Linear Transmit-Receive Beamforming Design for MIMO Relay Broadcast Channels

*Edwin Monroy, †Sunghyun Choi, and ‡Bijan Jabbari

*International IT Policy Program, College of Engineering, Seoul National University, Seoul, Korea
†Department of Electrical and Computer Engineering and INMC, Seoul National University, Seoul, Korea
‡Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, USA

Email: emonroy@mwnl.snu.ac.kr, schoi@snu.ac.kr, bjabbari@gmu.edu

Abstract—This paper studies the relay broadcast channel with one source, one decode-and-forward half-duplex relay, and two destinations, with one of them receiving data directly from the source and the other being cooperatively served by the source and the relay. Such a scenario models a situation that may arise in practice in the downlink channel of a single cell in cellular systems. We derive the achievable sum rate of the proposed scheme in the form of a non-convex optimization problem that is hard to solve. Thus, we propose a low-complexity, suboptimal solution that employs a linear transmit-receive beamforming scheme based on block diagonalization and interference alignment with resource allocation. The numerical results confirm that the proposed scheme achieves a higher rate than that of the reference schemes in a symmetric scenario, with the channel between source and one of the destinations being relatively poor.

Index Terms—MIMO relay broadcast channel, half-duplex relay, block diagonalization, interference alignment, repetition coding.

I. INTRODUCTION

The relay channel is a classic problem in network information theory that has been studied quite intensively in the past, and it still remains an open problem. One variation of the traditional three-node relay channel that is particularly applicable to cellular networks is the relay broadcast channel (RBC), first introduced in [1], in which a transmitter simultaneously sends messages to a number of receivers with the help of one or several relays. Kramer et al. [1] give an achievable rate region for RBCs with one source, one full-duplex decode-and-forward (DF) relay, and two destinations.

For the case when all or certain nodes have multiple antennas, several papers have introduced diverse precoding/decoding schemes considering a more practical half-duplex relay node. Some of them include [2], [3], in which multiple destinations receive data from the source via a single relay. However, the assumption there is that all destinations receive their signals from the relay, and the source does not serve any users directly, which is a scenario that could arise in practice, e.g., in the downlink of cellular systems. Additionally, these works ignore the direct link between source and destinations, and hence, do not take advantage of the possible increase in diversity or multiplexing gain that the direct link could offer when available. Other works that do include the direct link, such as [4], [5], do not consider the relay and source transmitting data simultaneously to separated groups of users, who are only served by either the base station or the relay. On the other hand, authors of [6] study a Multi-Input Multi-Output (MIMO) RBC where the destinations are separated in direct and relay users although the direct link of relay users is again neglected, so that there is no cooperation between base station and relay. Moreover, all the above-mentioned works study the case where the relay operates in amplify-and-forward (AF) mode only. In contrast, there are much fewer studies on the design of transmit/receive beamforming schemes for the DF RBC. Along with the information-theoretic work of Kramer et al. in [1], [7] considers an RBC with multiple DF relays and designs a centralized scheduling scheme to maximize the sum rate with limited feedback. However, the direct link is again not considered.

In this paper, we study a type of RBC with one source, one half-duplex DF relay, and two destinations. It is assumed that one of the destinations (denoted as D1) is distant from the source (e.g., on the cell edge) and requires the help of the relay to properly decode source’s transmission, whereas the other destination (D2) is relatively closer to the source and is therefore able to receive its signal directly. Both destinations are served simultaneously by the source, and by the relay in case of D1. We express the achievable sum rate of our model as an optimization problem that turns out to be highly complex and difficult to solve. Therefore, a low-complexity, linear transceiver design based on interference alignment (IA) and block diagonalization (BD) with resource (in terms of time-slot duration, transmit power, and spatial streams) allocation is proposed. Our scheme keeps the implementation complexity low by making use of practical techniques such as repetition coding, half-duplex relaying, linear transceiver, and non-iterative solutions. The remainder of the paper is structured as follows. We describe the system model in Section II. The achievable sum rate of the proposed scheme is formulated in Section III. Section IV presents the proposed design of the MIMO transceivers (i.e., transmit precoders and receive filters) based on BD and IA with resource allocation and its resulting sum rate. Finally, a set of numerical results is shown in Section V, while our conclusion follows in Section VI.
II. SYSTEM MODEL

The system under study in this paper consists of four nodes, namely, one source, one DF relay, and two destinations, each one with $N_S$, $N_R$, $N_{D1}$, and $N_{D2}$ antennas, respectively. The MIMO channels from source to relay (S-R), source to each destination (S-D1 and S-D2, also called direct links), and relay to each destination (R-D1 and R-D2) are referred to as $H_{SR} \in \mathbb{C}^{N_S \times N_R}$, $H_{SD1} \in \mathbb{C}^{N_S \times N_{D1}}$, $H_{SD2} \in \mathbb{C}^{N_S \times N_{D2}}$, $H_{RD1} \in \mathbb{C}^{N_R \times N_{D1}}$, and $H_{RD2} \in \mathbb{C}^{N_R \times N_{D2}}$, respectively, and are random and independent matrices whose elements are independent and identically distributed (i.i.d) zero mean, unit variance, circular symmetric complex Gaussian random variables. This means that these channel matrices are full rank with probability one. All the channels follow a frequency-flat block-fading model, and are assumed to be perfectly known at all nodes. The source and relay nodes are subject to maximum transmit power constraints $P_S$ and $P_R$, respectively. The noise at the receivers is an i.i.d complex Gaussian random vector with zero mean and identity covariance matrix, and independent of the transmitted signals.

In our system, each destination node is only interested in receiving its own signal, and the other node’s signal is regarded as interference. It is also assumed that the relay cannot transmit and receive at the same time so that it operates in half-duplex mode in time, i.e., any packet transmission is completed in two phases or time slots, and the channel matrices are assumed to remain constant during both time slots. The proposed transmission protocol is illustrated in Fig. 1. During the first time slot, the source transmits signal $x_S^{(1)} \in \mathbb{C}^{N_S \times 1}$, which consists of three independent parts, denoted as $x_S^{(1)}$, $x_{D1}$, and $x_{D2}$ that are respectively received by the relay and each destination. Specifically, the message to D1 is split into two parts: a cooperative part, $x_S^{(1)}$, that is first transmitted to the relay and then cooperatively sent to D1 by source and relay; and a direct part, $x_{D1}$, intended to be directly transmitted over the S-D1 link (similar to [8]). These three signals are opportunistically transmitted to maximize the achievable sum rate, as explained in Section IV-C.

We employ linear precoders to separate the three signals such that $x_S^{(1)} = F_F x_F^{(1)} + F_{D1} x_{D1} + F_{D2} x_{D2}$. Here, $F_F \in \mathbb{C}^{N_S \times m_f}$, $F_{D1} \in \mathbb{C}^{N_S \times m_{d1}}$, and $F_{D2} \in \mathbb{C}^{N_S \times m_{d2}}$ are the source precoding matrices in the first time slot; $x_F^{(1)} \in \mathbb{C}^{m_f \times 1}$, $x_{D1} \in \mathbb{C}^{m_{d1} \times 1}$, and $x_{D2} \in \mathbb{C}^{m_{d2} \times 1}$; and $m_f$, $m_{d1}$, and $m_{d2}$ represent the number of spatial streams (SSs) used to transmit each signal, with $m_f + m_{d1} + m_{d2} = N_S$. Signals $x_F^{(1)}$, $x_{D1}$, and $x_{D2}$ are independent, complex Gaussian vectors with zero mean and identity covariance matrix. The source precoding matrices shall satisfy the power constraint $\text{tr}(F_F^* F_F + F_{D1}^* F_{D1} + F_{D2}^* F_{D2}) \leq P_S$.

The received signals at relay, D1, and D2 in the first time slot are given, respectively, by

$$y_R = H_{SR} x_S^{(1)} + n_R^{(1)}$$
$$y_{D1} = H_{SD1} x_{D1}^{(1)} + n_{D1}^{(1)}$$
$$y_{D2} = H_{SD2} x_{D2}^{(1)} + n_{D2}^{(1)}$$

where $n_i^{(1)}$, $i = 1, 2$, represent the noise vectors at the relay and the destinations.

The relay processes $y_R$ to obtain an estimate of the source’s transmission, referred to as $x_R^{(2)}$, which is then precoded and forwarded to D1 as $x_R = G_R x_R^{(2)}$, where $G_R \in \mathbb{C}^{N_R \times m_f}$ is the relay precoding matrix, with $m_f$ being the number of SSs used to transmit $x_R$. In processing $y_R$, the relay does not decode $x_{D1}$ or $x_{D2}$ and regards them as interference. Similar treatment is given to $x_F^{(1)}$ and the other destination’s signal by D1 and D2. Hence, after the source’s transmission is completed in the first time slot, D1 has yet to receive a copy of the cooperative part of the source message. In order for the source and the relay to cooperatively transmit such message to D1, the source also sends a precoded copy of $x_F^{(2)}$ in the second time slot [8]. That is, repetition coding is used at source and relay for signal $x_F^{(2)}$.

In addition, the source transmits a new signal $x_{D2}^{(2)}$ toward D2.

Hence, the signal transmitted by the source during the second time slot is given by $x_S^{(2)} = F_F x_F^{(2)} + F_{D2} x_{D2}^{(2)}$, where $F_F \in \mathbb{C}^{N_S \times m_f}$ and $F_{D2} \in \mathbb{C}^{N_S \times m_{d2}}$ are the source precoding matrices in the second time slot; and $x_F^{(2)} \in \mathbb{C}^{m_f \times 1}$ and $x_{D2}^{(2)} \in \mathbb{C}^{m_{d2} \times 1}$ are independent, complex Gaussian vectors with zero mean and identity covariance matrix. Moreover, $m_f$ and $m_{d2}$ represent the number of SSs used to transmit each signal, with $m_f + m_f = N_S$. Note that $m_f = m_f$, since, as described above, both relay and source send the same signal $x_F^{(2)}$.

The precoding matrices at source and relay shall also meet the power constraints $\text{tr}(F_F^* F_F + F_{D2}^* F_{D2}) \leq P_S$, and $\text{tr}(G_R G_R^H) \leq P_R$, respectively.

Finally, the signals received at D1 and D2 in the second time slot are, respectively,

$$y_{D1}^{(2)} = H_{SD1} x_{D1}^{(2)} + H_{RD1} x_R + n_{D1}^{(2)}$$
$$y_{D2}^{(2)} = H_{SD2} x_{D2}^{(2)} + H_{RD2} x_R + n_{D2}^{(2)}$$

where $n_{D1}^{(2)}$, $i = 1, 2$, is the noise vector at Di.

In the following sections, we determine the optimal achievable sum rate for the system model described here and design the MIMO transceivers at each node in such a way that they eliminate the unwanted signals, while at the same time they achieve the highest rate possible by efficiently allocating the time-slot duration, the available SSs, and the transmit power.

III. ACHIEVABLE SUM RATE

From the description in the previous section, it can be seen that our system is equivalent to the RBC described in [1,
Section III.D] with no common message and no message from relay to D2. Hence, it can be proved that the capacity region of our system can be obtained from [1, Theorem 2] with $U_i = x_{i,2}$, $U_0 = x_{0,2}$, $i = 1, 2, 3$, and no I_{12} or I_{23}, and it is given by the convex closure of the set of rates $(R_1, R_2)$ to each destination D1 and D2 respectively, such that

$$R_1 < R_{1}^\text{max} = \min \left\{ t I \left( x_1^1; y_1 \right) + (1-t) I \left( x_2^1; y_1^1 \right) | x_1 \right\},$$

$$R_2 < R_{2}^\text{max} = t I \left( x_1^2; y_2 \right) + (1-t) I \left( x_2^2; y_2^2 \right) \right\} + t I \left( x_1^1; y_1^1 \right),$$

where $t$ denotes the duration of the first time slot and $1-t$, that of the second. Notice that $R_1$ is upper bounded by the achievable rate of a classic half-duplex decode-and-forward relay channel [1], whereas $R_2$ is upper bounded by the rate of a point-to-point MIMO channel. We are interested in the achievable sum rate of the system, denoted by $R_{\text{prop}}$, which is to be optimized over the time-slot duration, the source and relay precoding matrices, and the number of SSs allocated to the input signals:

$$R_{\text{prop}} = \max_{t \in (0,1), F_1, F_2, G_0, G_1, G_2} \left( R_1^\text{max} + R_2^\text{max} \right)$$

subject to

$$\text{tr} \left( F_1^H F_1 + F_1^H D_1 F_1^H \right) \leq P_S$$

$$\text{tr} \left( F_1^H D_2 F_2^H + F_2^H D_2^H \right) \leq P_S,$$

$$\text{det} \left( F_1^H H_{SR} F_1 + F_1^H H_{I1} F_1 + F_2^H D_2^H \right) = 0$$

The expressions for $R_{1}^\text{max}$ and $R_{2}^\text{max}$ can be obtained similarly to [8]. For $R_{1}^\text{max}$ we have

$$I(x_1^1; y_1^1) = \log_2 \left( \frac{\det(I_{N_R} + H_{SR} F_1 D_1 F_1^H + F_1^H F_1^H)}{\det(I_{N_R} + H_{SR} F_1 D_1 F_1^H)} \right)$$

$$I(x_1^2; y_1^2) = \log_2 \left( \frac{\det(I_{N_D1} + H_{SD1} F_1 D_1 F_1^H + F_2^H D_2^H)}{\det(I_{N_D1} + H_{SD1} F_1 D_1 F_1^H + F_2^H D_2^H)} \right)$$

and $H_{D1} = [H_{SD1} \quad H_{RD1}]$ and

$$Q^{(2)} = \begin{bmatrix} F_2^H F_2^H & F_2^H G_R H_R^H \\ G_R F_2^H & G_R G_R^H \end{bmatrix}.$$
\[ V_{D2}^{(1)} = \text{null}\left( \left[ H_{SR}^H \, H_{SD1}^H \right]^H \right), \] where \text{null}(A) denotes an orthonormal basis for the nullspace of the rectangular matrix \( A \). These matrices transform the three links into three independent, interference-free single-user MIMO (SU-MIMO) equivalent links over which SVD with water filling is optimal. Thus, we make \( B_{F1}^{(1)} = V_{SR}, \ B_{D1}^{(1)} = V_{SD1}, \) and \( B_{D2}^{(1)} = V_{SD2}, \) where the RHSs are the matrices of right-singular vectors of the equivalent channel matrices \( H_{SR} V_{F1}^{(1)}, \ H_{SD1} V_{D1}, \) and \( H_{SD2} V_{D2}^{(1)}. \) In turn, the power allocation matrices are obtained by water filling over the eigenmodes of these equivalent channels, with \( \text{tr}(P_F^{(1)} + P_{D1} + P_{D2}^{(1)}) \leq P_S. \)

**B. Second Time Slot**

In the second time slot, signal \( x_{D2}^{(2)} \) from relay and source causes interference at D2, and so does \( x_{D1}^{(2)} \) at D1. In order to remove these unwanted signals, we propose an approach inspired by IA [9] that, as in the first time slot with BD, creates interference-free equivalent channels to each destination, over which single-user decoding of the desired signals can be performed, as described in the following.

The proposed transceivers consist of two stages that are designed independently [10]: the first stage performs interference alignment at the transmitters and nullifying at the receivers [9]; the second stage performs single-user SVD and waterfilling over the resulting channel. Accordingly, we write \( F_F = F_{IA}^{(2)} F_{SU} \) and \( G_R = G_{R,IA} G_{R,SU} \) for the transmit precoders; and \( U_{D1}^{(2)} = U_{D1,IA}^{(2)} U_{D1,SU} \) and \( U_{D2}^{(2)} = U_{D2,IA}^{(2)} U_{D2,SU} \), for the receive decoders. Precoder \( F_{F} \) is not expressed in the same way and only consists of the single-user part, as will become clear in the following, so that it is just \( F_{D2}^{(2)} = F_{D2}^{(2),SU}. \) Here, the IA part of the transceivers is denoted by subindex \( IA \) while the single-user part is denoted by \( SU \).

We start by rewriting the receive signals at D1 and D2 at the output of the receive filters, considering only the IA stage of the transceivers, as

\[
\begin{align*}
U_{D1,IA}^{(2)} y_{D1}^{(2)} = & U_{D1,IA}^{(2)} \left( H_{SD1} F_{IA}^{(2)} x_F^{(2)} + H_{RD1} G_{R,IA} x_F^{(2)} \right) \\
& + H_{SD2} F_{D2}^{(2),SU} x_{D2}^{(2)} + n_{D1}^{(2)} \tag{3}
\end{align*}
\]

\[
\begin{align*}
U_{D2,IA}^{(2)} y_{D2}^{(2)} = & U_{D2,IA}^{(2)} \left( H_{SD2} F_{D2}^{(2),SU} x_{D2}^{(2)} + H_{SD2} F_{IA}^{(2)} x_F^{(2)} \right) \\
& + H_{RD2} G_{R,IA} x_F^{(2)} + n_{D2}^{(2)} \tag{4}
\end{align*}
\]

From these expressions, we first notice that D1 receives a single interference signal, so that we can choose its receiver filter such that \( U_{D1,IA}^{(2)} = \text{null}(H_{SD1} F_{D2}^{(2),SU})^H \), where \( F_{D2}^{(2),SU} \) is obtained later. At D2, on the other hand, we first have to select the source and relay precoders such that the interference is aligned in such a way that the dimensions of the interference signals is less than or equal to \( m_{D2}^{(2)} \), instead of \( m_{D2}^{(2)} + m_{D2}^{(2)} \) without alignment. By doing this, up to \( m_{D2}^{(2)} = M - m_{D2}^{(2)} \) dimensions are left for the intended signal \( x_{D2}^{(2)} \). Such an alignment can be expressed as \( C(H_{SD2} F_{IA}^{(2)}) = C(H_{RD2} G_{R,IA}) \), where \( C(A) \) denotes the column space of \( A \). We could solve this equation by means of an iterative algorithm over \( G_{R,IA} \) or \( F_{IA}^{(2)} \) until convergence of certain criterion (e.g., the Frobenius norm) is reached. Nonetheless, we decide to employ a non-iterative alternative instead with a view to keeping the complexity of the proposed scheme low. A reasonable guess for \( G_{R,IA} \) could be the first \( m_{D2}^{(2)} \) left-singular vectors from the SVD of \( H_{RD2} \). This allow us to readily determine \( F_{IA}^{(2)} = C(H_{RD2} G_{R,IA}), \) since the channel matrices are invertible, with \( O(A) \) being an orthonormal base for \( C(A) \). After the alignment, the decoding matrix at D2 can now be chosen in the same way as for D1, that is, \( U_{D2,IA}^{(2)} = \text{null}(H_{SD2} F_{D2}^{(2),SU})^H \), or equivalently \( U_{D2,IA}^{(2)} = \text{null}(H_{RD2} G_{R,IA})^H \).

We now focus on the second stage of the transceivers. By plugging the IA-based source and relay precoders and destination decoders into (3) and (4), and by including their corresponding single-user parts, the received signal at each destination now becomes

\[
\begin{align*}
U_{D1,SU}^{(2)} y_{D1}^{(2)} = & U_{D1,SU}^{(2)} \left( H_{SD1} F_{SU} x_{D1}^{(2)} + H_{RD1} G_{SU} x_{SU}^{(2)} \right) + n_{D1}^{(2)} \\
U_{D2,SU}^{(2)} y_{D2}^{(2)} = & U_{D2,SU}^{(2)} \left( H_{SD2} F_{SU} x_{D2}^{(2)} + H_{RD2} G_{SU} x_{SU}^{(2)} \right) + n_{D2}^{(2)}
\end{align*}
\]

where \( y_{D1}^{(2)} = U_{D1,IA}^{(2)} D1 \) and \( y_{D2}^{(2)} = U_{D2,IA}^{(2)} D2 \), \( H_{SD1} = U_{D1,IA}^{(2)} H_{D1,IA} \) and \( H_{SD2} = U_{D2,IA}^{(2)} H_{D2,IA} \), \( H_{RD1} = U_{D1,IA}^{(2)} H_{D1,IA} \) and \( H_{RD2} = \text{null}(H_{RD2} G_{R,IA}), \) and \( H_{RD2} = \text{null}(H_{RD2} G_{R,IA}). \) It can be seen that after applying the IA stage of the transceivers, both destinations see an equivalent channel with no interference. However, while at D2 such channel is a simple SU-MIMO channel over which only \( x_{D2}^{(2)} \) is received, D1 sees two channels, one from the source and one from the relay, over which two copies of \( x_{D2}^{(2)} \) are received and have to be appropriately combined. Therefore, at D2 we can apply SVD with water filling to obtain the second stage of the transceivers, such that

\[
\begin{align*}
U_{D2,SU}^{(2)} = U_{D2,SU}^{(2)}, \quad F_{D2,SU}^{(2)} = V_{SD2} G_{SU}^{(2)} \left( \begin{array}{c} 0 \\ 1 \end{array} \right), \quad \text{where} \ U_{D2,SU}^{(2)} \quad \text{and} \quad U_{SD2} \quad \text{are the matrix of left and right-singular vectors of} \quad H_{SD2} \quad \text{respectively,} \quad \text{and} \quad \tilde{P}_{SU} \quad \text{is a diagonal power allocation matrix obtained by water filling over the singular values of} \quad H_{SD2}, \quad \text{subject to} \quad tr(\tilde{P}_{SU}^{(2),SU}) \leq (1 - \beta) P_S, \quad \text{where} \ \beta \ \text{is defined as} \ \text{follows. Since the source transmits} \ x_{F}^{(2)} \ \text{and} \ x_{D2}^{(2)} \ \text{to each destination, it has to appropriately allocate its total transmit power} \ P_S \ \text{between them. Parameter} \ \beta \ \in [0, 1] \ \text{is defined as the fraction of} \ P_S \ \text{allocated to} \ x_{F}^{(2)}, \ \text{while the remaining} \ (1 - \beta) \ \text{is allocated to} \ x_{D2}^{(2)}.
\end{align*}
\]

As for D1, let us first assume that the relay and the source can pool their transmit power for \( x_{F}^{(2)} \) as \( \beta P_S + P_R \), and define the composite channel matrix \( H_{SU} = [H_{SD1} \, H_{SD2}] \), as in [12]. By doing this, the composite source-relay precoding matrix is obtained by SVD and water filling, such that \( F_{SU} = V_{SU,1} \beta P_D \), where \( V_{SU,1} \) is the matrix of right-singular vectors of the composite channel matrix, and \( P_D \) is the water-filling power matrix, such that \( \text{tr}(P_{SU}) \leq \beta P_S + P_R. \) Now, let \( F_{SU,1} \) be a matrix formed with the first \( m_{F}^{(2)} \) rows of \( F_{SU} \), and \( G_{SU,2} \), with the remaining \( m_{F}^{(2)} \) rows. We can finally obtain the single-user matrices as \( F_{SU}^{(2)} = \sqrt{\alpha_1} F_{SU,1} \) and \( G_{SU}^{(2)} = \sqrt{\alpha_2} G_{SU,2} \), where the factors \( \alpha_1 \) and \( \alpha_2 \) scale the composite source-relay precoding matrix to satisfy the individual power constraints of
the source and relay respectively, and are given by
\[ \alpha_1 = \frac{\beta P_s}{\text{tr}(F_{SU_1}F_{SU_1}^H)}, \quad \alpha_2 = \frac{P_R}{\text{tr}(G_{SU_2}G_{SU_2}^H)}. \]

Finally, \( U_{D1}^{(t)} = U_{D1} \), with \( U_{D1} \) being the matrix of left-singular vectors of \( H_{D1} \).

C. Suboptimal Sum Rate of the Proposed Scheme

Having obtained the MIMO transceivers for the first and second time slots for all the nodes in the system, we can return to problem (1) to determine the maximum sum rate that our scheme can achieve with the proposed transceivers. It is important to note that the decoders at the relay, D1, and D2 are unitary matrices, implying that the noise vectors remain statistically unaltered with the same mean and variance when pre-multiplied by them, and that the mutual information expressions are also invariable to such decoders. Therefore, we can plug the expressions for the proposed transceivers in (1), and after rearranging some terms, obtain \( R_{\text{prop}} \), the achievable sum rate of our scheme with the designed transceivers:

\[
R_{\text{prop}} = \max_{\epsilon \in (0,1], \beta \in [0,1]} \min_{m_1, m_2} \left( R_a, R_b \right)
\]

where \( R_a = a_1 t + a_2 (1 - t) \), and \( R_b = b_1 t + b_2 (1 - t) \), and

\[
a_1 = \log_2(\det(I_M + H_{SR}F_{SF}^H F_{FL}^H H_{SR}^H)) + b_1
\]

\[
a_2 = \log_2(\det(I_M + H_{SD}F_{DF}D_{SF}^H))
\]

\[
b_1 = \log_2(\det(I_M + H_{SD}F_{DF}D_{DF}^H D_{DF}^H H_{SD}^H)) + \log_2(\det(I_M + H_{SD}F_{DF}D_{DF}^H D_{DF}^H H_{SD}^H))
\]

\[
b_2 = \log_2(\det(I_M + H_{DI}Q_{DI}^H H_{DI}^H)) + a_2.
\]

This resulting optimization problem does not seem convex, particularly with respect to \( \beta \); moreover, it is not immediately clear what transformation or other manipulation would convert it into a convex form. Instead of trying to solve (5) via standard convex optimization techniques, we propose a suboptimal but much simpler way to separately obtain the three parameters of the problem, namely, the number of SSs for each transmitted signal, the source power allocation factor in the second time slot, and the duration of the time slots.

Parameters \( m_1, m_2 \), and \( m_2 \), \( i = 1, 2 \) could be found by exhaustive search, by evaluating (5) for all their possible combinations (while keeping the other two parameters fixed), and choosing the one that yields the highest rate. Clearly, this approach is practical only when \( M \) is a small number. Otherwise, they could be set to a certain fixed value, (e.g., equal to \( \frac{M}{2} \) in the first time slot and equal to \( \frac{M}{2} \) in the second) to reduce complexity. Yet another alternative is to employ more intelligent techniques such as multimode selection for MU-MIMO systems [13]. Parameter \( \beta \) is set to let the source allocate its power proportionally to the number of SSs over which each signal is transmitted, that is, \( \beta = \frac{m_2}{m_1} \). Finally, the time-slot duration \( t \) can be determined by noticing that when the other two decision variables are fixed, the optimization in (5) becomes a linear program of the form \( \max \min(a_1 + \beta t, b_2 + t) \), where \( m = a_1 - a_2 \) and \( n = b_1 - b_2 \), and \( m \geq n \).

Then, it can be shown that the optimal value of \( t \) depends on the signs of these slopes and their combinations:

\[
t = \begin{cases} 
1 & \text{if } m > 0, n \geq 0 \\
0 & \text{if } m \leq 0, n < 0 \\
\frac{b_2 - a_2}{m - n} & \text{if } m < 0, n > 0 \text{ or } m > 0, n < 0.
\end{cases}
\]

However, since the trivial cases \( t = 1 \) and \( t = 0 \) are of no practical interest in our scenario with a half-duplex relay, we set \( t = \frac{b_2 - a_2}{m - n} \), where it can be verified that \( 0 < t < 1 \), i.e., the relay is used.

V. NUMERICAL RESULTS

In this section we provide some numerical results to evaluate the achievable sum rate of the proposed scheme. Unless otherwise noted, \( M = 4 \), \( P_S = P_R = 1 \), and the results are averaged over 200 independent channel realizations. The SNR is defined as the average power at each receive antenna over the noise power. Since we have assumed that D1 sees a weak S-D1 channel, we set its SNR to \(-5 \) dB (unless stated otherwise), while that of all other links is the same and varied. Finally, we set \( m_1 = \lfloor \frac{M}{2} \rfloor \), \( m_2 = \lceil \frac{M}{2} \rceil \), \( i = 1, 2 \) and \( m_1 = 0 \) and do not optimize them for simplicity. We compare the sum rate of the proposed scheme obtained as described in Section IV-C (denoted by “Proposed” in the figures) with the following two schemes:

1) Direct link: The two destinations are served directly by the source, and the relay is not used. Such a setup corresponds to a MIMO BC channel with two receivers, so we can use the same technique employed in Section IV-A, now with only two receivers. The number of SSs allotted to each destination is the same as in the second time slot of the proposed scheme.

2) TDMA: The transmissions to the destinations are orthogonally separated in time and completed in two time slots of equal duration. In the first, the source transmits data to D1 with the help of the relay, and such transmission is completed, in turn, in two phases as well, whose duration is optimized in the same way as in Section IV-C. The transceivers are designed in the same way as in Section IV-A. In the second time slot, the source serves D2 directly.

Figure 2 shows that the proposed scheme outperforms the other two, especially at high SNR. Such behavior is expected because the BD and IA techniques employed in the first and second phases are known to be capacity-achieving in the high-SNR regime, while suffer from noise amplification introduced by zero-forcing at low SNR. In fact, in the low-SNR regime the system is noise-limited as opposed to interference-limited, so that eliminating interference has little impact on improving the performance.

Nevertheless, as the SNR increases, so does the gain of the proposed scheme over the other two schemes. This can be further observed in Fig. 3. When the SNR of S-D1 is \(-5 \) dB, the gain of the proposed scheme over direct link is as large
as around 40%, while it is more modest with respect to the TDMA scheme, around 17%. In both cases, the gain is reduced (especially with respect to the direct link) when the condition of the S-D1 link improves to 5 dB so that at low SNR the proposed scheme becomes useless with respect to the other schemes. Indeed, the relay is of any help only when D1 does not have a good direct channel with the source.

Finally, it can also be seen in Fig. 4 that as the number of antennas grows, the proposed scheme is able to better exploit the additional antennas to achieve a higher rate. The uneven shape of the curves is explained by the fact that when $M$ is odd, one additional SS is allocated to the signals transmitted to D2 over those to D1. This increases the achievable rate in a more pronounced way than when the number of SSs to D1 is increased by one, since the SNR of the S-D2 link is higher than that of the S-D1 link.

VI. CONCLUSION
In this paper, we have studied a specific scenario of the RBC that models the downlink of a single cell in a cellular system with relay. We have formulated the achievable sum rate of the system as an optimization problem, and designed a low-complexity, non-iterative, linear precoding and decoding beamforming scheme based on BD and IA that, although suboptimal in general, achieves the highest rate possible by completely removing the interference and efficiently allocating the available resources, namely, the transmit power, the duration of each time slot, and the total number of SSs. The numerical results confirm the superiority of our system over other schemes in terms of sum rate and achievable DoF, of as much as 40% in certain cases. As future work, we plan to extend our scheme to a more realistic scenario where transmitters have more antennas than receivers, and evaluate the rate of the proposed scheme when the SS allocation is optimized as mentioned in Section IV-C.

REFERENCES