Coordinated Direct and Relay Transmission with Linear Non-Regenerative Relay Beamforming

Fan Sun, Elisabeth de Carvalho, Petar Popovski and Chan Dai Truyen Thai

Abstract—Joint processing of multiple communication flows in wireless systems has given rise to a number of novel transmission techniques, notably the two-way relaying, but also more general traffic scenarios, such as coordinated direct and relay (CDR) transmissions. In a CDR scheme the relay has a central role in managing the interference and boosting the overall system performance. In this letter we consider the case in which an amplify-and-forward relay has multiple antennas and can use beamforming to support the coordinated transmissions. We focus on one representative traffic type with one uplink user and one downlink user. Two different criteria for relay beamforming are analyzed: maximal weighted sum-rate and maximization of the worst-case weighted SNR. We propose iterative optimal solutions, as well as low-complexity near-optimal solutions.

Index Terms—Analog network coding, relay beamforming.

I. INTRODUCTION

Joint processing of multiple flows is beneficial and has recently been a focus of extensive research, most notably in the area of two-way relaying (TWR), where throughput gains have been demonstrated by utilizing the ideas of wireless network coding. Various transmission designs have been proposed for multi-antenna relay beamforming in TWR systems [1] and the transceivers in multi-user TWR systems [2].

Leveraging on principles from wireless network coding, we have proposed coordinated direct/relay (CDR) transmission schemes with amplify-and-forward (AF) relaying in [3]. The CDR transmission considers wireless cellular scenarios with one direct user (UE) which is directly served by the base station (BS) and one relayed UE which requires a relay to facilitate the communication. The value of CDR lies in the fact that it extends the gains of the analog network coding introduced in the TWR scenarios: while the TWR scenarios feature a very specific two-way pattern, CDR shows how similar gain can be introduced for other traffic patterns.

The relayed UE is assumed to have no direct link to the BS due to large path loss and relies only on the amplified/forwarded signal from the relay. Each user might have a downlink or uplink traffic and thus there are different traffic types. We focus on one representative type with one relayed uplink UE and one direct downlink UE. In this case, we are using the principles of analog network coding for traffic flows that are not related to the TWR traffic in an obvious way. Furthermore it showcases the principle of overheard information where a node overhears a signal that is not intended to itself and uses it as side information to cancel interference in a previous transmission phase.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with the Department of Electronic Systems, Aalborg University, Denmark (e-mail: [fs, edc, petarp, ttc]@es.aau.dk). This work was supported by the Danish Research Council for Technology and Production, Grant 09-065920 and 09-065035.

Notation: ⊗ refers to the Kronecker product and \( ||\cdot||_F^2 \) denotes the Frobenius norm. \( I \) is the identity matrix.
II. SYSTEM OVERVIEW

We consider a relay with $M$ antennas and the BS and the UEs are equipped with one antenna. The transmission from the relayed UE to the relay has the same duration as the transmission from the relay to the BS. There are two information flows: the relayed UE (UE 1) delivers $x_1$ to the BS and the BS transmits $x_2$ to the direct UE (UE 2).

The direct UE wants to estimate the desired signal $x_2$ while $x_1$ is the interference from the other information flow. We cancel the interference linearly using a zero-forcing (ZF) receiver at UE 2 for analytical convenience and $\text{SNR}_2 = \frac{P_{\text{br}} h_{2R}^T W_{\text{br}} h_{R1} + n_{R}}{P_2 h_{2R}^T W_{h_{R1}}^2 + h_{2R}^T W_{\text{br}} h_{R1}}$.

Meanwhile, the direct UE uses $y_2[1]$ from the first slot and $y_2[2]$ from the second slot to form a virtual 2-antenna received signal vector $y_2 = [y_2[1]  y_2[2]]^T$.

A. SNR Expressions

From the network coding principle, $x_2$ is known at the BS and the related interference is mitigated via self-interference cancellation. Therefore, there is no interference when the BS wants to decode $x_1$. Notice we then use linear receivers in the CDR system to decode the desired signals. For the direct UE, after self-interference cancellation, we will have $y_2[2] = \sqrt{P}h_{2B} W_{\text{br}} h_{2R} x_1 + n_{R} + n_{B}$. Then the SNR at the BS to decode $x_1$ is expressed as $\text{SNR}_1 = \frac{P_{\text{br}} h_{2B}^T W h_{2R} h_{R1} + h_{2R}^T W_{\text{br}} h_{R1}}{h_{2R}^T W_{\text{br}} h_{R1}^2 + 1}$.

With the network-coded signal and forward the beamformed signal to the relay, the two traffic flows through two different optimization criteria: (1) the relayed UE and the transmission to the direct UE, can be balanced the two information flows, i.e. the transmission from both the BS and the direct UE. The central role of the relay to network-coded signal and forward the beamformed signal to the receiver at UE 2 for analytical convenience and $\text{SNR}_2 = \frac{P_{\text{br}} h_{2R}^T h_{R1} h_{2B} W_{\text{br}} h_{R1}^2 + h_{2B}^T h_{2R} W_{\text{br}} h_{R1}^2}{h_{2R}^T h_{2B} W h_{R1} h_{2R} + h_{2R}^T W_{\text{br}} h_{R1} + 1}$. Comparing CDR to the TWR in [1], the main difference is that in TWR both flows have to traverse the relay, while in this CDR scenario UE 2 has a direct-link signal component. Moreover, using ZF at UE 2 results in a different SNR form compared to TWR.

To rewrite the expressions in a simple way, the beamforming matrix $W$ is converted into a vector form using the vectorization operation, $w = \text{vec}(W)$ where $w$ is a $M^2 \times 1$ complex vector. With $\text{vec}(M) = (N^T \otimes M) \text{vec}(W)$, we have:

\[
\text{SNR}_1(w) = \frac{w^H G h_{2R}^T w + 1}{w^H f^2 w}
\]

\[
\text{SNR}_2(w) = \frac{w^H (C h_{R1}^T + b b^H) w + |h_{2R}|^2}{w^H f^2 w}
\]

with $e = \sqrt{P} h_{R1}^T h_{2B}^T h_{2R}^T$, $G = I \otimes h_{BR}$, $b = h_{R1}^T h_{2B}^T$, $C = h_{2B}^T h_{R1}$, $f^2 = \sqrt{P} h_{2B}^T h_{2R} + h_{2B}^T h_{R1} h_{2R}$). The relay power is written as $w^H D w + \text{E}^H \text{E} + 1$ with $D = \sqrt{P} (h_{R1}^T \otimes I)$, $\text{E} = \sqrt{P} (h_{R1}^T)$.

III. OPTIMIZING RELAY BEAMFORMING

The relay has the capability to beamform the received network-coded signal and forward the beamformed signal to both the BS and the direct UE. The central role of the relay to balance the two information flows, i.e. the transmission from the relayed UE and the transmission to the direct UE, can be observed in Fig. 2. We consider two ways for balancing the two traffic flows through two different optimization criteria: (1) weighted sum-rate maximization and (2) worst-case weighted SNR maximization.

A. Weighted Achievable Sum-Rate Maximization

The WSRS maximization problem can be formulated as

\[
\begin{align*}
\text{arg max}_w & \quad \mu_1 C_1(w) + \mu_2 C_2(w) \\
\text{s. t.} & \quad \text{w}^H D \text{w} + \text{E}^H \text{E} + 1 \leq P_R
\end{align*}
\]

where $C_1(w)$ and $C_2(w)$ denote the maximum achievable rates for the transmission of $x_1$ and $x_2$, respectively. And $\mu_1$ and $\mu_2$ are the corresponding rate rewards for $C_1$ and $C_2$. The rate expression for each information flow can be written as $R_1 \leq C_1(w) = \frac{1}{2} \log_2 [1 + \text{SNR}_1(w)]$ and $R_2 \leq C_2(w) = \frac{1}{2} \log_2 [1 + \text{SNR}_2(w)]$ where the factor $\frac{1}{2}$ comes from the 2 equal-duration phase transmissions. Using the monotonicity of the log function, the objective can be rewritten as $\text{arg max}_w \{ [1 + \text{SNR}_1(w)]^{\mu_1} [1 + \text{SNR}_2(w)]^{\mu_2} \}$. Because of the shapes of $\text{SNR}_1(w)$ and $\text{SNR}_2(w)$, the weighted
sum-rate maximization problem is a non-convex and NP-hard problem, where the global optimum solution is difficult to obtain within a reasonable computation time. This optimization problem has no general closed form solution.

Although the WSR problem in CDR differs from TWR due to the shape of SNR\(_2(w)\), we also resort to an alternative based on rate-profile similar to [1]. The rate-profile vector, defined as \(a = [\alpha_1 \alpha_2]^T\), regulates the ratio between \(R_1, R_2\) and the sum-rate \(R_{\text{sum}} = R_1 + R_2\) via \(\alpha_1 = \frac{R_1}{R_{\text{sum}}}\) and \(\alpha_2 = \frac{R_2}{R_{\text{sum}}}\). Each feasible \(a\) corresponds to a boundary point in the achievable rate region [1]. Given \(a\), we find the boundary point via (3), maximizing the sum-rate subject to the relay power and rate-profile constraints. Therefore, (2) can be solved by first discarding \(\mu_i (i = 1, 2)\) and determining all the boundary rate-pairs obtained from (3) with respect to all different \(a\) values and then select the pair with highest WSR.

\[
R(P_R, a) = \begin{cases} \max_w R_{\text{sum}} \\ \text{s. t.} \quad C_i(w) \geq \alpha_i R_{\text{sum}} (i = 1, 2) \\ w^H [D^H D + E^H E + ] w \leq P_R \end{cases}
\]

(3)

In Algorithm 1, we use a one-dimensional bisection search to efficiently obtain the optimal sum-rate for the NP-hard problem (3). Note that \(R_{\text{UB}}\) is an upper bound on the sum-rate value (see Section III-C) and \(\epsilon_R\) is a small positive constant to control the precision of the algorithm. At each bisection step, we need to verify whether it is feasible for the relay to support the rate-pair \(aR_0\), meaning whether the relay transmit power is sufficient. To do so, we determine the minimal transmit power at the relay subjecting to QoS constraints (SNR\(_i(w) \geq r_i = 2^{2\alpha_i R_0} - 1\)) in (4). A convergence proof of Algorithm 1 can be built based on Appendix D in [1].

**Algorithm 1** Sum-rate max relay beamforming (given \(a\))

For given \(a\), set \(R_{\text{min}} = 0\) and \(R_{\text{max}} = R_{\text{UB}}\)

repeat until \(R_{\text{max}} - R_{\text{min}} \leq \epsilon_R\)

I. set \(R_0 = \frac{1}{2} (R_{\text{min}} + R_{\text{max}})\)

II. solve (4) and obtain \(P_R\)

\(\text{if} \) (4) is feasible and \(P_R \leq P_R, \) \(R_{\text{min}} = R_0 \)

\(\text{else} \) \(R_{\text{max}} = R_0\)

Now we explain how to solve the relay power minimization:

\[
P_{\text{SR}}(R_0, a) = \min_w \begin{cases} P_R = w^H [D^H D + E^H E + ] w \\ \text{s. t.} \quad \frac{w^H e e^H w}{w^H f f^H w} \geq r_1 \\ \frac{w^H c c^H w + h_{21}^H + h_{21} h_{21}^H w}{w^H b b^H w} \geq r_2 \end{cases}
\]

(4)

which is a non-convex quadratically constrained quadratic programming (QCQP) problem and hence NP-hard. Introducing \(X = \text{ww}^H\), we can solve (4) by resorting to the widely used semidefinite programming (SDP) with rank relaxation by discarding the non-convex constraint Rank\((X) = 1\) [4], [5]:

\[
\min_X P_R = \text{Tr} \left\{ [D^H D + E^H E + I] X \right\} \\
\text{s. t.} \quad \text{Tr} \left\{ [e e^H - r_1 G^H G] X \right\} \geq 1, \quad X \succeq 0 \\
\quad \text{Tr} \left\{ [[f f^H - r_2 C^H C + b b^H] X] \right\} \geq |h_{21}|^2 r_2,
\]

where \(X \succeq 0\) means that \(X\) is positive semidefinite. Notice that the optimal \(X\) could have Rank\((X) > 1\), where the randomization technique [5] can be applied for near global optimal solution. However, for this special structure based on

**Theorem 3.2** in [1], we can efficiently construct an optimal rank-one solution from the optimal \(X\). Therefore, the exact optimal solution for (4) can be efficiently obtained.

**B. Worst-Case Weighted SNR Maximization**

Denote \(\beta_i > 0 (i = 1, 2)\) to be the priority factor associated with SNR\(_i\) \((i = 1, 2)\). The max-min problem is written as

\[
S(P_R) = \begin{cases} \max_w \min_i \text{SNR}(w) \\ \text{s. t.} \quad w^H [D^H D + E^H E + I] w \leq P_R \end{cases}
\]

which maximizes the worst-case weighted SNR. In point-to-point MIMO systems, an algorithm to optimize the minimal SNR based on the connection with power minimization is proposed in [6]. Denote \(\gamma_0\) to be the target weighted SNR value, we extend the connection to the CDR scenario via

\[
P_{\text{SNR}}(\gamma_0) = \begin{cases} \min_w \frac{P_R}{w^H [D^H D + E^H E + I] w} \\ \text{s. t.} \quad \frac{1}{2} \beta_1 \frac{w^H G^H G w}{w^H {\tilde{f}} f^H w} \geq \gamma_0 \\
\quad \frac{1}{2} \beta_2 \frac{w^H \epsilon \epsilon^H w}{w^H b b^H w} \geq \gamma_0 \end{cases}
\]

(5)

(6)

The constraints in (6) indicate that we target the same weighted SNR requirements for both flows while the constraints in (4) focus on different SNR requirements. We introduce the connection between (5) and (6) in Proposition 1 which can be proved by contradiction (similar to Theorem 3 in [6]).

**Proposition 1.** The relay power minimization problem (6) and the worst-case weighted SNR problem (5) are inverse problems. The optimal value for each problem is continuous and strictly monotonic increasing.

To solve (5), the strict monotonicity and continuity guarantees a one-dimensional bisection search applied to (6) will obtain the optimal worst-case weighted SNR value. The algorithm is shown in Algorithm 2. Note that WSNR\(_{\text{min}}\) and WSNR\(_{\text{max}}\) define the range of weighted SNR values with detailed derivations in Section III-C. \(\epsilon_r\) is a small positive constant. The convergence proof of Algorithm 2 can be built by contradiction and letting \(\epsilon_r \to 0\), similar to Appendix D in [1]. To solve (5), we apply the same relaxed SDP as for (4).

**Algorithm 2** Minimal weighted SNR max relay beamforming

set \(r_{\text{min}} = \text{WSNR}_{\text{min}}\) and \(\text{r}_{\text{max}} = \text{WSNR}_{\text{max}}\)

repeat until \(r_{\text{max}} - r_{\text{min}} \leq \epsilon_r\)

I. set \(r_0 = \frac{1}{2} (r_{\text{min}} + r_{\text{max}})\)

II. solve (6) and obtain \(P_R\)

\(\text{if} \) (6) is feasible and \(P_R \leq P_R, \) \(r_{\text{min}} = r_0 \)

\(\text{else} \) \(r_{\text{max}} = r_0\)

**C. Sum-Rate and Weighted SNR Ranges**

The initial values for the two optimization algorithms are now discussed. Let us first consider two benchmarks: the relay beamforming designs targeting either SNR\(_1(w)\) or SNR\(_2(w)\) maximization. It can be easily proved that an increase of the relay power will increase both SNR values. Therefore SNR\(_1(w)\) and SNR\(_2(w)\) are optimized when the relay transmits at full power. We further introduce \(J^H J \triangleq D^H D + E^H E + I\) and \(\tilde{w} = J w\) to transform the relay power constraint into \(w^H \tilde{w} = P_R\). Then, SNR\(_1(w)\) and SNR\(_2(w)\) are expressed as generalized Rayleigh quotients

\[
g_1(\tilde{w}) = \frac{\tilde{w}^H J^H J \tilde{w} - 1}{\tilde{w}^H J^H J \tilde{w} + \frac{1}{\gamma_0}}
\]

\[
\tilde{w}^H J^H G G J^H + \frac{1}{\gamma_0} \tilde{w}
\]
\[ g_2(\tilde{w}) = \frac{\tilde{w}^H J^{-H} f f^H J^{-1} \tilde{w}}{\tilde{w}^H \left[ J^{-H} (C^H C + b b^H) J^{-1} + \frac{b_{21}^2 + b_{22}^2}{b_{21}^2} \right] \tilde{w}}. \]

The two vectors \( \tilde{w}_1 \) and \( \tilde{w}_2 \) maximizing \( g_1 \) and \( g_2 \) separately can be easily obtained. Thus, \( R_{UB} = R_1(\tilde{w}_1) + R_2(\tilde{w}_2) \). From [3], a suitable SNR range are \( g_1(\tilde{w}_2) \leq \text{SNR}_1 \leq g_1(\tilde{w}_1), \quad g_2(\tilde{w}_1) \leq \text{SNR}_2 \leq g_2(\tilde{w}_2) \) because one flow is optimized when we are sacrificing the other flow. For the weighted SNR problem, \( \text{WSNR}_{\text{min}} = \min \left\{ \frac{g_1(\tilde{w}_2)}{\beta_1}, \frac{g_2(\tilde{w}_1)}{\beta_2} \right\} \)

and \( \text{WSNR}_{\text{max}} = \max \left\{ \frac{g_1(\tilde{w}_1)}{\beta_1}, \frac{g_2(\tilde{w}_2)}{\beta_2} \right\} \).

Meanwhile, from the bisection principle, the number of iterations required for Algorithm 1 or Algorithm 2 can be obtained directly from the initial values and the threshold for the stopping criteria, i.e. \( \epsilon_R \) or \( \epsilon_r \).

D. Low-Complexity Design

The optimal solutions involve SDP optimization and optimal rank-one construction via linear programming (LP) [1] inside each bisection search step. We also propose low-complexity suboptimal designs for both problems, inspired by [7]. The design is termed to be the linear space spanning (LSS), where the solution is chosen to lie in the linear space spanned by \( \tilde{w}_1 \) and \( \tilde{w}_2 \), \( \tilde{w}_L = a \tilde{w}_1 + b \tilde{w}_2 \) (\( a \) and \( b \) are real value parameters).

It is obvious that any scaling of \( a \) does not change the values for \( g_1(\tilde{w}_L) \) and \( g_2(\tilde{w}_L) \). Therefore, \( \tilde{w}_L \) is further simplified by letting \( a = 1 \) and \( \tilde{w}_L = \tilde{w}_1 + b \tilde{w}_2 \). Therefore, the suboptimal designs will perform maximization of the WSR objective and the weighted SNR objective over \( b_{\text{WSR}} \) and \( b_{\text{SNR}} \) without other constraints, respectively. For either problem, a grid search using the Nelder-Mead method [7] is applied to efficiently find the solution. Then the obtained beamforming solution is scaled to satisfy the relay power constraint, which does not change the WSR and weighted SNR values.

The optimization in the suboptimal design is over one real variable. Thus the complexity reduction is huge compared to the optimal solutions. However, this advantage is obtained at the expense of a performance loss shown in Section IV.

IV. NUMERICAL RESULTS

In this section, we present simulation results with \( M = 2 \) antennas at the relay and all transmission nodes having the same transmit power, i.e. \( P_R = P \). The thresholds for the two algorithms are \( \epsilon_R = \epsilon_r = 0.01 \). The two benchmarks targeting individual SNR maximization are included. Fig. 3 shows the sum-rate with \( \mu_1 = \mu_2 = 1 \). The proposed optimal and suboptimal solutions provide non-trivial gains compared to the two benchmarks. LSS gives performance close to the optimal solution. The conventional two orthogonal uplink and downlink transmission with three equal slots is constrained to use the same total duration as CDR and the multi-antenna relay helps only the flow from UE 1; it has a huge sum-rate loss compared to the optimal CDR transmission. We also show the worst-case SNR performance with \( \beta_1 = \beta_2 = 1 \) in Fig. 4 where LSS is observed to perform close to the optimal solution. The conventional transmission is not suitable to be included for SNR comparison.

REFERENCES


