_i\}$. Therefore, we have $\upsilon/\zeta > \sum_{i=1}^{2} \mathbb{E}\{Q_i\} = \sum_{i=1}^{2} \mathbb{E}\{D_i\}\rho_i$ using the boundedness of the first moment according to Theorem 5. This means that the bound grows asymptotically like $1/\zeta$ as the arrival rate vector $\boldsymbol{\rho}$ is pushed towards the boundary of the ergodic rate region. A similar discussion is given in [11].

To clarify the results of this section in Figure 5 we illustrate and discuss numerical queuing simulation results of a simple example with uniformly distributed channel processes. In order to indicate the performance gain, we present a comparison between the proposed optimal time-division (OpT) with optimal relay power distribution (optimal achievable rates - OpR) approach, equal time-division (EqT) with optimal relay power distribution (see [8] for more details), as well as Round-Robin (RR) scheduling with optimal and equal time-division (OpT/EqT). A protocol is called a Round-Robin scheduling when for any communication between two nodes the time-slot is subdivided into exclusive time-intervals. The analysis of those scenarios is similar and gives no further insights.

6. CONCLUSION

Two phase bidirectional relaying using interference cancellation is spectral efficient since it avoids the inherent spectral loss of unidirectional protocols. The principle results found for bidirectional relaying with equal time division also hold for the optimized time division. However, the analytic characterization of the convex achievable rate region gets more involved.

With relay selection we can achieve the same multi-user space diversity order $O(log(log(N)))$ as distributed beamforming, which additionally requires coherent transmission between distributed relay nodes. From an additional multicast of a relay message on the bidirectional relaying scheme we see that significant performance gains can be expected if multiple routing schemes are considered jointly. Finally, the complete characterization of the bidirectional achievable rate region allows the design of a throughput optimal rate allocation policy in a cross-layer design. It shows that a coordinated resource allocation achieves a significant larger stability region than a simple Round Robin strategy.

7. REFERENCES

[1] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[2] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler, "Fading relay channels: Performance limits and space-time signal design," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[3] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Asilomar Conference on Signals, Systems, and Computers 2005*, Nov. 2005.

[4] T. J. Oechtering and H. Boche, "Optimal Resource Allocation for a Bidirectional Regenerative Half-duplex Relaying," in *IEEE International Symposium on Information Theory and its Applications (ISITA 06)*, Seoul, Korea, 2006, pp. 528 – 533.

[5] R. W. Yeung, "Network Coding Theory Part I/II: Single/Multiple Source," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 4–5, pp. 241–381, 2005.

[6] A. F. Dana, M. Sharif, R. Gowaikar, B. Hassibi, and M. Effros, "Is broadcast plus multi-access optimal for gaussian wireless networks," in *Asilomar Conference on Signals, Systems, and Computers 2003*, Nov. 2003.

[7] T. J. Oechtering and H. Boche, "Piggyback a Common Message on," in *Proc. Workshop on Wireless Networks - Communication, Cooperation and Competition (WNC3-'07)*, 2007, accepted.

[8] ——, "Stability Region of an Efficient Bidirectional Regenerative Half-duplex Relaying Protocol," in *IEEE Information Theory Workshop (ITW 06)* , Chengdu, China, Oct. 2006, pp. 380–384.

[9] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 466–478, 1993.

[10] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. London: Springer-Verlag, 1993.

[11] M. J. Neely, E. Modiano, and C. E. Rohrs, "Power Allocation and Routing in Multibeam Satellites with Time-Varying Channels," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 138–152, Feb. 2003.

[12] P. R. Kumar and S. P. Meyn, "Stability of queueing networks and scheduling policies," *IEEE Transactions on Automatic Control*, vol. 40, pp. 251–260, Feb. 1995.