LINEAR TRANSCEIVER DESIGN FOR RELAY-ASSISTED BROADCAST SYSTEMS WITH DIAGONAL SCALING

Chao Zhao and Benoit Champagne

Dept. Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, Quebec, Canada, H3A 0C6
Email: chao.zhao@mail.mcgill.ca, benoit.champagne@mcgill.ca

ABSTRACT

In this paper, we study the linear transceiver design for the downlink of a cellular network assisted by a multi-antenna relay. A diagonal scaling scheme is proposed in which multiple single-antenna users apply different complex-valued scaling to their signals before decoding, as represented by an equivalent *diagonal* equalizer matrix. This equalizer is designed together with a linear precoder at the base station (BS) and a linear processing matrix at the relay. The objective is to minimize the weighted minimum mean square error (MMSE) between the precoder input and the equalizer output, subject to power constraints at the BS and the relay. In particular, the optimal relaying matrix is first derived in closed form as a function of the precoder and the equalizer. The latter two can then be jointly designed in an efficient iterative manner. Simulation results demonstrate lower bit-error rates (BERs) than previous design methods.

Index Terms— Broadcast channel, wireless relaying, transceiver design, weighted MMSE, diagonal scaling

1. INTRODUCTION AND RELATION TO PRIOR WORK

The use of relays to assist communications within cellular networks has attracted interests in both academic and industrial worlds [1]. For relay stations equipped with multiple antennas, the simplest and perhaps also the most practical relaying strategy is half-duplex amplify-and-forward (AF). In this scheme, the source-to-relay and relay-to-destination transmissions occur in orthogonal degrees of freedoms (time slots or frequency bins), and the relay applies linear processing to its received signals before retransmitting them. For single-user one-source–one-relay–one-destination (1S-1R-1D) systems, the optimal relay processing takes the form of singular value decomposition (SVD) under a wide variety of criteria [2]. This can be extended to the multiple access channels (MAC) without difficulty [3].

This SVD-based relaying framework, however, cannot be readily generalized to relay-assisted broadcast channels (BC), i.e., downlink transmissions of cellular systems. This is because the mobile users are not collocated and hence cannot jointly process their received signals. In general, the problems of transceiver design for relay-assisted BC fall into two categories: minimizing the weighted sum power of the base station (BS) and the relay subject to quality of service (QoS) constraints [4], or optimizing a selected performance criterion such as the sum rate or mean square error (MSE) subject to power constraints [5–11]. The maximum sum-rate design was studied in [5] with special structures, such as zero forcing (ZF), dirty paper coding (DPC) and QR transceiver, and in [8] using quadratic programming. The minimum mean square error (MMSE) design was considered in [6] with single-antenna users, and in [7,9,11] with multi-antenna users. The latter case usually requires

This work was funded by a CRD grand from NSERC of Canada with industrial support from InterDigital, Inc.

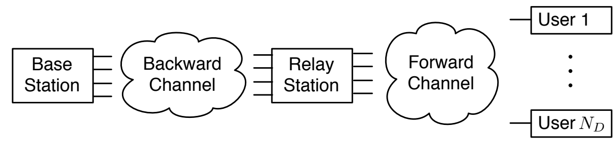


Fig. 1. A relay-assisted broadcast channel (downlink) with single-antenna users.

complicated algorithms that iterate through the precoder, the relaying matrix and every equalizers multiple times. Consequently, treating multi-antenna users as multiple single-antenna users, although suboptimal, simplifies the transceiver design significantly.

With this in mind, we study in this paper the relay-assisted BC with single-antenna users as shown in Fig. 1. In particular, we consider the joint design of a linear precoder at the BS, a linear processing matrix at the relay, and an equalizer for the destination users. The equalizer matrix employs a diagonal scaling scheme which provides more flexibility by allowing different users to apply their own amplitude scaling and phase rotation before decoding, in contrast to [6] which assumes the same scaling for these users. The objective here is to minimize the weighted MSE between the equalizer output and the source input, subject to power constraints at the BS and the relay. Specifically, we first derive the optimal relaying matrix as a function of the precoder and the equalizer. As a consequence, the latter two can then be jointly designed in an iterative manner. As a special case, our approach provides a closed form for the optimal solution when the users apply the same scaling, whereas only an iterative approach was used in [6]. Furthermore, simulation results demonstrate lower bit-error rate (BER) with the proposed diagonal scaling scheme than previous methods such as [6].

In Sec. 2, we present the system model and formulate the optimization problem, which is then solved step by step in Sec. 3. The numerical results are shown in Sec. 4, followed by conclusion in Sec. 5. Notations: superscripts *, ^{T} and ^{H} denote conjugate, transpose and Hermitian transpose, respectively; $\|\cdot\|$ stands for the Euclidean norm; $\mathrm{E}(\cdot)$ takes expectation; $\mathbb{C}$ denotes the complex field.

2. SYSTEM MODEL AND PROBLEM FORMULATION

In the downlink of a cellular system as shown in Fig. 1, the multi-antenna BS is sending multiple symbol streams simultaneously to their intended single-antenna users, through the aid of a multi-antenna relay. This system operates in a two-hop half-duplex mode: in the first hop, the BS transmits signals to the relay through the backward channel; in the second hop, the relay forwards these signals to the users via the forward channel. The relay applies a linear transformation to its received signals before retransmitting. The direct links between the BS and the users are neglected due to high levels of attenuation. The numbers of antennas at the BS and the relay are respectively N_S and N_R. The number of users is N_D.

We consider a discrete-time complex baseband-equivalent signal model as shown in Fig. 2. The input symbol vector $\mathbf{b} \in \mathbb{C}^{N_D \times 1}$,

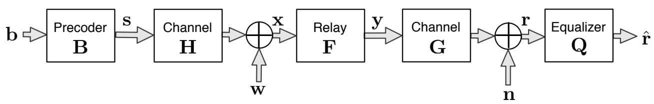


Fig. 2. The system and signal model

with zero mean and covariance $\mathbf{R}_b = \mathbf{I}_{N_D}$, consists of N_D statistically independent symbols to be transmitted to the corresponding users. At the BS, this vector is preprocessed by a linear *precoder* matrix $\mathbf{B} \in \mathbb{C}^{N_S \times N_D}$ to generate the transmitted signal vector

$$\mathbf{s} = \mathbf{B}\mathbf{b}. \tag{1}$$

The backward channel between the BS and the relay is represented by matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_S}$. The signal vector $\mathbf{x} \in \mathbb{C}^{N_R \times 1}$ received at the relay is therefore

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w}, \tag{2}$$

where $\mathbf{w} \in \mathbb{C}^{N_R \times 1}$ is an additive, zero-mean, circularly symmetric complex Gaussian noise with covariance $\mathbf{R}_w$.

In this baseband-equivalent model, the linear processing at the relay is represented by a matrix $\mathbf{F} \in \mathbb{C}^{N_R \times N_R}$. That is, the relay retransmits its received noisy signal $\mathbf{x}$ as in

$$\mathbf{y} = \mathbf{F}\mathbf{x}. \tag{3}$$

The signal received by the kth single-antenna user is

$$r_k = \mathbf{g}_k^T \mathbf{y} + n_k = \mathbf{g}_k^T \mathbf{F}\mathbf{H}\mathbf{B}\mathbf{b} + \mathbf{g}_k^T \mathbf{F}\mathbf{w} + n_k, \tag{4}$$

in which $\mathbf{g}_k \in \mathbb{C}^{N_R \times 1}$ denotes the channel vector from the relay to the kth user and the scalar n_k is the receiver noise. These received signals can be stacked into a single vector

$$\mathbf{r} \triangleq [r_1, \ldots, r_{N_D}]^T = \mathbf{G}\mathbf{F}\mathbf{H}\mathbf{s} + \mathbf{G}\mathbf{F}\mathbf{w} + \mathbf{n}, \tag{5}$$

where $\mathbf{G} \triangleq [\mathbf{g}_1, \ldots, \mathbf{g}_{N_D}]^T$ is the compounded *forward* channel matrix from the relay to the destination users. The noise term, $\mathbf{n} \triangleq [n_1, \ldots, n_{N_D}]^T$, is independent from $\mathbf{b}$ and $\mathbf{w}$, and modeled as a circularly symmetric complex Gaussian random vector with zero mean and covariance $\mathbf{R}_n$.

Considering the possible difference in pathloss experienced by different users, we propose a diagonal scaling scheme in which each destination user scales its signal r_k by its *own* complex factor q_k before decoding, as represented by a *diagonal* compound equalizer matrix $\mathbf{Q} \triangleq \mathrm{diag}(q_1, \ldots, q_{N_D})$. This is in contrast with previous works such as [6] that considered the same scaling for different users. The output of the equalizer $\mathbf{Q}$ is

$$\hat{\mathbf{r}} = \mathbf{Q}\mathbf{r} = \mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{H}\mathbf{B}\mathbf{b} + \mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{w} + \mathbf{Q}\mathbf{n}. \tag{6}$$

We consider the problem of optimizing the relaying matrix $\mathbf{F}$, together with the source precoder $\mathbf{B}$ and the diagonal equalizer $\mathbf{Q}$. The optimality criterion is chosen as the weighted MSE between the precoder input and the equalizer output

$$\mathrm{MSE}(\mathbf{F}, \mathbf{B}, \mathbf{Q}) \triangleq \mathrm{E}\big\{(\hat{\mathbf{r}} - \mathbf{b})^H \mathbf{W} (\hat{\mathbf{r}} - \mathbf{b})\big\}, \tag{7}$$

in which the diagonal weight matrix $\mathbf{W}$ provides different priority to different users. Our purpose is to minimize the above weighted MSE subject to the following two power constraints simultaneously. One is the expected transmit power of the BS

$$\mathrm{E}\big\{\|\mathbf{s}\|_2^2\big\} = \mathrm{tr}(\mathbf{R}_s) = \mathrm{tr}(\mathbf{B}\mathbf{B}^H) \leq P_S \tag{8}$$

and the other is the expected transmit power of the relay

$$\mathrm{E}\big\{\|\mathbf{y}\|_2^2\big\} = \mathrm{tr}(\mathbf{R}_y) = \mathrm{tr}(\mathbf{F}\mathbf{R}_x\mathbf{F}^H) \leq P_R, \tag{9}$$

where $\mathbf{R}_x = \mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_w$. We assume the availability of perfect channel state information (CSI), i.e., the channel matrices and the covariance matrices are known.

3. OPTIMAL TRANSCEIVER DESIGN

We take a step-by-step approach to design the optimal combination of the relaying matrix $\mathbf{F}$, the precoder $\mathbf{B}$ and the diagonal equalizer $\mathbf{Q}$. The first step is to derive the optimal $\mathbf{F}_o$ as a closed-form function of $\mathbf{B}$ and $\mathbf{Q}$, thereby removing the constraint (9) from the problem. After substituting this $\mathbf{F}_o$ into the objective function, the problem is reduced to the joint design of $\mathbf{B}$ and $\mathbf{Q}$ subject to the constraint in (8). As it will be shown in Sec. 3.2 and 3.3, it is straightforward to design either one of $\mathbf{B}$ and $\mathbf{Q}$ while treating the other as a constant matrix. The joint design will be discussed in Sec. 3.4.

3.1. Optimal Design of the Relaying Matrix

We start by designing the relaying matrix $\mathbf{F}$ as a function of the precoder $\mathbf{B}$ and the equalizer $\mathbf{Q}$. The constraint in (8) does not depend on $\mathbf{F}$ and henceforth need not to be considered for now. Based on the Lagrangian duality and the Karush-Kuhn-Tucker (KKT) conditions [12], the optimal relaying matrix can be derived as

$$\mathbf{F}_o = (\mathbf{G}^H\mathbf{Q}^H\mathbf{W}\mathbf{Q}\mathbf{G} + \lambda^\star\mathbf{I})^{-1}\mathbf{G}^H\mathbf{Q}^H\mathbf{W}\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}. \tag{10}$$

However, the duality parameter $\lambda^\star$ has no closed form and must be obtained by solving a nonlinear equation. Our approach to circumvent this difficulty is to introduce a linear scaling $\eta > 0$ in the equalizer. If we replace any given $\mathbf{Q}$ with a scaled version $\eta^{-1}\mathbf{Q}$, the duality parameter $\lambda^\star$, the optimal relaying matrix $\mathbf{F}_o$ and the corresponding minimum MSE are all functions of η. It turns out that for the optimal η_o leading to the smallest MSE, these quantities can all be expressed in closed forms, as presented in the following theorem:

Theorem 1 *For any equalizer* $\mathbf{Q}$*, there exists a scaled version* $\eta_o^{-1}\mathbf{Q}$ *so that:*

(a) The optimal relaying matrix is in the closed form

$$\mathbf{F}_o = \eta_o(\mathbf{G}^H\mathbf{Q}^H\mathbf{W}\mathbf{Q}\mathbf{G} + \theta\mathbf{I})^{-1}\mathbf{G}^H\mathbf{Q}^H\mathbf{W}\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}, \tag{11}$$

where $\theta \triangleq \mathrm{tr}(\mathbf{W}\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)/P_R$ *and* $\eta_o > 0$ *satisfies* $\mathrm{tr}(\mathbf{F}_o\mathbf{R}_x\mathbf{F}_o^H) = P_R$.

(b) The minimum weighted MSE with $\mathbf{F} = \mathbf{F}_o$ *takes the form*

$$\mathrm{MSE}(\mathbf{Q}, \mathbf{B}) = \mathrm{tr}(\mathbf{W}) - \mathrm{tr}\big(\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}\mathbf{H}\mathbf{B} \; \mathbf{W}\mathbf{Q}\mathbf{G}\,(\mathbf{G}^H\mathbf{Q}^H\mathbf{W}\mathbf{Q}\mathbf{G} + \theta\mathbf{I})^{-1}\mathbf{G}^H\mathbf{Q}^H\mathbf{W}^H\big). \tag{12}$$

(c) Any other choice of $\eta^{-1}\mathbf{Q}$ *together with the corresponding* $\mathbf{F}_o$ *in* (10) *would lead to an MSE no smaller than* (12)*. In addition,* (11) *and* (12) *are invariant to linear scaling of* $\mathbf{Q}$.

Proof The proof is omitted due to space limitations.

In the sequel, we suppose that the optimal $\mathbf{F}_o$ and η_o in Theorem 1 are implicitly chosen. Hence, the problem now becomes that of designing $\mathbf{B}$ and $\mathbf{Q}$ which can minimize (12) subject to (8).

3.2. Optimal Design of the Precoder B for Fixed Equalizer Q

In this subsection, we assume the equalizer $\mathbf{Q}$ is fixed and design the optimal precoder $\mathbf{B}$. Define the eigenvalue decomposition (EVD)

$$\mathbf{W}\mathbf{Q}\mathbf{G}\,\big(\mathbf{G}^H\mathbf{Q}^H\mathbf{W}\mathbf{Q}\mathbf{G} + \theta\mathbf{I}\big)^{-1}\mathbf{G}^H\mathbf{Q}^H\mathbf{W}^H = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H, \tag{13}$$

where $\mathbf{U}$ is unitary and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_{N_D})$ with its diagonal entries sorted in non-increasing order. Similarly, define the EVD

$$\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H} = \mathbf{V}\boldsymbol{\Sigma}_{\mathbf{H}}\mathbf{V}^H, \tag{14}$$

where $\mathbf{V}$ is unitary and $\mathbf{\Sigma_H} = \mathrm{diag}(a_1, \cdots, a_{N_S})$. Based on

$$\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}\mathbf{H}\mathbf{B} = \mathbf{I} - \left(\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\mathbf{B}\right)^{-1}, \tag{15}$$

the problem of minimizing (12) subject to the power constraint (8) is equivalent to that of minimizing

$$f(\mathbf{B}) \triangleq \mathrm{tr}\left(\mathbf{\Sigma}(\mathbf{I} + \mathbf{U}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\mathbf{B}\mathbf{U})^{-1}\right) \tag{16}$$

subject to (8). Next, we show that the optimal precoder $\mathbf{B}_o$ diagonalizes the matrix in (16).

Lemma 2 *For any* $\mathbf{B}$ *satisfying* (8)*, there exists another* $\tilde{\mathbf{B}}$ *so that:*

(a) $\tilde{\mathbf{B}}$ *also satisfies the power constraint* (8)*;*

(b) $\mathbf{U}^H\tilde{\mathbf{B}}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\tilde{\mathbf{B}}\mathbf{U}$ *is diagonal with non-increasing main diagonal entries;*

(c) $f(\tilde{\mathbf{B}}) \leq f(\mathbf{B})$.

Proof Define the EVD $\mathbf{U}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\mathbf{B}\mathbf{U} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{U}}^H$, with the eigenvalues, $\tilde{\sigma}_k$, sorted in the non-increasing order. The objective function is hence a weighted sum of the nonnegative diagonal entries of $\tilde{\mathbf{U}}(\mathbf{I} + \tilde{\mathbf{\Sigma}})^{-1}\tilde{\mathbf{U}}^H$, denoted as x_k:

$$f(\mathbf{B}) = \mathrm{tr}\left(\mathbf{\Sigma}\ \tilde{\mathbf{U}}(\mathbf{I} + \tilde{\mathbf{\Sigma}})^{-1}\tilde{\mathbf{U}}^H\right) = \sum_{k=1}^{N_D} \sigma_k x_k, \tag{17}$$

$$= \sigma_{N_D}\sum_{l=1}^{N_D} x_l + \sum_{k=1}^{N_D-1}\left((\sigma_k - \sigma_{k+1})\sum_{l=1}^{k} x_l\right). \tag{18}$$

Let $\tilde{\mathbf{B}} \triangleq \mathbf{B}\mathbf{U}\tilde{\mathbf{U}}\mathbf{U}^H$. Since both $\mathbf{U}$ and $\tilde{\mathbf{U}}$ are unitary, $\tilde{\mathbf{B}}$ satisfies $\mathrm{tr}(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H) = \mathrm{tr}(\mathbf{B}\mathbf{B}^H) \leq P_S$, and $\mathbf{U}^H\tilde{\mathbf{B}}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\tilde{\mathbf{B}}\mathbf{U} = \tilde{\mathbf{\Sigma}}$ is diagonal. Hence, $f(\tilde{\mathbf{B}})$ can be obtained by replacing x_k in (17) or (18) with $t_k \triangleq (\tilde{\sigma}_k + 1)^{-1}$. For the same matrix $\tilde{\mathbf{U}}(\mathbf{I} + \tilde{\mathbf{\Sigma}})^{-1}\tilde{\mathbf{U}}^H$, t_k is a non-decreasing sequence comprising of its eigenvalues, whereas the sequence x_k includes its diagonal entries. The latter can be sorted into a non-decreasing sequence x_{π_k}. According to [13, Thm. 4.3.26], we have

$$\sum\nolimits_{l=1}^{k} t_l \leq \sum\nolimits_{l=1}^{k} x_{\pi_l} \leq \sum\nolimits_{l=1}^{k} x_l, \tag{19}$$

for any $1 \leq k \leq N_D$ (majorization). As per (18), $f(\tilde{\mathbf{B}}) \leq f(\mathbf{B})$. ■

Up to now, we can assume without loss of generality that $\mathbf{B}$ diagonalizes $\mathbf{U}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\mathbf{B}\mathbf{U}$. Furthermore, for any such $\mathbf{B}$, it can be shown that there exists a $\hat{\mathbf{B}}$ satisfying $\mathbf{U}^H\hat{\mathbf{B}}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\hat{\mathbf{B}}\mathbf{U} = \mathbf{U}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\mathbf{B}\mathbf{U}$ and $\mathrm{tr}(\hat{\mathbf{B}}\hat{\mathbf{B}}^H) \leq \mathrm{tr}(\mathbf{B}\mathbf{B}^H)$. Specifically, $\hat{\mathbf{B}} = \mathbf{V}\mathbf{\Sigma_B}\mathbf{U}^H$, where $\mathbf{\Sigma_B} \triangleq [\mathrm{diag}(b_1, \cdots, b_{N_D}), \mathbf{0}]^T \in \mathbb{C}^{N_D \times N_S}$. The weighted sum MSE in (16) is now reduced to

$$f(\mathbf{B}) = \sum_{k=1}^{N_D} \sigma_k a_k/(1 + a_k b_k^2), \tag{20}$$

subject to $\mathrm{tr}(\mathbf{\Sigma_B}\mathbf{\Sigma_B}^H) = \sum_{k=1}^{N_D} b_k^2 \leq P_S$. This convex problem can easily be solved. We skip the details to present the following theorem directly:

Theorem 3 *For a fixed equalizer* $\mathbf{Q}$*, with the optimal* $\mathbf{F}_o$ *in* (11)*, the optimal precoder* $\mathbf{B}$ *is of the SVD form*

$$\mathbf{B}_o = \mathbf{V}\mathbf{\Sigma_B}\mathbf{U}^H. \tag{21}$$

The diagonal entries b_k *of* $\mathbf{\Sigma_B}$ *satisfy*

$$b_k^2 = \left(\sqrt{\frac{\sigma_k}{a_k}}\frac{1}{\sqrt{\gamma}} - \frac{1}{a_k}\right)^+, \tag{22}$$

where $x^+ \triangleq \max(x, 0)$ *and* $\gamma > 0$ *is the unique solution of* $\sum_{k=1}^{N_D} b_k^2 = P_S$.

3.3. Optimal Design of the Equalizer Q for Fixed Precoder B

We can derive the following equation

$$\begin{aligned}&\mathbf{W}\mathbf{Q}\mathbf{G}\left(\mathbf{G}^H\mathbf{Q}^H\mathbf{W}\mathbf{Q}\mathbf{G} + \theta\mathbf{I}\right)^{-1}\mathbf{G}^H\mathbf{Q}^H\mathbf{W}^H \\ &= \mathbf{W} - (\mathbf{Q}\mathbf{G}\mathbf{G}^H\mathbf{Q}^H/\theta + \mathbf{W}^{-1})^{-1},\end{aligned} \tag{23}$$

so that optimizing $\mathbf{Q}$ for a fixed $\mathbf{B}$ is equivalent to minimizing

$$g(\mathbf{Q}) \triangleq \mathrm{tr}\left((\mathbf{Q}\mathbf{G}\mathbf{G}^H\mathbf{Q}^H/\theta + \mathbf{W}^{-1})^{-1}\ \mathbf{C}\right), \tag{24}$$

where $\mathbf{C} \triangleq \mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}\mathbf{H}\mathbf{B}$. This problem can efficiently be solved using steepest descent. Define $\mathbf{q} \triangleq \mathrm{diag}(\mathbf{Q}) = [q_1, \ldots, q_{N_D}]^T$. The gradient of $g(\mathbf{Q})$ with respect to $\mathbf{q}^*$ is derived as

$$\begin{aligned}\bigtriangledown_{\mathbf{q}^*} g &= [\partial g/\partial q_1^* \ \cdots \ \partial g/\partial q_{N_D}^*]^T \\ &= \mathrm{tr}(\mathbf{Q}^H\mathbf{E}\mathbf{C}\mathbf{E}\mathbf{Q}\mathbf{G}\mathbf{G}^H)\mathrm{diag}(\mathbf{W}\mathbf{Q}\mathbf{R}_n) \\ &\quad - \mathrm{tr}(\mathbf{W}\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)\mathrm{diag}(\mathbf{E}\mathbf{C}\mathbf{E}\mathbf{Q}\mathbf{G}\mathbf{G}^H),\end{aligned} \tag{25}$$

where $\mathbf{E} \triangleq (\mathbf{Q}^H\mathbf{G}\mathbf{G}^H\mathbf{Q}^H + \theta\mathbf{W}^{-1})^{-1}$. The steepest descent method starts from an initial $\mathbf{q}_0$ (or equivalently $\mathbf{Q}_0$) and searches along the opposite direction of the gradient

$$\mathbf{q}_{k+1} = \mathbf{q}_k - \alpha_k \bigtriangledown_{\mathbf{q}^*} g|_{\mathbf{q}=\mathbf{q}_k}, \tag{26}$$

until $\mathbf{q}_k$ converges to a local optima. The step size α_k can be chosen according to Wolfe conditions [14].

3.4. Joint Design

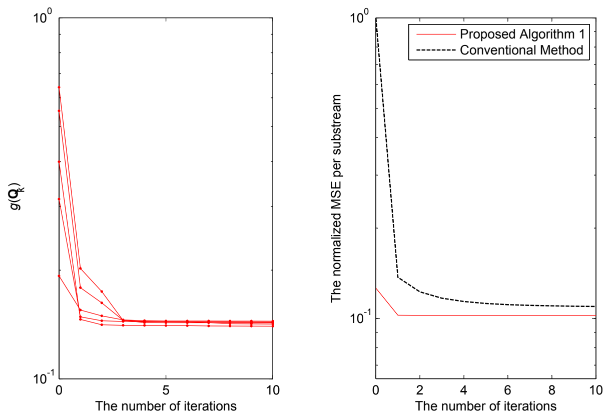
Based upon the theoretical development in Sec. 3.1, 3.2 and 3.3, we now propose an iterative algorithm for the joint design of $\mathbf{B}$, $\mathbf{F}$ and $\mathbf{Q}$ (cf. Sec. 2). The first step is to design $\mathbf{B}$ and $\mathbf{Q}$ that minimize (12) subject to (8). Specifically, our algorithm starts from initial $\mathbf{Q}_0$ and $\mathbf{B}_0$, and updates $\mathbf{B}$ and $\mathbf{Q}$ repeatedly based on the methods proposed in Sec. 3.2 and 3.3. This process generates a non-increasing sequence of MSE values and is therefore guaranteed to converge to a local optima. The second step is to compute the optimal $\mathbf{F}_o$ and η_o as in Theorem 1. These procedures are summarized in Algorithm 1.

Algorithm 1: Joint Design

Initiate $\mathbf{Q}_0, \mathbf{B}_0, m = 0, \epsilon > 0$;
repeat
  Increment the counter $m \leftarrow m + 1$;
  Compute $\mathbf{B}_m$ according to Theorem 3;
  Compute $\mathbf{Q}_m$ using the steepest descent method in Sec. 3.3;
until $|\mathrm{MSE}(\mathbf{Q}_m, \mathbf{B}_m) - \mathrm{MSE}(\mathbf{Q}_{m-1}, \mathbf{B}_{m-1})| \leq \epsilon$;
Compute $\mathbf{F}_o$ and η_o according to Theorem 1;
Let $\mathbf{B}_o \leftarrow \mathbf{B}_m$, $\mathbf{Q}_o \leftarrow \eta_o^{-1}\mathbf{Q}_m$.

Algorithm 1 is different from a conventional iterative approach widely used to solve similar problems [6, 7, 16]. The latter needs to iterate through *all three* matrices $\mathbf{B}$, $\mathbf{F}$ and $\mathbf{Q}$: if $\mathbf{F}$ and $\mathbf{Q}$ are fixed, $\mathbf{B}$ is updated as the optimal precoder for the equivalent multiple-input multiple-output (MIMO) channel [15]; if $\mathbf{B}$ and $\mathbf{Q}$ are held constant, $\mathbf{F}$ is optimized according to (11); if $\mathbf{B}$ and $\mathbf{F}$ are preserved, each diagonal entry of $\mathbf{Q}$ is the optimal MMSE receiver for the scalar channel of the corresponding user. Our numerical results will show that Algorithm 1 converges significantly faster than this approach.

A special case of the joint design is when $\mathbf{Q} = \mathbf{I}_{N_D}$ and $\mathbf{W} = \mathbf{I}_{N_D}$, which is in fact the same problem as that studied in [6]. However, the optimal $\mathbf{B}_o$ and $\mathbf{F}_o$ were obtained using an iterative approach in [6], whereas we derived simpler closed-form expressions for them as per Theorem 3 and Theorem 1.



(a) Steepest descent for $\mathbf{Q}$ (Sec. 3.3) (b) Joint design of $\mathbf{B}$, $\mathbf{F}$ and $\mathbf{Q}$

Fig. 3. Speed of convergence for iterative algorithms

4. NUMERICAL SIMULATIONS

We first study the convergence behavior of the numerical algorithms discussed in Sec. 3 and then compare the BER performance of different designs. The following configurations and parameters are used for the relay-assisted system: $N_S = N_R = N_D = 4$, $\mathbf{R}_w = \sigma_w^2\mathbf{I}$, $\mathbf{R}_n = \sigma_n^2\mathbf{I}$ and $\sigma_w^2 = \sigma_n^2$. We define two signal-to-noise ratio (SNR) $\rho_1 = P_S/(N_S\sigma_w^2)$ and $\rho_2 = P_R/(N_R\sigma_n^2)$ that represent the link quality of the first-hop and second-hop transmissions, respectively.

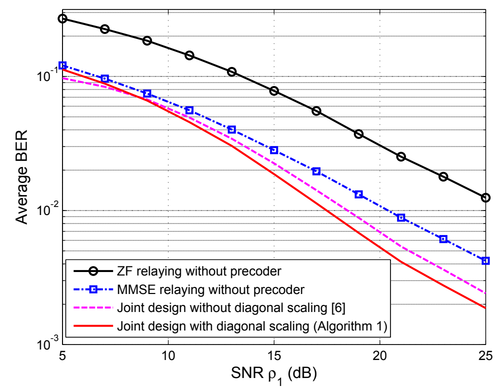
Convergence of the steepest descent method: For a typical channel realization, Fig. 3(a) shows the values of $g(\mathbf{Q})$ in (24) versus the number of iterations with five different randomly generated $\mathbf{Q}_0$. $\rho_1 = \rho_2 = 15$dB and $\mathbf{W} = \mathbf{I}$. In general, the steepest descent method of designing $\mathbf{Q}$ converges rapidly. With different initial matrices $\mathbf{Q}_0$, the algorithm may converges to different local optima but with insignificant difference in g.

Convergence of joint design approaches: We compare the speed of convergence between the two approaches mentioned in Sec. 3.4: the proposed Algorithm 1 and the conventional approach. The initial values are $\mathbf{B}_0 \propto \mathbf{I}$, $\mathbf{Q}_0 \propto \mathbf{I}$ and $\mathbf{F}_0 \propto \mathbf{I}$.[1] As shown in Fig. 3(b), by removing $\mathbf{F}$ from the iterating process, Algorithm 1 converges much faster than the conventional algorithm, in fact after just one iteration. A possible explanation is that at high SNRs, the second terms on the right-hand side of both (15) and (23) are close to zero. Hence, it follows from (12) that the optimal $\mathbf{B}_o$ would almost be independent of $\mathbf{Q}_o$ and vice versa.

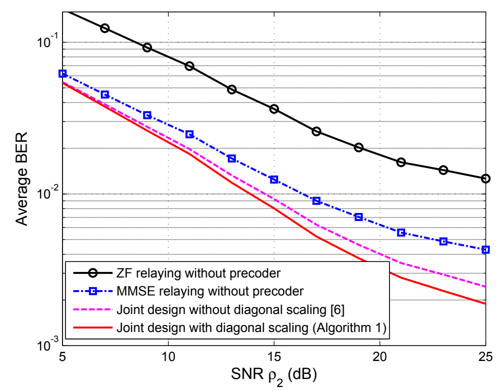
Comparison of BER: In the simulations, the BS transmits four independent uncoded QPSK symbol streams to their corresponding destination users. The wireless channel between the BS and the relay satisfies the Kronecker model: $\mathbf{H} = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}$. Herein, $\mathbf{H}_w$ has zero-mean, unit-variance, circularly symmetric complex Gaussian entries that are statistically independent. The (i, j)-th entries of $\mathbf{R}_r$ and $\mathbf{R}_t$ are both $0.7^{|i-j|}$. The forward channel satisfies $\mathbf{G} = \mathbf{D}\mathbf{G}_w\mathbf{R}_t^{1/2}$, where $\mathbf{G}_w$ has the same statistical characteristics as $\mathbf{H}_w$ and the diagonal matrix $\mathbf{D}$ represents relative pathlosses of 0, 0, 3 and 6dB for different users.

The methods under comparison are: 1) ZF relaying [17]: $\mathbf{B} = \sqrt{P_S/N_S}\mathbf{I}$, $\mathbf{F} = \eta\mathbf{G}^\dagger\mathbf{H}^\dagger$ and $\mathbf{Q} \propto \mathbf{I}$. 2) MMSE relaying without precoder [18]: $\mathbf{B} = \sqrt{P_S/N_S}\mathbf{I}$, $\mathbf{F}$ as in (11) and $\mathbf{Q} \propto \mathbf{I}$. 3) Joint design of $\mathbf{B}$ and $\mathbf{F}$ without diagonal scaling ($\mathbf{Q} \propto \mathbf{I}$) [6]; 4)

[1]The initial MSE values are different for the two methods because the optimal $\mathbf{F}_o$ for $\mathbf{B}_0$ and $\mathbf{Q}_0$ has already been implicitly selected in Algorithm 1.



(a) BER versus ρ_1 with $\rho_2 = 25$dB.



(b) BER versus ρ_2 with $\rho_1 = 25$dB.

Fig. 4. BER comparison of a 4-user BC (QPSK, $N_S = N_R = N_D = 4$, $\mathbf{W} = \mathbf{I}$).

Proposed joint design of $\mathbf{B}$, $\mathbf{F}$ and $\mathbf{Q}$ with diagonal scaling (Algorithm 1). The average BERs of the multiple users are shown in Fig. 4(a) and 4(b), where we let $\rho_1 = 25$dB or $\rho_2 = 25$dB and vary the other SNR between 5 and 25dB. As expected, MMSE relaying without precoder always outperforms ZF relaying without precoder. The joint design of $\mathbf{B}$ and $\mathbf{F}$ leads to an SNR gain of 2–3dB over the MMSE relaying without precoder. Furthermore, by including the diagonal $\mathbf{Q}$ in the joint design, the proposed diagonal scaling scheme enables an additional SNR gain of 0.5–2dB at mid-to-high SNRs.

5. CONCLUSIONS

In this paper, we have studied the joint transceiver design for the downlink of a cellular network assisted by a multi-antenna relay. Considering the possible differences in pathloss, we proposed a diagonal scaling scheme which allows the multiple users to apply their own complex-valued scaling factors to their received signals before decoding. The corresponding diagonal equalizer matrix is optimized together with a precoder at the BS and a linear processing matrix at the relay, to minimize the weighted MSE under power constraints. In particular, the optimal relaying matrix was first derived in closed form as a function of the precoder and the equalizer. This enabled the development of an efficient iterative algorithm (Algorithm 1) for the joint design. Simulation results demonstrate lower BER with the proposed diagonal scaling scheme than previous methods.

6. REFERENCES

[1] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, "Relay architectures for 3GPP LTE-Advanced," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, Article ID 618787, 14 pages, 2009.

[2] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4837–4851, Dec. 2009.

[3] B. K. Chalise and L. Vandendorpe, "MIMO relaying for multiaccess communication in cellular networks," in *Proc. 5th IEEE Sensor Array and Multichannel Signal Process. Workshop*, Darmstadt, Germany, Jul. 2008, pp. 146–150.

[4] R. Zhang, C. C. Chai, and Y. C. Liang, "Joint beamforming and power control for multiantenna relay broadcast channel with QoS constraints," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 726–737, Feb. 2009.

[5] C. B. Chae, T. Tang, R. W. Heath, and S. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay networks," *IEEE Trans. Signal Process.*, vol. 56, pp. 727–738, Feb. 2008.

[6] G. Li, Y. Wang, T. Wu, and J. Huang, "Joint linear filter design in multi-user non-regenerative MIMO-relay systems," in *Proc. IEEE Int. Conf. Commun.*, Dresden, Germany, Jun. 2009, pp. 1–6.

[7] Y. Cai, D. Le Ruyet, and D. Roviras, "Joint interference suppression and power allocation techniques for multiuser multi-antenna relay broadcast systems," in *Proc. 7th Int. Symp. Wireless Commun. Sys.*, York, UK, Sep. 2010, pp. 265–269.

[8] W. Xu, X. Dong, and W.-S. Lu, "Joint precoding optimization for multiuser multi-antenna relaying downlinks using quadratic programming," *IEEE Trans. Commun.*, vol. 59, pp. 1228–1235, May 2011.

[9] J. Liu, Z. Liu, and Z. Qiu, "Joint MMSE transceiver design for multiuser non-regenerative MIMO relay downlink systems," in *Proc. 7th Int. Wireless Commun. Mobile Comput. Conf.*, Istanbul, Turkey, Jul. 2011, pp. 877–882.

[10] B. Zhang, Z. He, K. Niu, and L. Zhang, "Robust linear beamforming for MIMO relay broadcast channel with limited feedback," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 209–212, Feb. 2010.

[11] C. Xing, S. Ma, M. Xia, and Y.-C. Wu, "Cooperative beamforming for dual-hop amplify-and-forward multi-antenna relaying cellular networks," *Signal Process.*, vol. 92, no. 11, pp. 2689–2699, 2012.

[12] Stephen P. Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, UK, 2004.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, Cambridge, UK, 1985.

[14] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer Verlag, New York, USA, 2nd edition, 2006.

[15] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, pp. 2700–2712, Aug. 2005.

[16] S. Ma, C. Xing, Y. Fan, Y.-C. Wu, T.-S. Ng, and H. V. Poor, "Iterative transceiver design for MIMO AF relay networks with multiple sources," in *Proc. Military Commun. Conf.*, San Jose, California, USA, Oct.-Nov. 2010, pp. 369–374.

[17] R. Louie, Yonghui Li, and B. Vucetic, "Zero forcing in general two-hop relay networks," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 191–202, Jan. 2010.

[18] O. Oyman and A. J. Paulraj, "Design and analysis of linear distributed MIMO relaying algorithms," *IEE Proc.-Commun.*, vol. 153, pp. 565–572, Apr. 2006.