A Unified Approach to Optimal Transceiver Design for Nonregenerative MIMO Relaying

Chao Zhao and Benoît Champagne, *Senior Member, IEEE*

***Abstract*—We propose a unified approach to transceiver optimization for nonregenerative multiple-input–multiple-output (MIMO) relay networks. This approach leads to new transceiver designs and reduces algorithmic complexity with adaptive implementations. First, we formulate a generic system model that accommodates various network topologies by imposing structural constraints on the source precoder, the relaying matrix, and the destination equalizer. Based on the minimum-mean-square-error (MMSE) criterion, we derive the optimal relaying matrix as a function of the other two matrices, thereby freeing the optimization problem from this matrix and its associated power constraint with no loss of optimality. Subsequently, we study how to optimize either the precoder or the equalizer under different structural constraints and propose an alternating algorithm for the joint design. When optimizing the equalizer, the optimum from the previous transmission block is chosen as the initial search point to speed up convergence and, hence, to reduce computational complexity. The proposed framework is further explained and numerically validated within the context of different system configurations.**

***Index Terms*—Multiple-input multiple-output (MIMO), minimum mean square error (MMSE), relaying, transceiver design.**

I. Introduction

RELAY-ASSISTED communications have attracted interest in both academia and in industry [1]. The introduction of multiple-input multiple-output (MIMO) technology into the relaying framework, through the use of multiple antennas at the sources, relays, or destinations, brings further advantages in performance [2]. For multiantenna relays, the simplest and perhaps also the most practical strategy is two-hop half-duplex amplify-and-forward relaying. Herein, the source users transmit signals to the relays in the first time slot; the relays then apply linear processing to their received signals and retransmit them to the destinations in the second time slot. There exist various relaying topologies with different numbers of sources, relays, and destinations. For convenience, we use these numbers to identify such configurations, e.g., 1S-MR-1D represents a relay network with one source, multiple relays, and one destination.

Manuscript received January 29, 2014; revised June 27, 2014; accepted August 13, 2014. Date of publication August 26, 2014; date of current version July 14, 2015. This work was supported in part by InterDigital Canada Ltée and in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The review of this paper was coordinated by Prof. W. A. Krzymien.

C. Zhao is with Microsemi Corporation, Ottawa, ON K2K 3H4, Canada (e-mail: chao.zhao@mail.mcgill.ca).

B. Champagne is with the Department of Electrical and Computer Engineering, McGill University, Montréal, QC H3A 0E9, Canada (e-mail: benoit.champagne@mcgill.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2351802

The optimal design of relaying matrices (possibly together with source precoders and destination equalizers) has been extensively studied; see [2] for a comprehensive survey of the existing literature. This problem can be defined in either of the following two ways: 1) to minimize the power usage subject to various quality of service (QoS) constraints [3], [4]; and 2) to minimize (or maximize) a certain performance criterion subject to power constraints. In this paper, we concentrate on the latter approach.

For the 1S-1R-1D topology, the optimal relaying matrix takes the form of a singular value decomposition (SVD) under a wide variety of criteria [5]–[10]. For other topologies, the major difficulty comes from structural constraints such as block diagonality on the underlying mathematical models. The SVD-based relaying framework cannot be readily generalized, and most existing works take either one of the following approaches to optimal transceiver design. The first type of approaches imposes special structures on the matrices under design [11]–[17]. For example, in [11] and [17], the relaying matrices are generated by cascading two substructures, akin to an equalizer for the backward channel and a precoder for the forward channel. The second type of approaches optimizes the precoders, the relaying matrices, and the equalizers in turn, in a way that guarantees convergence to a local optimum [18]–[35]. The third type of approaches turns to general optimization techniques such as semidefinite programming or second-order cone programming [36]–[42]. The first type is far from optimal, whereas the latter two may suffer from slow convergence and do not provide much physical insight. Moreover, these works are confined to an individual transmission block. In practice, the underlying wireless channels change only slightly across successive blocks. The proper exploitation of this property may simplify algorithms that are complex when the neighboring blocks are viewed in isolation.

Motivated by the goal of achieving mathematical simplicity, computation efficiency, and providing physical insights, we further investigate the transceiver design problem in this paper. Our previous work [43] focuses on the 1S-MR-1D topology. The essential feature of this topology is that the signals from different relays should be coherently combined at the destinations, thereby leading to a distributed array gain. However, since the relays can only process their own signals, the relaying matrix has to be *block diagonal*. Most recent works on the multirelay problem rely on numerical algorithms such as gradient descent (GD) [31], bisection [32], and other iterative schemes [33] to obtain this matrix. In [43], we were able to derive a simple closed-form expression for it, which brought physical insights and simplifications to the design problem. However,

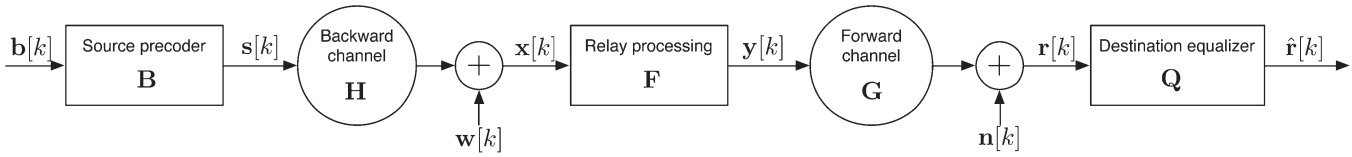


Fig. 1. Unified system model for nonregenerative MIMO relaying.

we did not consider the presence of a precoder or the previously mentioned interblock relationship.

In this paper, we propose a transceiver optimization framework that allows us to treat not only 1S-MR-1D but also other topologies in a unified way. First, we formulate a generic system model that includes a source precoder (which is not considered in [43]), a relaying matrix, and a destination equalizer as design variables (matrices). This model can accommodate various relaying topologies by imposing structural constraints such as block diagonality on these matrices. Then, we design those matrices to minimize the mean square error (MSE) between the source input and the destination output, subject to power constraints. More specifically, we first derive the optimal relaying matrix as a function of the other two matrices, thereby freeing the optimization problem from this matrix and its corresponding power constraint, without loss of optimality. This is the common step for point-to-point and multiuser systems. Subsequently, we study optimization of either the precoder or the equalizer under different structural constraints and propose an alternating algorithm for the joint design of these two matrices. When optimizing the equalizer, the optimal equalizer from the previous block can serve as the initial search point for the current block, which speeds up convergence and, henceforth, reduces computational complexity significantly.

For various topologies, this unified framework is consistent in the choice of the performance metric, formulation of the mathematical model, treatment of the relaying matrix, and exploitation of interblock adaptation. We explain how the proposed framework can accommodate different relaying topologies and validate the performance of the resulting designs through simulations over fading channels. In many instances, this unified framework leads to new algorithms with lower complexity or better performance. For example, for the 1S-1R-MD topology with single-antenna users, a new diagonal scaling scheme is obtained, which provides better bit error rate (BER) performance. Diagonal scaling allows different users to apply their own amplitude scaling and phase rotation before decoding, in contrast to [19], which assumes the same scaling for these users. 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 previously used.

The organization of this paper is as follows. Section II formulates the unified system model. In Section III, we define the optimization problem and then derive the optimal relaying matrix. In Section IV, we study in detail how to optimize either the precoder or the equalizer with various structural constraints. Subsequently, the precoder and the equalizer are jointly designed based on an alternating algorithm. Interblock adaptation is introduced in Section V to exploit the slow variations between successive transmission blocks. Section VI makes further comments on typical network topologies. The numerical results are presented and discussed in Section VII, followed by a brief conclusion in Section VIII.

The following notations are used: Italic, boldface lowercase, and boldface uppercase letters represent scalars, vectors, and matrices; $\bar{(\cdot)}$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^\dagger$ denote conjugate, transpose, Hermitian transpose, and Moore–Penrose pseudo-inverse, respectively; tr() refers to the trace of a matrix; $\|\cdot\|_2$ ($\|\cdot\|_{\mathrm{F}}$) stands for the Euclidean (or Frobenius) norm of a vector; $\mathrm{col}()$ stacks many column vectors into a single vector, $\mathrm{vec}()$ stacks the columns of a matrix into a vector, and $\mathrm{unvec}()$ is its inverse operator; $\mathrm{diag}()$ forms a diagonal (or block diagonal) matrix from multiple scalars (or matrices); $\otimes$ represents the Kronecker product; $\mathbf{I}_N$ is an identity matrix of dimension N; E{} refers to mathematical expectation; $\mathbb{C}$ denotes the set of complex numbers, and $\mathbb{Z}$ is the set of integers.

II. System Model

Fig. 1 shows a generic system model that can represent different types of MIMO relaying systems. We first explain its operations in terms of 1S-1R-1D and then show how it can be modified to accommodate more complex topologies by imposing structural constraints on its constituent building blocks. This system operates in a two-hop half-duplex mode: In the first time slot, the source transmits signals to the relay through the backward channel; in the second time slot, the relay forwards these signals to the user via the forward channel. The relay applies a linear transformation matrix to its received signals before retransmitting them. The direct source–destination link is neglected due to the assumed high level of attenuation. The numbers of antennas at the source, relay, and destination are, respectively, N_S, N_R, and N_D.

Assuming flat block fading for the wireless channels, we propose the following discrete-time complex baseband signal model. Herein, the signal and noise vectors incessantly change and, therefore, are modeled as discrete-time random processes with time index $k \in \mathbb{Z}$. In contrast, the various channel and system matrices are assumed to remain constant within each block and change only across successive blocks. For now, we focus on the optimal transceiver design for a single block and, therefore, omit the dependence of those matrices on the block index. However, this dependence will be reintroduced in Section V when we consider neighboring blocks together. The input symbol vector $\mathbf{b}[k] \in \mathbb{C}^{N_B \times 1}$, with zero mean and covariance matrix $\mathbf{R}_b = \mathbf{I}_{N_B}$, consists of N_B statistically independent symbols. This vector is preprocessed by a linear *precoder* matrix $\mathbf{B} \in \mathbb{C}^{N_S \times N_B}$ to generate the transmitted signal vector, i.e.,

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

TABLE I
TYPICAL SYSTEM CONFIGURATIONS

	$\mathbf{Q}$ *is arbitrary*	$\mathbf{Q}$ *is diagonal/blockdiagonal*
$\mathbf{B}$ *is arbitrary*	Point-to-point channel	Broadcast channel (BC)
$\mathbf{B}$ *is diagonal/blockdiagonal*	Multiple access channel (MAC)	Interference channel (IC)

*$\mathbf{F}$ is arbitrary for single-relay systems, and diagonal/blockdiagonal for multi-relay systems.

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

$$\mathbf{x}[k] = \mathbf{H}\mathbf{s}[k] + \mathbf{w}[k] \tag{2}$$

where $\mathbf{w}[k] \in \mathbb{C}^{N_R \times 1}$ 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 $\mathbf{x}[k]$ as in

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

The signal received by the destination user is

$$\mathbf{r}[k] = \mathbf{G}\mathbf{x}[k] + \mathbf{n}[k] = \mathbf{G}\mathbf{F}\mathbf{H}\mathbf{s}[k] + \mathbf{G}\mathbf{F}\mathbf{w}[k] + \mathbf{n}[k] \tag{4}$$

in which $\mathbf{G} \in \mathbb{C}^{N_D \times N_R}$ denotes the forward channel matrix from the relay to the destination. The noise term $\mathbf{n}[k]$ is independent from $\mathbf{b}[k]$ and $\mathbf{w}[k]$ and modeled as a circularly symmetric complex Gaussian random vector with zero mean and covariance $\mathbf{R}_n$. The destination applies a linear equalizer $\mathbf{Q} \in \mathbb{C}^{N_B \times N_D}$ whose output is

$$\hat{\mathbf{r}}[k] = \mathbf{Q}\mathbf{r}[k]. \tag{5}$$

It is straightforward to extend the given model to other relaying topologies. Let the numbers of physically separated sources, relays, and destinations be L, M, and N, respectively. The channel matrix $\mathbf{H}$ now consists of $M \times L$ blocks, *viz.,*

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{H}_{1,1} & \cdots & \mathbf{H}_{1,L} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{M,1} & \cdots & \mathbf{H}_{M,L} \end{bmatrix} \tag{6}$$

where $\mathbf{H}_{m,l}$, the (m,l)th block, corresponds to the MIMO channel from source l to relay m. Similary, matrix $\mathbf{G}$ has $N \times M$ blocks, *viz.,*

$$\mathbf{G} \triangleq \begin{bmatrix} \mathbf{G}_{1,1} & \cdots & \mathbf{G}_{1,M} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_{N,1} & \cdots & \mathbf{G}_{N,M} \end{bmatrix} \tag{7}$$

where $\mathbf{G}_{n,m}$ is the channel between relay m and destination n. The processing matrices $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{Q}$ are block diagonal, i.e.,

$$\mathbf{B} \triangleq \operatorname{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_L) \tag{7a}$$

$$\mathbf{F} \triangleq \operatorname{diag}(\mathbf{F}_1, \ldots, \mathbf{F}_M) \tag{7b}$$

$$\mathbf{Q} \triangleq \operatorname{diag}(\mathbf{Q}_1, \ldots, \mathbf{Q}_N). \tag{7c}$$

These matrices become diagonal if all the corresponding nodes use a single antenna. The noise vectors are the stacked versions of the individual vectors, *viz.,*

$$\mathbf{w}[k] \triangleq \operatorname{col}(\mathbf{w}_1[k], \ldots, \mathbf{w}_M[k]) \tag{7d}$$

$$\mathbf{n}[k] \triangleq \operatorname{col}(\mathbf{n}_1[k], \ldots, \mathbf{n}_N[k]) \tag{7e}$$

where $\mathbf{w}_m[k]$ is the additive noise induced at the mth relay, and $\mathbf{n}_n[k]$ is the noise at the nth destination. The relationship between structural constraints and system configurations is summarized in Table I.

III. UNIFIED FRAMEWORK FOR TRANSCEIVER DESIGN

A. Problem Formulation

We consider the general problem of optimizing relaying matrix $\mathbf{F}$, source precoder $\mathbf{B}$, and equalizer $\mathbf{Q}$, to minimize the MSE between the precoder input and the equalizer output, as given by

$$\begin{aligned} \operatorname{MSE}(\mathbf{F},\mathbf{B},\mathbf{Q}) &\triangleq \mathrm{E}\left\{\|\hat{\mathbf{r}}[k] - \mathbf{b}[k]\|^2\right\} \\ &= \operatorname{tr}\left(\mathrm{E}\left\{(\hat{\mathbf{r}}[k] - \mathbf{b}[k])(\hat{\mathbf{r}}[k] - \mathbf{b}[k])^H\right\}\right) \\ &= \operatorname{tr}\left((\mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{H}\mathbf{B} - \mathbf{I})(\mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{H}\mathbf{B} - \mathbf{I})^H\right) \\ &\quad + \operatorname{tr}(\mathbf{Q}\mathbf{G}\mathbf{F}\mathbf{R}_w\mathbf{F}^H\mathbf{G}^H\mathbf{Q}^H) + \operatorname{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H). \end{aligned} \tag{8}$$

Two power constraints are simultaneously imposed. The first constraint is on the expected transmit power of the source, i.e.,

$$\mathrm{E}\left\{\|\mathbf{s}[k]\|_2^2\right\} = \operatorname{tr}(\mathbf{R}_s) = \operatorname{tr}(\mathbf{B}\mathbf{B}^H) \le P_s. \tag{9}$$

For multisource topologies, a per-user constraint should be used instead, which will be discussed later. The other constraint is on the expected transmit power of the relay, i.e.,

$$\mathrm{E}\left\{\|\mathbf{y}[k]\|_2^2\right\} = \operatorname{tr}(\mathbf{R}_y) = \operatorname{tr}(\mathbf{F}\mathbf{R}_x\mathbf{F}^H) \le P_r \tag{10}$$

where $\mathbf{R}_x \triangleq \mathrm{E}\{\mathbf{x}[k]\mathbf{x}[k]^H\} = \mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_w$. For multirelay systems, $\mathbf{F}$ is block diagonal, and this constraint corresponds to the sum power of the relays.

A popular approach to solving similar problems starts from equalizer $\mathbf{Q}$. For any choice of $\mathbf{B}$ and $\mathbf{F}$, the optimal equalizer $\mathbf{Q}^\star$ is the minimum MSE (MMSE) equalizer [44], i.e.,

$$\mathbf{Q}^\star = \mathbf{R}_{rb}^H\mathbf{R}_r^{-1} \tag{11}$$

where the cross-correlation matrix $\mathbf{R}_{rb}$ and the covariance matrix $\mathbf{R}_r$ are defined as

$$\mathbf{R}_{rb} \triangleq \mathrm{E}\{\mathbf{r}\mathbf{b}^H\} = \mathbf{GFHB} \tag{12}$$

$$\mathbf{R}_r \triangleq \mathrm{E}\{\mathbf{r}\mathbf{r}^H\} = \mathbf{GFHBB}^H\mathbf{H}^H\mathbf{F}^H\mathbf{G}^H + \mathbf{GFR}_w\mathbf{F}^H\mathbf{G}^H + \mathbf{R}_n. \tag{13}$$

However, in our case, after substituting $\mathbf{Q}^\star$ into the MSE expression in (8), the remaining problem is nonconvex, *and* the number of constraints remains the same as in the original problem, which is difficult to solve. Moreover, this approach does not work in the presence of multiple destinations because $\mathbf{Q}$ has to be diagonal or block diagonal.

In this paper, we take a different approach, which begins with the relaying matrix $\mathbf{F}$. The first step is to derive the optimal $\mathbf{F}^\star$ as a closed-form function of $\mathbf{B}$ and $\mathbf{Q}$, thereby removing constraint (10) from the problem. This alternative approach offers two important advantages.

1) Various network topologies can now be treated in a unified manner because we do not need to consider their individual structural constraints until the next step.
2) After substituting the optimal $\mathbf{F}^\star$ into (8), the remaining optimization problem is unconstrained with respect to $\mathbf{Q}$. This fosters the use of numerical algorithms such as GD, which, in turn, makes it both easy to handle structural constraints on $\mathbf{Q}$ and convenient to exploit the slow variations between successive blocks.

B. Optimal Relaying Matrix for Single-Relay Topologies

Here, we derive the optimal relaying matrix $\mathbf{F}^\star$ as a function of precoder $\mathbf{B}$ and equalizer $\mathbf{Q}$ for single-relay topologies. This step is the same for different network topologies. The constraint in (9) does not depend on $\mathbf{F}$ and, henceforth, does not need to be considered for now.

The convex optimization problem can be solved by defining the Lagrangian multiplier, i.e.,

$$\mathcal{L}(\mathbf{F}, \lambda) = \mathrm{MSE}(\mathbf{F}, \mathbf{B}, \mathbf{Q}) + \lambda\left(\mathrm{tr}(\mathbf{FR}_x\mathbf{F}^H) - P_r\right) \tag{14}$$

where $\lambda > 0$. The dual problem is to

$$\begin{aligned} \text{maximize} \quad & D(\lambda) \triangleq \inf_{\mathbf{F}} \mathcal{L}(\mathbf{F}, \lambda) \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned} \tag{15}$$

Let the solution of the primal problem be $\mathbf{F}^\star$ and define $p^\star \triangleq \mathrm{MSE}(\mathbf{F}^\star)$; let the solution of the dual problem be $\lambda^\star$ and $d^\star \triangleq D(\lambda^\star)$. For a convex primal problem, strong duality holds (i.e., the duality gap $p^\star - d^\star$ is zero) if Slater's condition is satisfied [45, p. 226]. Here, the primal problem is convex, and Slater's condition is always satisfied ($\mathbf{F} = \mathbf{0}$ is strictly feasible: $0 < P_r$). Henceforth, the Karush–Kuhn–Tucker (KKT) conditions are *necessary and sufficient* for the optimal primal–dual pair $(\mathbf{F}^\star, \lambda^\star)$, *viz.*,[1]

$$\left.\frac{\partial \mathcal{L}}{\partial \bar{\mathbf{F}}}\right|_{\mathbf{F}=\mathbf{F}^\star} = \mathbf{0} \tag{16}$$

$$\mathrm{tr}(\mathbf{F}^\star\mathbf{R}_x\mathbf{F}^{\star H}) - P_r \leq 0 \tag{17}$$

$$\lambda^\star \geq 0 \tag{18}$$

$$\lambda^\star\left(\mathrm{tr}(\mathbf{F}^\star\mathbf{R}_x\mathbf{F}^{\star H}) - P_r\right) = 0. \tag{19}$$

The first-order necessary condition in (16) is expressed as

$$(\mathbf{G}^H\mathbf{Q}^H\mathbf{QG} + \lambda^\star\mathbf{I})\mathbf{F}^\star\mathbf{R}_x = \mathbf{G}^H\mathbf{Q}^H\mathbf{B}^H\mathbf{H}^H. \tag{20}$$

If $\lambda^\star > 0$, there is a unique solution; if $\lambda^\star = 0$, the solution is not unique when $\mathbf{G}^H\mathbf{Q}^H\mathbf{QG}$ is not of full rank. In the latter case, we are interested in the particular solution (among all solutions) that leads to the smallest value of $\mathrm{tr}(\mathbf{FR}_x\mathbf{F}^H) = \|\mathbf{FR}_x^{1/2}\|_F^2$ and, hence, would most likely satisfy (17). This solution is obtained by taking the pseudo-inverse. Therefore, in both cases, the optimal solution has the same form, i.e.,

$$\mathbf{F}^\star(\lambda^\star) = (\mathbf{G}^H\mathbf{Q}^H\mathbf{QG} + \lambda^\star\mathbf{I})^\dagger\mathbf{G}^H\mathbf{Q}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1} \tag{21}$$

which is an explicit function of $\lambda^\star$. Two scenarios may apply.

1) If $\mathrm{tr}(\mathbf{F}^\star\mathbf{R}_x\mathbf{F}^{\star H})|_{\lambda^\star=0} \leq P_r$, $\lambda^\star = 0$ satisfies (17)–(19). That is, the power constraint at the relay is inactive, and the solution is equal to that of the unconstrained problem.
2) If $\mathrm{tr}(\mathbf{F}^\star\mathbf{R}_x\mathbf{F}^{\star H})|_{\lambda^\star=0} > P_r$, $\lambda^\star = 0$ is not a solution and according to (19), $\lambda^\star > 0$ has to satisfy $\mathrm{tr}(\mathbf{F}^\star\mathbf{R}_x\mathbf{F}^{\star H}) = P_r$. It is straightforward to prove that the left-hand side is a monotonically decreasing function of $\lambda^\star$. The value is larger than P_r at $\lambda^\star = 0$ and converges to zero as $\lambda^\star \to \infty$. Therefore, a unique solution exists and can be obtained via bisection or Newton's method.

The given formulation has been proposed in previous works such as [10]. We rephrase it in our own notations here because it is necessary for the following mathematical development.

Based on the strong duality, the MMSE can be obtained by substituting (21) into (14), *viz.*,

$$\mathrm{MSE}_{\min}(\mathbf{B}, \mathbf{Q}) = -\mathrm{tr}\left(\mathbf{QG}(\mathbf{G}^H\mathbf{Q}^H\mathbf{QG} + \lambda^\star\mathbf{I})^\dagger\mathbf{G}^H\mathbf{Q}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}\mathbf{HB}\right) + N_B + \mathrm{tr}(\mathbf{QR}_n\mathbf{Q}^H) - \lambda^\star P_r. \tag{22}$$

The given expression in (22) depends on parameter $\lambda^\star$, which, in turn, is an *implicit* function of $\mathbf{Q}$. This lack of an explicit formula would complicate later design, and our approach to overcoming this problem 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 $\mathbf{F}^\star$, and the corresponding MMSE are all functions of η. It turns out that for the optimal $\eta^\star$ leading to the smallest MSE, these quantities can all be expressed in closed forms, as presented in the following theorem (see the Appendix for proof).

[1] We could have taken the gradient with respect to either one of $\mathbf{F}$, $\bar{\mathbf{F}}$, or $\mathbf{F}^H$. Here, $\bar{\mathbf{F}}$ was chosen because it leads to the most convenient expression. Similar approaches were used in references such as [8].

Theorem 1: Any $\mathbf{Q}$ can be replaced by a particular scaled version $\eta^{\star-1}\mathbf{Q}$ ($\eta^\star > 0$) such that we have the following:

a) The optimal relaying matrix is in the closed form, i.e.,

$$\mathbf{F}^\star = \eta^\star(\mathbf{G}^H\mathbf{Q}^H\mathbf{Q}\mathbf{G} + \theta\mathbf{I})^{-1}\mathbf{G}^H\mathbf{Q}^H\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1} \quad (23)$$

where $\theta \triangleq \mathrm{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)/P_r$.

b) $\eta^\star$ is the unique number satisfying $\mathrm{tr}(\mathbf{F}^\star\mathbf{R}_x\mathbf{F}^{\star H}) = P_r$.

c) The MMSE with $\mathbf{F} = \mathbf{F}^\star$ takes the form

$$\mathrm{MSE}_{\min}(\mathbf{B},\mathbf{Q}) = \mathrm{tr}\left(\mathbf{I}_{N_B}\right) - \mathrm{tr}\left(\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}\mathbf{H}\mathbf{B}\mathbf{Q}\mathbf{G}(\mathbf{G}^H\mathbf{Q}^H\mathbf{Q}\mathbf{G} + \theta\mathbf{I})^{-1}\mathbf{G}^H\mathbf{Q}^H\right). \quad (24)$$

d) Any other choice of $\eta^{-1}\mathbf{Q}$ together with the corresponding $\mathbf{F}^\star$ in (21) would lead to an MSE no smaller than (24).

In a nutshell, although a closed form does not exist for $\mathbf{F}^\star$ under an arbitrary $\mathbf{Q}$, it is always possible to find another equalizer $\eta^{\star-1}\mathbf{Q}$, which leads to not only better performance but also a closed-form expression for $\mathbf{F}^\star$. We may visualize η^2 as a target signal power level at the destination, which is similar to some QoS-related works [4], [46]. For small η values ($\eta \leq \eta_c$, where η_c is defined in the proof in the Appendix), the power budget at the relay is sufficient to support the unconstrained optimal solution ($\lambda^\star = 0$). In this case

$$\mathbf{Q}\mathbf{G}\mathbf{F}^\star = \eta\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}$$

which is, in fact, the MMSE equalizer for the first-hop transmission. This would minimize the portion of the MSE due to the first hop, including the interstream interferences and the noise term $\mathbf{w}$. However, the part due to the noise term $\mathbf{n}$ (as in $\eta^{-2}\mathrm{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)$) is large. Once η exceeds the threshold η_c, the power budget P_r becomes insufficient, and therefore, the power regularization term $\lambda^\star\eta^2\mathbf{I}$ is introduced. This slightly increases the first-hop MSE, but the overall MSE still decreases because the part due to the noise term $\mathbf{n}$ is reduced by more. Nonetheless, there is a critical and therefore optimal $\eta^\star$ above which the latter can no longer completely compensate for the former.

In essence, the scaling factor η plays the pivotal role of balancing the performance impairment due to the first-hop transmission and the second-hop transmission. Therefore, in an alternative approach, the scaling factor in $\eta^{-1}\mathbf{Q}$ could have been introduced earlier in the definition of the MSE in (8), which is similar to the approach taken in [47] for a point-to-point MIMO system. Without η, the portion of the MSE due to the noise term $\mathbf{n}$ would be predetermined for any fixed $\mathbf{Q}$. In addition, it is worth mentioning that although one may be tempted to solve the resulting problem by setting to zero the partial derivatives with respect to both η and $\bar{\mathbf{F}}$, this does not formally guarantee optimality because the MSE is not a convex function of *both* η and $\mathbf{F}$.

From now on, we suppose that for any given $\mathbf{Q}$, the corresponding $\eta^{\star-1}\mathbf{Q}$ and $\mathbf{F}^\star$ in Theorem 1 are selected. Hence, the optimization problem is reduced to that of designing $\mathbf{B}$ and $\mathbf{Q}$ to minimize (24) subject to (9). For convenience, we define

$$\mathbf{E}_h \triangleq \left(\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\mathbf{B}\right)^{-1} \quad (25)$$

$$\mathbf{E}_g \triangleq (\mathbf{I} + \mathbf{Q}\mathbf{G}\mathbf{G}^H\mathbf{Q}^H/\theta)^{-1}. \quad (26)$$

It can be proved that

$$\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}\mathbf{H}\mathbf{B} = \mathbf{I} - \mathbf{E}_h \succeq \mathbf{0} \quad (27)$$

$$\mathbf{Q}\mathbf{G}(\mathbf{G}^H\mathbf{Q}^H\mathbf{Q}\mathbf{G} + \theta\mathbf{I})^{-1}\mathbf{G}^H\mathbf{Q}^H = \mathbf{I} - \mathbf{E}_g \succeq \mathbf{0} \quad (28)$$

in which $\succeq \mathbf{0}$ means positive semidefinite. Substituting the given equations into (24), the objective function becomes

$$\mathrm{MSE}_{\min}(\mathbf{B},\mathbf{Q}) = \mathrm{tr}(\mathbf{E}_h) + \mathrm{tr}(\mathbf{E}_g) - \mathrm{tr}(\mathbf{E}_h\mathbf{E}_g). \quad (29)$$

In fact, the optimal relaying matrix in (23) can be viewed as a cascade of an MMSE equalizer ($\mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}$) for the backward channel and an MMSE precoder ($\eta^\star(\mathbf{G}^H\mathbf{Q}^H\mathbf{Q}\mathbf{G} + \theta\mathbf{I})^{-1}\mathbf{G}^H\mathbf{Q}^H$) for the forward channel [47]. Correspondingly in (29), the first term $\mathrm{tr}(\mathbf{E}_h)$ is the MMSE achieved by the first-hop transmission alone [44], and $\mathrm{tr}(\mathbf{E}_g) - \mathrm{tr}(\mathbf{E}_h\mathbf{E}_g)$ is the penalty due to the second-hop transmission. The last term in (29), i.e., $-\mathrm{tr}(\mathbf{E}_h\mathbf{E}_g)$, is determined by both $\mathbf{B}$ and $\mathbf{G}$, implying that the joint design cannot be easily decoupled.

C. Optimal Relaying Matrix for Multirelay Topologies

Optimal design of the relaying matrix $\mathbf{F}$ for multirelay topologies is more complicated than single-relay topologies due to block diagonality, that is, $\mathbf{F} = \mathrm{diag}(\mathbf{F}_1,\ldots,\mathbf{F}_M)$ and $\mathbf{R}_w = \mathrm{diag}(\mathbf{R}_{w_1},\ldots,\mathbf{R}_{w_M})$. For convenience, we define $\mathbf{H}_m \triangleq [\mathbf{H}_{m,1},\ldots,\mathbf{H}_{m,L}]$ and $\mathbf{G}_m \triangleq [\mathbf{G}_{m,1}^T,\ldots,\mathbf{G}_{m,N}^T]^T$, for all $1 \leq m \leq M$. The Lagrangian function is obtained by substituting the block diagonal $\mathbf{F}$ into (14), *viz.*,

$$\mathcal{L}(\mathbf{F}_1,\ldots,\mathbf{F}_M,\lambda) \triangleq \mathcal{L}(\mathbf{F},\lambda)|_{\mathbf{F}=\mathrm{diag}(\mathbf{F}_1,\ldots,\mathbf{F}_M)}. \quad (30)$$

Similar to the single-relay topologies, the KKT conditions are necessary and sufficient for the optimal relaying matrices $\mathbf{F}_1^\star,\ldots,\mathbf{F}_M^\star$. The first-order necessary condition in (16) becomes

$$\left(\mathbf{G}_m^H\mathbf{Q}^H\mathbf{Q}\mathbf{G}_m + \lambda^\star\mathbf{I}\right)\mathbf{F}_m^\star\left(\mathbf{H}_m\mathbf{B}\mathbf{B}^H\mathbf{H}_m^H + \mathbf{R}_{w_m}\right) + \sum_{k\neq m}\mathbf{G}_m^H\mathbf{Q}^H\mathbf{Q}\mathbf{G}_k\mathbf{F}_k^\star\mathbf{H}_k\mathbf{B}\mathbf{B}^H\mathbf{H}_m^H = \mathbf{G}_m^H\mathbf{Q}^H\mathbf{B}^H\mathbf{H}_m^H \quad (31)$$

where $m = 1,\ldots,M$. By defining $\mathbf{f}_m^\star \triangleq \mathrm{vec}(\mathbf{F}_m^\star)$ and applying $\mathrm{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{C}^T\otimes\mathbf{A})\mathrm{vec}(\mathbf{B})$, the given equation is equivalent to

$$\left(\left(\mathbf{H}_m\mathbf{B}\mathbf{B}^H\mathbf{H}_m^H + \mathbf{R}_{w_m}\right)^T\otimes\left(\mathbf{G}_m^H\mathbf{Q}^H\mathbf{Q}\mathbf{G}_m + \lambda^\star\mathbf{I}\right)\right)\mathbf{f}_m^\star + \sum_{n\neq m}\left(\left(\mathbf{H}_n\mathbf{B}\mathbf{B}^H\mathbf{H}_m^H\right)^T\otimes\mathbf{G}_m^H\mathbf{Q}^H\mathbf{Q}\mathbf{G}_n\right)\mathbf{f}_n^\star = \mathrm{vec}\left(\mathbf{G}_m^H\mathbf{Q}^H\mathbf{B}^H\mathbf{H}_m^H\right). \quad (32)$$

These equations can be written in the following compact block-matrix form:

$$\mathbf{\Psi}(\lambda^\star)\mathbf{f}^\star = \mathbf{b} \tag{33}$$

in which the (m,n)th block of $\mathbf{\Psi}(\lambda^\star)$ is $(\mathbf{H}_n\mathbf{B}\mathbf{B}^H\mathbf{H}_m^H)^T \otimes \mathbf{G}_m^H\mathbf{Q}^H\mathbf{Q}\mathbf{G}_n$ if $m \neq n$ and $(\mathbf{H}_m\mathbf{B}\mathbf{B}^H\mathbf{H}_m^H + \mathbf{R}_{w_m})^T \otimes (\mathbf{G}_m^H\mathbf{Q}^H\mathbf{Q}\mathbf{G}_m + \lambda^\star\mathbf{I})$ if $m = n$. The vector $\mathbf{b}$ is obtained by stacking $\mathrm{vec}(\mathbf{G}_m^H\mathbf{Q}^H\mathbf{B}^H\mathbf{H}_m^H)$, $1 \le m \le M$ into a single column vector.

Similar to single-relay topologies, the optimal relaying *vector* $\mathbf{f}^\star$ can now be expressed in a closed form. However, the dual variable $\lambda^\star$ can be obtained only by solving a nonlinear equation based on numerical methods. Fortunately, as in the single-relay scenario, we can avoid this hurdle by introducing a similar complex scaling η^{-1} in equalizer $\mathbf{Q}$. Since the theoretical development is similar to that for single-relay topologies, we skip the details and present the main conclusions in the following theorem.

Theorem 2: Any equalizer $\mathbf{Q}$ can be replaced by a scaled version $\eta^{\star-1}\mathbf{Q}$ $(\eta^\star > 0)$ such that we have the following.

1) The optimal solution is

$$\mathbf{f}^\star = \eta^\star\mathbf{\Psi}(\theta)^{-1}\mathbf{b} \tag{34}$$

where $\theta \triangleq \mathrm{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)/P_r$.

2) $\eta^\star$ is the unique number satisfying $\mathrm{tr}(\mathbf{F}^\star\mathbf{R}_x\mathbf{F}^{\star H}) = P_r$.
3) The MMSE with $\mathbf{f}^\star$ takes the form

$$\mathrm{MSE}_{\min} = N_B - \mathbf{b}^H\mathbf{\Psi}(\theta)^{-1}\mathbf{b}. \tag{34a}$$

4) Any other choice of $\eta^{-1}\mathbf{Q}$ together with the corresponding $\mathbf{f}^\star$ would lead to an MSE no smaller than (35).

The sum power constraint seems not very practical because each relay should have its own power constraint. However, due to the randomness of the channels and users, this problem with the ideal narrowband configuration is probably not as important. First, most modern communication systems are based on a multicarrier scheme such as orthogonal frequency-division multiplexing. A relay station may transmit less power on one subcarrier but more on another, so that the variation in the total power is smaller. Second, the multiple relays are simultaneously serving several randomly located users (with different subcarriers or time intervals). This will further reduce the disparity between the transmit power of different relays. Finally, if the expected transmit power of a particular relay is abnormally small, the problem likely comes from an inappropriate network layout, and the relays should be relocated instead.

The use of multiple relays helps achieve a distributed array gain but, at the same time, increases both computational complexity and communication overhead. Due to space limits, we concentrate on the single-relay topologies in the following sections; however, in principle, the methodology is applicable to multirelay networks if their complexity is acceptable under the circumstances.

IV. Optimization of the Precoder and the Equalizer With Structural Constraints

Here, we consider the joint design of precoder $\mathbf{B}$ and equalizer $\mathbf{Q}$. First, we study the optimal design for either of the two matrices while the other is fixed. The structural constraints imposed by different system configurations must be appropriately handled. Then, we propose an alternating algorithm for the joint design.

A. Optimization of the Precoder $\mathbf{B}$ With Fixed Equalizer $\mathbf{Q}$

From (29), the objective is to minimize

$$g(\mathbf{B}) \triangleq \mathrm{tr}\,(\mathbf{E}_h(\mathbf{I} - \mathbf{E}_g)) \tag{35}$$

subject to (9). We consider two scenarios when $\mathbf{B}$ is either structurally unconstrained or diagonal. For the latter case, the power constraint is slightly different from (9).

Precoder with arbitrary structure. Define the eigenvalue decomposition (EVD)

$$\mathbf{I} - \mathbf{E}_g = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{36}$$

where $\mathbf{U}$ is unitary, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{N_D}) \succeq \mathbf{0}$ with its diagonal entries sorted in descending order. Similarly, define the EVD

$$\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H} = \mathbf{V}\mathbf{\Lambda}'\mathbf{V}^H \tag{37}$$

where $\mathbf{V}$ is unitary, and $\mathbf{\Lambda}' = \mathrm{diag}(\lambda_1', \ldots, \lambda_{N_S}') \succeq \mathbf{0}$. The problem of minimizing (29) subject to power constraint (9) is equivalent to that of minimizing

$$f(\mathbf{B}) \triangleq \mathrm{tr}\left(\mathbf{\Lambda}(\mathbf{I} + \mathbf{U}^H\mathbf{B}^H\mathbf{V}\mathbf{\Lambda}'\mathbf{V}^H\mathbf{B}\mathbf{U})^{-1}\right) \tag{38}$$

subject to (9). Based on similar techniques as in [44], we can prove the following theorem.

Theorem 3: For a fixed equalizer $\mathbf{Q}$, with the optimal $\mathbf{F}^\star$ in (23), the optimal precoder $\mathbf{B}$ that minimizes (29) is in the SVD form, i.e.,

$$\mathbf{B}^\star = \mathbf{V}\mathbf{\Sigma}_B\mathbf{U}^H. \tag{39}$$

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

$$b_k^2 = \left(\sqrt{\frac{\lambda_k}{\lambda_k'}}\frac{1}{\sqrt{\gamma}} - \frac{1}{\lambda_k'}\right)^+ \tag{40}$$

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

Diagonal precoder: This corresponds to network topologies with multiple single-antenna source users. It is impossible for the users to cooperate in terms of signaling because they are spatially distributed. Henceforth, precoder $\mathbf{B}$ has to be diagonal, and each user has its own power constraint, which can be expressed by the positive semidefinite ordering, i.e.,

$$\mathbf{B}\mathbf{B}^H \preceq \mathbf{P}_s = \mathrm{diag}(P_{s,1}, \ldots, P_{s,N_s}). \tag{41}$$

It is straightforward to prove that each user should transmit its maximum allowable power, that is

$$\mathbf{B} = \mathbf{P}_s^{\frac{1}{2}}. \tag{42}$$

B. Optimization of the Equalizer $\mathbf{Q}$ With Fixed Precoder $\mathbf{B}$

Since neither the power constraint in (9) nor matrix $\mathbf{E}_h$ depends on $\mathbf{Q}$, the problem is equivalent to that of minimizing

$$f(\mathbf{Q}) \triangleq \operatorname{tr}\left((\mathbf{I} - \mathbf{E}_h)\mathbf{E}_g\right). \tag{43}$$

This *unconstrained* optimization problem can be solved by a *gradient-based line search* algorithm. Starting from an initial matrix $\mathbf{Q}_0$, this method generates a sequence of iterates $\{\mathbf{Q}_n\}_{n=0}^{\infty}$ until a solution has been approximated with sufficient accuracy. Specifically, it chooses a direction $\Delta\mathbf{Q}_n$ and searches along it from the current $\mathbf{Q}_n$ for a new iterate $\mathbf{Q}_{n+1}$ so that $f(\mathbf{Q}_{n+1}) < f(\mathbf{Q}_n)$. The distance to move along $\Delta\mathbf{Q}_n$ should satisfy criteria such as Wolfe's conditions [48]. In particular, *GD* uses the opposite direction of the gradient

$$\Delta\mathbf{Q}_n = -\nabla_{\bar{\mathbf{Q}}}\mathrm{MSE}_{\min}|_{\mathbf{Q}=\mathbf{Q}_n}. \tag{44}$$

This method is summarized in Algorithm 1.

Algorithm 1: GD

1: Choose P_r and $\epsilon > 0$;
2: Choose line search parameters $\bar{\alpha} > 0$, $\rho, c \in (0,1)$;
3: Initialize the counter $n = -1$ and the equalizer $\mathbf{Q}_0$;
4: **repeat**
5: Increment counter $n \leftarrow n + 1$;
6: Compute $\Delta\mathbf{Q}_n$ from (44);
7: Set $\alpha \leftarrow \bar{\alpha}$;
8: **repeat**
9: $\alpha \leftarrow \rho\alpha$;
10: **until** $\mathrm{MSE}_{\min}(\mathbf{Q}_n + \alpha\Delta\mathbf{Q}_n) \le \mathrm{MSE}_{\min}(\mathbf{Q}_n) - c\alpha\|\Delta\mathbf{Q}_n\|_{\mathrm{F}}^2$;
11: Update $\mathbf{Q}_{n+1} \leftarrow \mathbf{Q}_n + \alpha\Delta\mathbf{Q}_n$;
12: **until** $\|\Delta\mathbf{Q}_n\|_{\mathrm{F}}^2 < \epsilon$.

This algorithm can handle different structural constraints conveniently. In the general case when $\mathbf{Q}$ is not constrained, the gradient is

$$\nabla_{\bar{\mathbf{Q}}}\mathrm{MSE}_{\min} = \eta^{-2}\mathbf{Q}\mathbf{R}_n - \theta^{-1}\mathbf{E}_g\mathbf{C}\mathbf{E}_g\mathbf{Q}\mathbf{G}\mathbf{G}^H \tag{45}$$

where $\mathbf{C} \triangleq \mathbf{B}^H\mathbf{H}^H\mathbf{R}_x^{-1}\mathbf{H}\mathbf{B}$. When $\mathbf{Q}$ is *diagonal*, we define $\mathbf{q} \triangleq \operatorname{diag}(\mathbf{Q}) = [q_1, \ldots, q_{N_D}]^T$. The gradient of $f(\mathbf{q})$ with respect to $\bar{\mathbf{q}}$ is simply the vector that includes all the diagonal entries of the given equation, that is

$$\nabla_{\bar{\mathbf{q}}}\mathrm{MSE}_{\min} = \eta^{-2}\operatorname{diag}(\mathbf{Q}\mathbf{R}_n) - \theta^{-1}\operatorname{diag}(\mathbf{E}_g\mathbf{C}\mathbf{E}_g\mathbf{Q}\mathbf{G}\mathbf{G}^H). \tag{46}$$

When $\mathbf{Q}$ is *block diagonal*, the gradient is equal to the Hadamard (element-wise) product of (45) and a block diagonal matrix $\mathbf{I}_{BD}$

$$\nabla_{\bar{\mathbf{Q}}}\mathrm{MSE}_{\min} = (\eta^{-2}\mathbf{Q}\mathbf{R}_n - \theta^{-1}\mathbf{E}_g\mathbf{C}\mathbf{E}_g\mathbf{Q}\mathbf{G}\mathbf{G}^H) \odot \mathbf{I}_{BD} \tag{47}$$

in which the diagonal blocks of $\mathbf{I}_{BD}$ are matrices whose entries are all one.

Initial search point when $\mathbf{Q}$ is diagonal: The choice of the initial $\mathbf{Q}_0$ for GD is very important. For this purpose, we propose to choose a phase rotation vector that can speed up the algorithms significantly. In particular, we temporarily assume that the diagonal entries of $\mathbf{Q}$ are all phase rotation terms, that is

$$\mathbf{Q} = \operatorname{diag}\{\mathbf{q}\} = \begin{bmatrix} e^{j\phi_1} & & \\ & \ddots & \\ & & e^{j\phi_{N_D}} \end{bmatrix}. \tag{48}$$

By defining the EVD $\mathbf{G}\mathbf{G}^H = \mathbf{U}_G\mathbf{\Lambda}_G\mathbf{U}_G^H$ with the eigenvalues $\lambda_1, \ldots, \lambda_{N_D}$, the objective function in (43) becomes

$$f(\mathbf{q}) = \operatorname{tr}\left(\mathbf{A}(\mathbf{I} + \mathbf{\Lambda}_G/\theta)\right)^{-1}) = \sum_{k=1}^{N_S} \frac{a_k}{\theta^{-1}\lambda_k + 1} \tag{49}$$

where $\mathbf{A} \triangleq \mathbf{U}_G^H\mathbf{Q}^H(\mathbf{I} - \mathbf{E}_h)\mathbf{Q}\mathbf{U}_G$. The diagonal entries of $\mathbf{A}$ are

$$a_k = \mathbf{u}_k^H\mathbf{Q}^H(\mathbf{I} - \mathbf{E}_h)\mathbf{Q}\mathbf{u}_k = \mathbf{q}^H\mathbf{U}_{(k)}^H(\mathbf{I} - \mathbf{E}_h)\mathbf{U}_{(k)}\mathbf{q} \tag{50}$$

where $\mathbf{U}_G = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_D}]$, and $\mathbf{U}_{(k)} \triangleq \operatorname{diag}(\mathbf{u}_k)$. We then have to minimize

$$f(\mathbf{q}) = \mathbf{q}^H\mathbf{\Psi}\mathbf{q}, \quad \text{where } \mathbf{\Psi} \triangleq \sum_{k=1}^{N_S} \frac{\mathbf{U}_{(k)}^H(\mathbf{I} - \mathbf{E}_h)\mathbf{U}_{(k)}}{\theta^{-1}\lambda_k + 1}. \tag{51}$$

This is not a trivial problem because the entries of $\mathbf{q}$ are all of norm one [49]. Without loss of generality, we can assume $\phi_1 = 0$. Define the EVD $\mathbf{\Psi} = \mathbf{U}_\Psi\mathbf{\Lambda}_\Psi\mathbf{U}_\Psi^H$, with the eigenvalues sorted in descending order. The objective function becomes the following weighted sum:

$$f(\mathbf{q}) = \sum_{k=1}^{N_D} \lambda_{\Psi,k}\left\|\mathbf{u}_{\Psi,k}^H\mathbf{q}\right\|^2 \tag{52}$$

where $\mathbf{u}_{\Psi,k}$ is the kth column of $\mathbf{U}_\Psi$. Since the entries of $\mathbf{q}$ must have the same magnitude, the solution cannot be the eigenvector of $\mathbf{\Psi}$ corresponding to its smallest eigenvalue $(\mathbf{u}_{\Psi,N_D})$. A heuristic suboptimal approach is to maximize the term with the smallest weight λ_{Ψ,N_D}. This is achieved by letting

$$\phi_k = \frac{\mathbf{u}_{\Psi,N_D}(k)}{|\mathbf{u}_{\Psi,N_D}(k)|}. \tag{53}$$

We should bear in mind that global optimality is not of much importance here because this ϕ_k is merely used as the initial search point.

C. Joint Design

Since optimization of either precoder $\mathbf{B}$ or equalizer $\mathbf{Q}$ has been discussed, we now propose an alternating algorithm for the joint design of these two matrices. This algorithm optimizes $\mathbf{B}$ and $\mathbf{Q}$ interchangeably while holding the other matrix constant. This process generates a descending sequence of MSE values and is therefore guaranteed to converge to a local optimum. Afterward, the optimal $\mathbf{F}^\star$ and $\eta^\star$ are computed as in Theorem 1. This framework is described in Algorithm 2.

Algorithm 2: Joint optimization of $\mathbf{B}$, $\mathbf{Q}$ and $\mathbf{F}$

1: Initialize $\mathbf{Q}_0$, $n = 0$, $\epsilon > 0$;
2: **repeat**
3: Increment the counter $n \leftarrow n + 1$;
4: Optimize and update $\mathbf{B}_n$;
5: Optimize and update $\mathbf{Q}_n$;
6: **until** $|\mathrm{MSE}_{\min}(\mathbf{Q}_n, \mathbf{B}_n) - \mathrm{MSE}_{\min}(\mathbf{Q}_{n-1}, \mathbf{B}_{n-1})| \leq \epsilon$;
7: Compute $\mathbf{F}^\star$ and $\eta^\star$ according to Theorem 1;
8: Let $\mathbf{B}^\star \leftarrow \mathbf{B}_n$, $\mathbf{Q}^\star \leftarrow \eta^{\star -1}\mathbf{Q}_n$.

We remark that the removal of the relaying matrix $\mathbf{F}$ from the optimization problem does not jeopardize the optimality of the final solution. The suboptimality comes from the alternate algorithm, which is well known to converge to a local optimum. As of now, the optimal solution for the joint design problem is not known in the current literature, and therefore, the performance gap is not measurable. The best we can do is test the algorithm by choosing different initial search points. As it will shown in Section VII-A, different initial points almost lead to the same performance.

Algorithm 2 is different from a conventional alternating approach widely used to solve similar problems [19], [20], [27], [50]. The latter needs to alternate between *all three* matrices $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{Q}$. As it will be numerically shown in Section VII-A, the proposed algorithm converges significantly faster by removing the relaying matrix $\mathbf{F}$ from this alternating process. Moreover, in the conventional approach, updating $\mathbf{B}$ also changes the transmit power of the *relay* such that the combination of $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{Q}$ may temporarily violate the power constraint in (10). In the next step, after $\mathbf{F}$ is updated, the new combination would satisfy (10) again. Since the former combination may use more transmit power than it should have, the latter can lead to worse performance. In other words, the MSE is not necessarily a monotonically decreasing sequence, which may compromise the optimality of the final solution. This problem does not exist for Algorithm 2 because the power constraints are satisfied throughout the whole process. One possible disadvantage of Algorithm 2 is that the optimization of $\mathbf{Q}$ relies upon the GD method, whose parameters may need fine tuning to achieve the optimal speed of convergence.

V. Interblock Adaptation

Optimal transceiver design generally requires knowledge of the underlying wireless channels. In practice, the entire transmission period is divided into blocks, and in each block, the channel coefficients are estimated and then used for transceiver optimization, followed by the actual transmission of payload data, as shown in Fig. 2. The block length is generally selected to be smaller than the coherence time, such that the channel coefficients remain almost constant within each block. Otherwise, the optimal transceiver design would become outdated within the current block, and this model mismatch would deteriorate the performance significantly [2]. This implies that both the channel and the corresponding optimal transceiver change slowly across successive blocks. This important aspect of the wireless channels and its implications on the design of relay transceiver has been overlooked in previous development and evaluation of iterative algorithms. Indeed, all the previous works are confined to an individual transmission block. The proper exploitation of this property may simplify algorithms that are otherwise complex when the neighboring blocks are viewed in isolation.

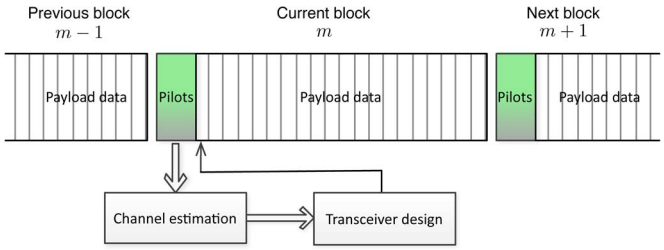


Fig. 2. Anatomy of a transmission block.

To exploit the slow variation between successive blocks, we propose an interblock adaptive implementation. For convenience, we introduce a block index m to differentiate these blocks. Under the given assumption of slow-fading channels, the optimal equalizers satisfy the following mathematical relationship:

$$\mathbf{Q}^\star[m] = \mathbf{Q}^\star[m-1] + \triangle\mathbf{Q}[m] \tag{54}$$

where $\|\triangle\mathbf{Q}[m]\|_F \leq \epsilon$, and $\epsilon > 0$ is a small positive number. Hence, a straightforward approach to interblock adaptation is to set the initial search point in Algorithm 1 to the optimal equalizer for the *previous* transmission block, *viz.*,

$$\mathbf{Q}_0[m] \Longleftarrow \mathbf{Q}^\star[m-1]. \tag{55}$$

As it will be shown in Section VII-B, interblock adaptation speeds up convergence of the iterative optimization process significantly. Faster convergence leads to much lower computational complexity because for any iterative algorithm, the complexity is proportional to the number of steps needed for convergence.

The fact that $\mathbf{Q}[m]$ does not show up in any power constraint is extremely important. Otherwise, the initial point $\mathbf{Q}_0[m]$ and the following iterates would not satisfy the constraints and, therefore, would not be eligible solutions. If we had optimized $\mathbf{Q}[m]$ in the first place, both $\mathbf{B}[m]$ and $\mathbf{F}[m]$ would be involved in their corresponding power constraints (9) and (10), making it impossible to employ the above interblock implementation. This improvement is possible only if the iterative algorithms are designed with interblock adaptation in mind in the first place.

VI. Further Comments on Typical Topologies

Here, we make further comments on the proposed unified framework within the context of a few typical relaying topologies (cf. Table I). These topologies may differ in terms of channel state information (CSI) availability and the numbers of sources, relays, or destinations.

1) Point-to-Point Topology (1S-1R-1D): The optimal MMSE relaying matrix for the 1S-1R-1D topology is in the well-known SVD form [7], [8]. The downside of this solution is the need to perform SVDs on the channel matrices for *every* transmission block. We note that computing the SVD of a matrix from that of another matrix when the two matrices differ by a small perturbation is not much easier than from scratch [51]. Consequently, with the help of interblock adaptation, the proposed numerical algorithm such as Algorithm 2 becomes competitive with the closed-form SVD approach in terms of BER performance and complexity.

2) MAC (MS-1R-1D): The SVD-based transceiver for 1S-1R-1D is also applicable to the MS-1R-1D topology with single-antenna source users or with multiple-antenna source users but without source CSI. For the former case, as discussed before, precoder $\mathbf{B}$ has to be *diagonal*, and each user should transmit the maximum allowable power. For the latter case, although the source users may use precoders, they cannot rely on the knowledge of the channels, and hence, the block-diagonal precoding matrix $\mathbf{B}$ must be predetermined.

3) BC (1S-1R-MD): The MMSE design for BC was considered in [19] with single-antenna users and in [20], [23], [25], and [27] with multiantenna users. The latter case usually requires complex algorithms that iterate multiple times through the precoder, the relaying matrix, and every equalizer. Our proposed general framework can be advantageously applied to these two scenarios. For single-antenna users, different users may apply their own amplitude scaling and phase rotation before decoding. This *diagonal scaling* scheme provides more flexibility than the design in [19], which assumes the same scaling for these users ($\mathbf{Q} = \mathbf{I}$). For this particular case when $\mathbf{Q} = \mathbf{I}$, our approach provides a closed-form expression for the optimal solution, whereas $\mathbf{B}$ and $\mathbf{F}$ were designed using an alternating algorithm in [19]. The optimal design for multi-antenna users is similar to that for single-antenna users, except that $\mathbf{Q}$ is block diagonal, and hence, the gradient has a different mathematical form as in (47). As discussed in Section IV-C, our approach should perform slightly better than previous iterative methods in [18], [20], and [21] and converges much faster.

4) IC (MS-1R-MD): In interference networks, multiple sources communicate with their intended destinations through a relay station. This network topology is more likely to be found in ad hoc networks than in cellular systems. The source users probably do not have access to CSI due to the communication overhead caused by CSI feedback. Therefore, precoder $\mathbf{B}$ does not need to be optimized, and the joint design of $\mathbf{F}$ and $\mathbf{Q}$ can be done using Algorithm 1.

5) Multirelay Topologies: In [43], we derived a simple closed-form expression for $\mathbf{F}$ for fixed equalizer $\mathbf{Q}$ and proposed two preliminary algorithms for the joint design, but did not consider the successive transmission blocks simultaneously. In this paper, we can start from the conclusions in Section III-C and derive a GD algorithm similar to Algorithm 1. The use of interblock adaptation will lead to much lower complexity than the algorithms in [43]. Moreover, this method applies to the 1S-MR-MD topology as well, in which case $\mathbf{Q}$ is block diagonal, and the gradient becomes the Hadamard product of the gradient for 1S-MD-1D and $\mathbf{I}_{BD}$ [similar to (47)].

TABLE II
Parameters of Fading Channels

Doppler Spectrum	Jakes
Rician K-factor	4
Sampling interval	0.5μs
Sampling frequency	2MHz
Carrier frequency	2GHz
Doppler spread	$f_d = 100$Hz (velocity of 15m/s)
Approximate coherence time	$0.5/f_d$ =0.005s
Delay spread	NA (flat fading)
Spatial correlation	0.7
Transmission block	1/20 $\times$ coherence time

VII. Numerical Results

In Section VII-A and B, we study the convergence and tracking behaviors of the proposed numerical algorithms, followed by performance comparison of representative topologies in Section VII-C. For convenience, we define two signal-to-noise ratio (SNR), i.e., $\rho_1 = P_s/(N_S\sigma_w^2)$ and $\rho_2 = P_r/(N_R\sigma_n^2)$, that represent the link quality of the first-hop and second-hop transmissions, respectively. The following parameters are chosen throughout this section: $N_B = N_S = N_R = N_D = 4$, $\mathbf{R}_w = \sigma_w^2\mathbf{I}$, $\mathbf{R}_n = \sigma_n^2\mathbf{I}$, and $\sigma_w^2 = \sigma_n^2$.

We assume slow- and flat-fading MIMO wireless channels whose temporal characteristics of the channels are summarized in Table II. The spatial characteristics of the channel between the base station (BS) and the relay is described by 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 $0.7^{|i-j|}$. Similarly, the forward channel is $\mathbf{G} = \mathbf{R}_d^{1/2}\mathbf{G}_w\mathbf{R}_r^{1/2}$, where $\mathbf{G}_w$ has the same statistical characteristics as $\mathbf{H}_w$, and the (i,j)th entry of $\mathbf{R}_d$ is $0.7^{|i-j|}$. For 1S-1R-MD with single-antenna destinations, $\mathbf{R}_d$ is a diagonal matrix representing relative pathloss of 0, 0, 3, and 6 dB for different users.

A. Joint Optimization

We generate a typical channel realization and compare the speed of convergence between the two approaches mentioned in Section IV-C: Algorithm 2 and the conventional approach. The former only updates $\mathbf{B}$ and $\mathbf{Q}$ alternatively, whereas the latter iterates through all three matrices $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{Q}$. We assume that $\mathbf{Q}$ is diagonal, and the initial values are $\mathbf{B}_0 \propto \mathbf{I}$, $\mathbf{Q}_0 \propto \mathbf{I}$, and $\mathbf{F}_0 \propto \mathbf{I}$. As shown in Fig. 3, by removing $\mathbf{F}$ from

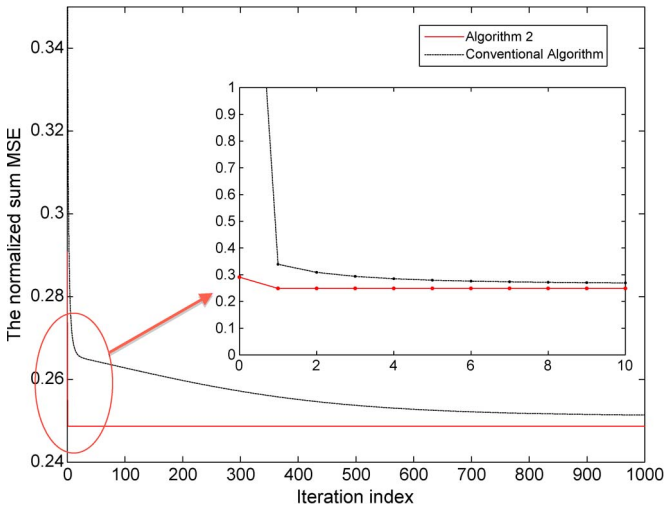


Fig. 3. Convergence behaviors of joint design.

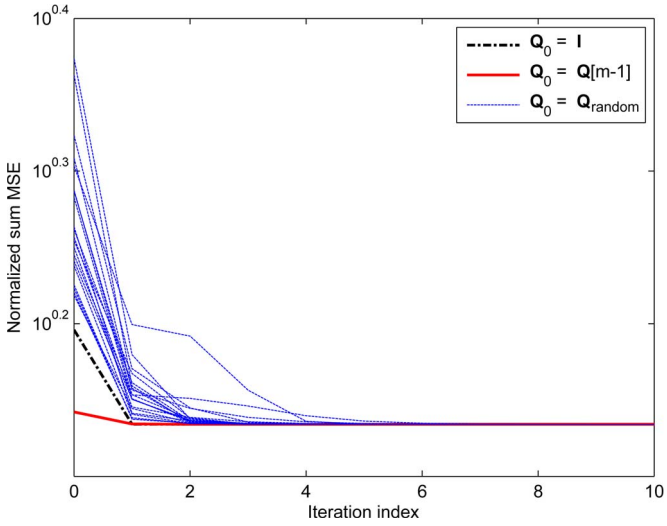


Fig. 4. Comparison of different initializations of $\mathbf{Q}_0$.

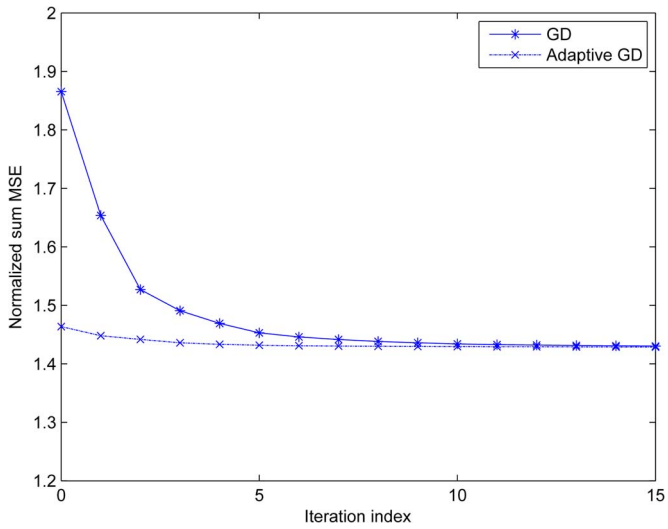


Fig. 5. Speed of convergence of Algorithm 1 when $\mathbf{Q}$ is structurally unconstrained.

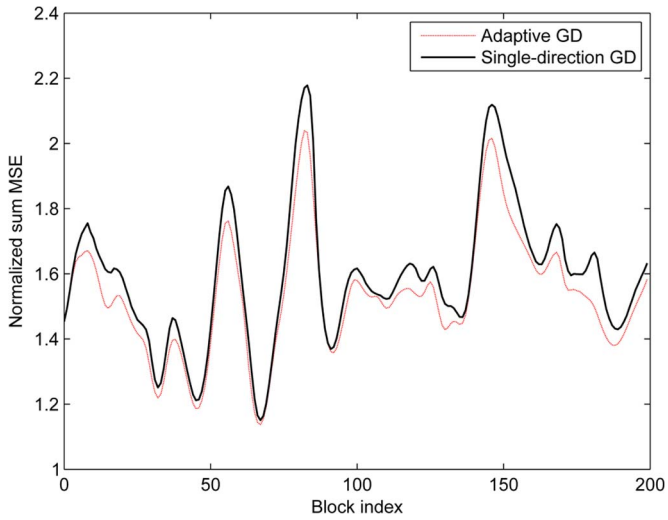


Fig. 6. Tracking behaviors of the proposed adaptive algorithms when $\mathbf{Q}$ is structurally unconstrained.

the iterating process, Algorithm 2 converges much faster than the conventional algorithm, in fact after just one iteration. A possible explanation is that at high SNRs, the third term on the right-hand side of (29) is negligible. Hence, it follows that the optimal $\mathbf{B}^\star$ would almost be independent of $\mathbf{Q}^\star$, and vice versa.

As pointed out in Section IV-C, different choices of $\mathbf{Q}_0$ in Algorithm 2 may lead to performance difference. Fig. 4 shows the convergence behaviors of 20 random choices of $\mathbf{Q}_0$ (i.e., diagonal matrices with independent and identically distributed diagonal entries taken from standard normal distribution). These random initial points almost lead to the same performance, which was also observed in [27] for the algorithms therein.

B. Interblock Adaptation

Next, we demonstrate how much interblock adaptation can improve the speed of convergence for Algorithm 1, that is, the GD algorithm.

Structurally Unconstrained $\mathbf{Q}$*:* Assume $\rho_1 = \rho_2 = 15$ dB and $\mathbf{B} \propto \mathbf{I}$. The parameters in Algorithm 1 are chosen as $\rho = 0.9$ and $c = 0.01$. In Fig. 5, we compare the speed of convergence between nonadaptive and adaptive versions of GD for a typical channel realization. As explained earlier, the former version searches from $\mathbf{I}$, whereas the latter version searches from the optimal $\mathbf{Q}^\star[m-1]$ for the previous transmission block. As shown by the curves, this interblock adaptation speeds up convergence considerably. The computational complexity is therefore reduced because it is proportional to the required number of search directions.

Next, we consider a low-complexity version of the adaptive GD algorithm, which we call single-direction GD. For each transmission block (and hence channel realization), this algorithm computes the gradient only once and then carries out a line search along its opposite direction for a suboptimal solution. This algorithm is compared with the standard adaptive GD algorithm, which searches along multiple directions for each transmission block. As shown in Fig. 6, single-direction GD is able to track GD closely as time evolves, although with slight SNR loss (within 0.6 dB).

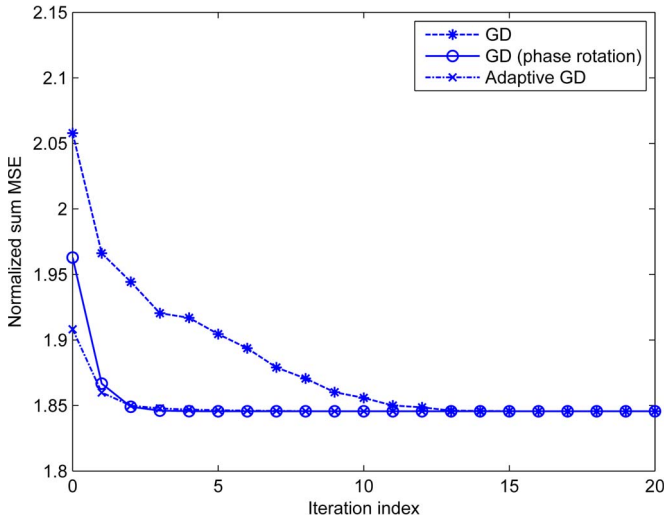


Fig. 7. Speed of convergence of Algorithm 1 with diagonal $\mathbf{Q}$.

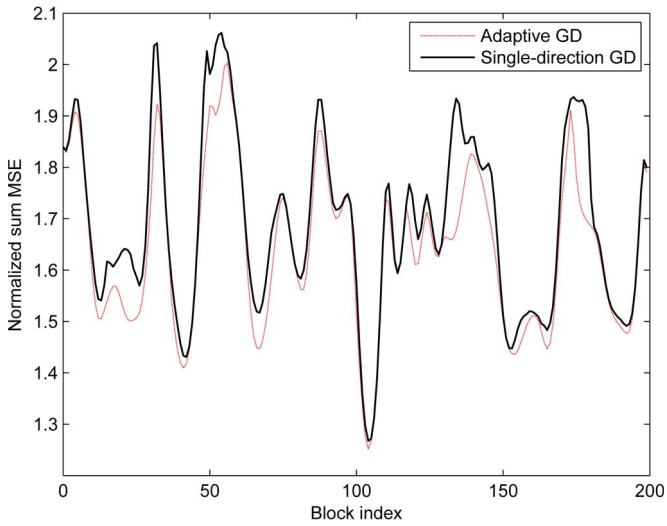


Fig. 8. Tracking behaviors of the proposed algorithms when $\mathbf{Q}$ is diagonal.

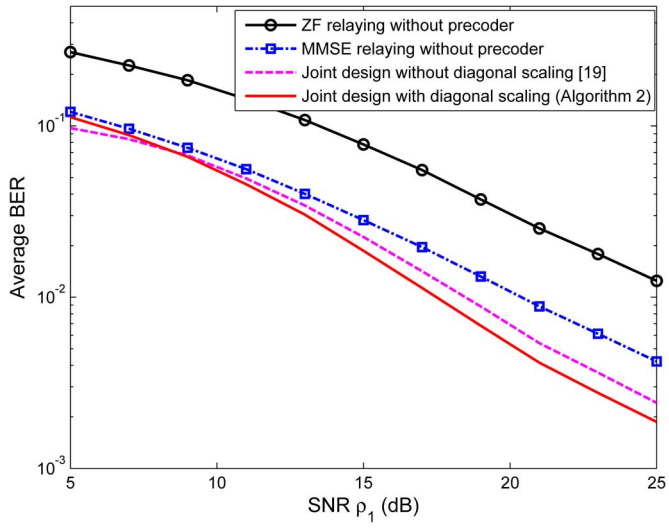


Fig. 9. BER versus ρ_1 with $\rho_2 = 25$ dB for single-antenna users.

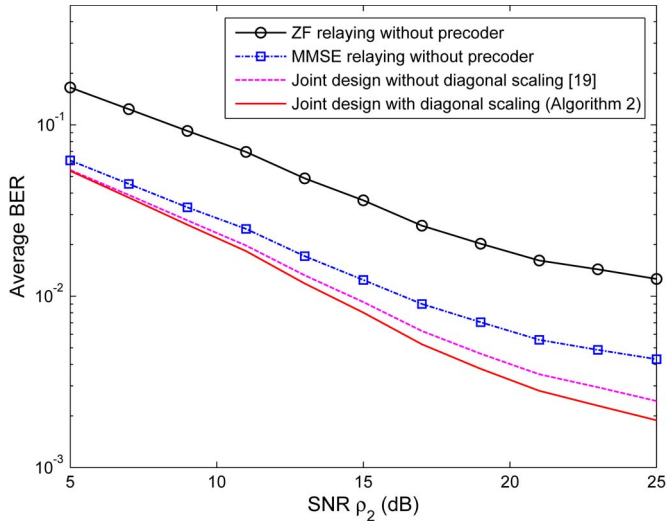


Fig. 10. BER versus ρ_2 with $\rho_1 = 25$ dB for single-antenna users.

Diagonal $\mathbf{Q}$*:* We choose the same settings as in the previous scenario. In addition to the original and adaptive versions, we consider searching from the phase rotation matrix in (48) as well. As shown in Fig. 7, the phase rotation version converges considerably faster and is suitable for the first transmission block. The adaptive version is still the best candidate for the following blocks. As in the previous scenario, single-direction GD tracks adaptive GD very closely, as shown in Fig. 8.

C. Performance Comparison

We evaluate the BER performance of two different relay network topologies by simulation. We first consider a 1S-1R-4D BC system, where the BS transmits four independent uncoded quaternary phase-shift keying symbol streams to their corresponding *single-antenna* destination users. The methods under comparison are as follows.

- Zero-forcing (ZF) relaying [52]: $\mathbf{B} = \sqrt{P_S/N_S}\mathbf{I}$, $\mathbf{F} = \eta\mathbf{G}^\dagger\mathbf{H}^\dagger$, and $\mathbf{Q} \propto \mathbf{I}$.
- MMSE relaying without precoder [11]: $\mathbf{B} = \sqrt{P_S/N_S}\mathbf{I}$, $\mathbf{F}$ as in (23), and $\mathbf{Q} \propto \mathbf{I}$.
- Joint design of $\mathbf{B}$ and $\mathbf{F}$ without diagonal scaling ($\mathbf{Q} \propto \mathbf{I}$) [19].
- Joint design of $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{Q}$ with diagonal scaling (Algorithm 2).

The average BERs of the multiple users are shown in Figs. 9 and 10, where we let $\rho_1 = 25$ dB or $\rho_2 = 25$ dB and vary the other SNR between 5 and 25 dB. 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 2–3 dB over the MMSE relaying without precoder. Furthermore, by including the diagonal $\mathbf{Q}$ in the joint design, the diagonal scaling scheme enables an additional SNR gain of 0.5–2 dB at mid-to-high SNRs.

We next consider a 1S-1R-2D system in which each destination user has two antennas. Algorithm 2 is compared with the iterative method in [27]. The BER curves in Fig. 11 show a 1–2-dB SNR gain for the proposed algorithm. This is consistent with our analysis in Section IV-C.

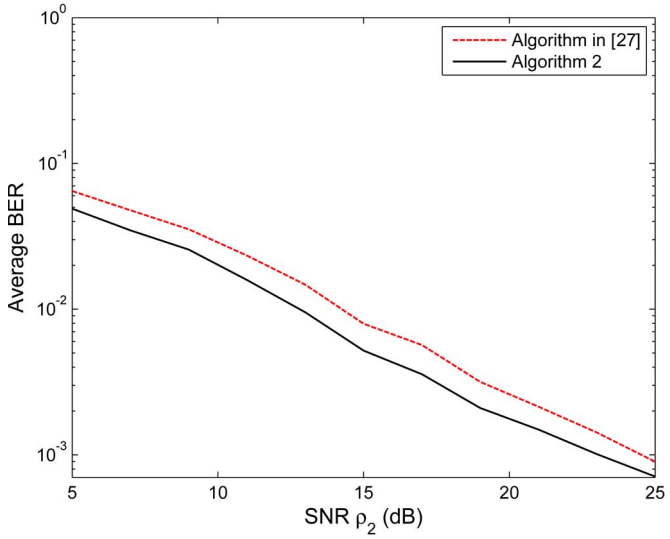


Fig. 11. BER versus ρ_2 with $\rho_1 = 25$ dB for multiantenna users.

VIII. Conclusion

In this paper, we have proposed a unified transceiver optimization framework for different MIMO relaying topologies. First, we formulated a generic model that can accommodate these topologies by imposing proper structural constraints on the precoder, the relaying matrix, and the equalizer. Then, we considered the joint design of these modules based on the MMSE criterion. More specifically, the optimal relaying matrix was derived as a closed-form expression in terms of the other two matrices. This is the common step for point-to-point and multiuser systems. Subsequently, we studied optimization of either the precoder or the equalizer under different structural constraints and proposed an alternating algorithm (see Algorithm 2) for the joint design. This algorithm allowed us to choose the optimal equalizer from the previous block as the initial search point for the current block, which was shown to speed up convergence and, henceforth, to reduce computational complexity significantly. We explained how this framework can accommodate different relaying topologies and validated the performance of the resulting designs through simulations over fading channels. In many instances, this unified framework led to new algorithms with lower complexity or better performance.

Appendix
Proof of Theorem 1

For any $\eta^{-1}\mathbf{Q}$ with $\eta > 0$, the optimal relaying matrix is obtained by replacing $\mathbf{Q}$ with $\eta^{-1}\mathbf{Q}$ in (21). Henceforth, $\mathbf{F}^\star$ and the corresponding MMSE are functions of η and $\lambda^\star$. In turn, $\lambda^\star$ is an implicit piecewise function of η, whose critical point $\eta_c > 0$ satisfies

$$P_r = \mathrm{tr}(\mathbf{F}^\star \mathbf{R}_x \mathbf{F}^{\star H})|_{\lambda^\star = 0}$$
$$= \eta_c^2 \mathrm{tr}\left(\mathbf{B}^H \mathbf{H}^H \mathbf{R}_x^{-1} \mathbf{H} \mathbf{B} \mathbf{Q} \mathbf{G} \left((\mathbf{G}^H \mathbf{Q}^H \mathbf{Q} \mathbf{G})^\dagger \right)^2 \mathbf{G}^H \mathbf{Q}^H \right).$$

As a result, the MMSE itself is also a piecewise function of η. The next step is to find the particular $\eta^\star$ leading to the smallest MMSE.

If $\eta \le \eta_c$, $\lambda^\star = 0$ is the solution to the KKT conditions. Substituting $\eta^{-1}\mathbf{Q}$ into the MMSE in (22), we found that only the third term, i.e., $\eta^{-2}\mathrm{tr}(\mathbf{Q}\mathbf{R}_n\mathbf{Q}^H)$, depends on η, and hence, the MMSE is a decreasing function of η in the interval $(0, \eta_c]$. Therefore, the smallest MMSE must be achieved when $\eta \ge \eta_c$.

If $\eta \ge \eta_c$, $\lambda^\star$ has to satisfy $P_r = \mathrm{tr}(\mathbf{F}^\star \mathbf{R}_x \mathbf{F}^{\star H})$. Using this equality to replace η^{-2} in the third term of (22) and P_r in the last term, we have

$$\mathrm{MSE}_{\min}(\eta) = \mathrm{tr}(\mathbf{I}) - \mathrm{tr}\left(\mathbf{Q}\mathbf{G} (\mathbf{G}^H \mathbf{Q}^H \mathbf{Q} \mathbf{G} + \lambda^\star \eta^2 \mathbf{I})^{-2} \times \left(\mathbf{G}^H \mathbf{Q}^H \mathbf{Q} \mathbf{G} + (2\lambda^\star \eta^2 - \theta)\mathbf{I} \right) \mathbf{G}^H \mathbf{Q}^H \mathbf{B}^H \mathbf{H}^H \mathbf{R}_x^{-1} \mathbf{H} \mathbf{B} \right) \tag{56}$$

which becomes a function of $\gamma \triangleq \lambda^\star \eta^2$. Define the SVD

$$\mathbf{Q}\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^H.$$

The problem is then reduced to that of maximizing

$$g(\gamma) = \mathrm{tr}\left(\mathbf{S} (\mathbf{S}^H \mathbf{S} + \gamma \mathbf{I})^{-2} \left(\mathbf{S}^H \mathbf{S} + (2\gamma - \theta)\mathbf{I} \right)^{-1} \mathbf{S}^H \mathbf{A} \right) \tag{57}$$

where $\mathbf{A} \triangleq \mathbf{U}^H \mathbf{B}^H \mathbf{H}^H \mathbf{R}_x^{-1} \mathbf{H} \mathbf{B} \mathbf{U}$. Define $\beta_k \ge 0$ as the kth diagonal entry of $\mathbf{S}\mathbf{S}^H$ and $a_k \ge 0$ as the (k,k)th entry of $\mathbf{A}$, we have

$$g(\gamma) = \sum_{k=1}^{N_D} a_k \beta_k \frac{2\gamma - \theta + \beta_k}{(\gamma + \beta_k)^2} \tag{58}$$

whose first-order derivative is

$$\frac{\mathrm{d}g}{\mathrm{d}\gamma} = \sum_{k=1}^{N_D} 2 a_k \beta_k \frac{\theta - \gamma}{(\gamma + \beta_k)^3} = 0. \tag{59}$$

If $0 < \gamma < \theta$, this derivative is positive, and $g(\gamma)$ is monotonically increasing. If $\gamma > \theta$, the derivative is negative, and $g(\gamma)$ is monotonically decreasing. Therefore, $\gamma^\star = \theta$ is the unique solution to maximize $g(\gamma)$. This corresponds to a particular combination of $\eta^\star$ and $\lambda^\star$. The optimal relaying matrix in (23) can be obtained by replacing $\mathbf{Q}$ with $\mathbf{Q}/\eta^\star$ in (21). Since $\eta \ge \eta_c$, the power constraint is tightly satisfied, and therefore, $\eta^\star$ satisfies $\mathrm{tr}(\mathbf{F}^\star \mathbf{R}_x \mathbf{F}^{\star H}) = P_r$. Substituting $\gamma^\star = \lambda^\star \eta^{\star 2} = \theta$ into (56) leads to (24).

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, no. 1, pp. 618 787-1–618 787-14, Jul. 2009.

[2] L. Sanguinetti, A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sep. 2012.

[3] Y. Rong, "Multihop nonregenerative MIMO relays QoS considerations," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 290–303, Jan. 2011.

[4] L. Sanguinetti and A. D'Ámico, "Power allocation in two-hop amplify-and-forward MIMO relay systems with QoS requirements," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2494–2507, May 2012.

[5] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, Jun. 2007.

[6] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

[7] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, Jul. 2008.

[8] 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, no. 12, pp. 4837–4851, Dec. 2009.

[9] C. Song, K.-J. Lee, and I. Lee, "MMSE based transceiver designs in closed-loop non-regenerative MIMO relaying systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2310–2319, Jul. 2010.

[10] C. Xing, S. Ma, and Y.-C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[11] O. Oyman and A. J. Paulraj, "Design and analysis of linear distributed MIMO relaying algorithms," *Proc. Inst. Elect. Eng.—Commun.*, vol. 153, no. 4, pp. 565–572, Apr. 2006.

[12] H. Shi, T. Abe, T. Asai, and H. Yoshino, "A relaying scheme using QR decomposition with phase control for MIMO wireless networks," in *Proc. IEEE Int. Conf. Commun.*, Seoul, Korea, May 2005, vol. 4, pp. 2705–2711.

[13] 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, no. 2, pp. 727–738, Feb. 2008.

[14] H. Kim, S. Lee, K. Kwak, H. Min, and D. Hong, "On the design of ZF and MMSE Tomlinson–Harashima precoding in multiuser MIMO amplify-and-forward relay system," in *Proc. 20th IEEE PIMRC*, Tokyo, Japan, Sep. 2009, pp. 2509–2513.

[15] W. Liu, C. Li, J.-D. Li, and L. Hanzo, "Block diagonalisation-based multiuser multiple input multiple output-aided downlink relaying," *IET Commun.*, vol. 6, no. 15, pp. 2371–2377, Oct. 2012.

[16] C. Zhai, X. Li, and Y. Hei, "A novel decomposed transceiver design for multiuser MIMO relay downlink systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Paris, France, Apr. 2012, pp. 1921–1924.

[17] C. Zhao and B. Champagne, "A low-complexity hybrid framework for combining-type non-regenerative MIMO relaying," *Wireless Pers. Commun.*, vol. 72, no. 1, pp. 635–652, Sep. 2013.

[18] S. Jang, J. Yang, and D.-K. Kim, "Minimum MSE design for multiuser MIMO relay," *IEEE Commun. Lett.*, vol. 14, no. 9, pp. 812–814, Sep. 2010.

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

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

[21] W. Xu, X. Dong, and W.-S. Lu, "Joint optimization for source and relay precoding under multiuser MIMO downlink channels," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 2010, pp. 1–5.

[22] M. Khandaker and Y. Rong, "Joint source and relay optimization for multiuser MIMO relay communication systems," in *Proc. 4th Int. Conf. Signal Process. Commun. Syst.*, Gold Coast, Australia, Dec. 2010, pp. 1–6.

[23] M. Khandaker and Y. Rong, "Joint transceiver optimization for multiuser MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5977–5986, Nov. 2012.

[24] W. Xu, X. Dong, and W.-S. Lu, "MIMO relaying broadcast channels with linear precoding and quantized channel state information feedback," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5233–5245, Oct. 2010.

[25] 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.

[26] 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.

[27] 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, Nov. 2012.

[28] G. Okeke, W. Krzymien, and Y. Jing, "Beamforming in non-regenerative MIMO broadcast relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6641–6654, Dec. 2012.

[29] C. Zhao and B. Champagne, "Linear transceiver design for relay-assisted broadcast systems with diagonal scaling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 4844–4848.

[30] Y. Fu, L. Yang, and W.-P. Zhu, "A nearly optimal amplify-and-forward relaying scheme for two-hop MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 14, no. 3, pp. 229–231, Mar. 2010.

[31] K.-J. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.

[32] Y. Rong, "Joint source and relay optimization for two-way MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1329–1331, Dec. 2011.

[33] Y. Izi and A. Falahati, "Amplify-forward relaying for multiple-antenna multiple relay networks under individual power constraint at each relay," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 50-1–50-10, Feb. 2012.

[34] H.-J. Choi, K.-J. Lee, C. Song, H. Song, and I. Lee, "Weighted sum rate maximization for multiuser multirelay MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 885–889, Feb. 2013.

[35] C. Xing, S. Li, Z. Fei, and J. Kuang, "How to understand linear minimum mean-square-error transceiver design for multiple-input–multiple-output systems from quadratic matrix programming," *IET Commun.*, vol. 7, no. 12, pp. 1231–1242, Aug. 2013.

[36] 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.

[37] 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, no. 5, pp. 1228–1235, May 2011.

[38] A. El-Keyi and B. Champagne, "Cooperative MIMO-beamforming for multiuser relay networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, NV, USA, Mar./Apr. 2008, pp. 2749–2752.

[39] B. K. Chalise, L. Vandendorpe, and J. Louveaux, "MIMO relaying for multi-point to multi-point communication in wireless networks," in *Proc. 2nd IEEE Int. Workshop CAMPSAP*, St. Thomas, VI, USA, Dec. 2007, pp. 217–220.

[40] B. K. Chalise and L. Vandendorpe, "MIMO relay design for multipoint-to-multipoint communications with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2785–2796, Jul. 2009.

[41] B. Chalise and L. Vandendorpe, "Optimization of MIMO relays for multipoint-to-multipoint communications: Nonrobust and robust designs," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6355–6368, Dec. 2010.

[42] A. Phan, H. Tuan, H. Kha, and H. Nguyen, "Beamforming optimization in multi-user amplify-and-forward wireless relay networks," *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 1510–1520, Apr. 2012.

[43] C. Zhao and B. Champagne, "Joint design of multiple non-regenerative MIMO relaying matrices with power constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4861–4873, Oct. 2013.

[44] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx–Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[45] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[46] L. Sanguinetti and M. Morelli, "Non-linear pre-coding for multiple-antenna multi-user downlink transmissions with different QoS requirements," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 852–856, Mar. 2007.

[47] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

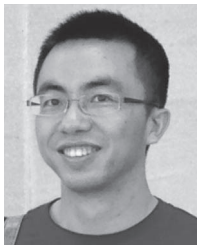
[48] J. Nocedal and S. Wright, *Numerical Optimization,* 2nd ed. New York, NY, USA: Springer Science+Business Media, LLC, 2006.

[49] Y. Ye, "Approximating quadratic programming with bound and quadratic constraints," *Math. Programm.*, vol. 84, no. 2, pp. 219–226, Feb. 1999.

[50] S. Ma *et al.*, "Iterative transceiver design for MIMO AF relay networks with multiple sources," in *Proc. Mil. Commun. Conf.*, Nov. 2010, pp. 369–374.

[51] P. I. Davies and M. I. Smith, "Updating the singular value decomposition," *J. Comput. Appl. Math.*, vol. 170, no. 1, pp. 145–167, Sep. 2004.

[52] R. Louie, Y. Li, and B. Vucetic, "Zero forcing in general two-hop relay networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 191–202, Jan. 2010.



Chao Zhao received the B.E. degree in electronics information engineering and the M.S. degree in information and communication engineering from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, and the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 2014.

He is currently an Engineer with Microsemi Corporation, Ottawa, ON, Canada. His research interests include array signal processing, multiple-input–multiple-output relay communications, computer architectures, network synchronization, and timing.

Benoît Champagne (SM'03) received the B.Ing. degree in engineering physics from the École Polytechnique de Montréal, Montréal, QC, Canada, in 1983; the M.Sc. degree in physics from the Université de Montréal, in 1985; and the Ph.D. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1990.

From 1990 to 1999, he was an Assistant Professor and then an Associate Professor with the Institut National de la Recherche Scientifique—Telecommunications, Université du Quebec, Montréal. In 1999, he joined McGill University, Montréal, where he is currently a Full Professor with the Department of Electrical and Computer Engineering. From 2004 to 2007, he was an Associate Chairman of Graduate Studies with the Department. His research focuses on the study of advanced algorithms for the processing of information-bearing signals by digital means. His interests span many areas of statistical signal processing, including detection and estimation, sensor array processing, adaptive filtering, and applications thereof to broadband communications and audio processing, where he has coauthored more than 200 referred publications. His research has been funded by the Natural Sciences and Engineering Research Council of Canada, the "Fonds de Recherche sur la Nature et les Technologies" from the Government of Quebec, Prompt Quebec, as well as some major industrial sponsors, including Nortel Networks, Bell Canada, InterDigital, and Microsemi.

Dr. Champagne has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the EURASIP *Journal on Applied Signal Processing*. He has also served on the Technical Committees of several international conferences in the fields of communications and signal processing. In particular, he was a Cochair, Wide Area Cellular Communications Track, for the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (Toronto, ON, Canada, September 2011), a Cochair, Antenna and Propagation Track, for the IEEE Vehicular Technology Conference—Fall (Los Angeles, CA, USA, September 2004), and a Registration Chair for the IEEE International Conference on Acoustics, Speech, and Signal Processing (Montréal, QC, Canada, May 2004).