

Tomlinson Harashima Precoding Design for Non-regenerative MIMO Relay Networks

Andrew P. Millar, Stephan Weiss, Robert W. Stewart
Department of Electronic & Electrical Engineering
University of Strathclyde
Glasgow, G1 1XW, Scotland, UK
Email: (andrew.millar, stephan, r.stewart)@eee.strath.ac.uk

Abstract—In this paper we consider the design of minimum mean square error (MMSE) transceivers for non-regenerative multiple input multiple output (MIMO) relay systems. Our design utilises Tomlinson Harashima precoding (THP) at the source along with linear processors in each stage of the network. Assuming full channel state information (CSI) is available at each node in the network the various processors are jointly optimised to minimise the system arithmetic mean square error (MSE) whilst abiding by average power constraints at both the source and relay terminals in the network. Simulations show that the proposed schemes outperform existing methods in terms of bit error ratio (BER).

Index Terms—MIMO, non-regenerative relay, transceiver design, Tomlinson Harashima precoding, minimum mean square error

I. INTRODUCTION

It is well known that relaying techniques can extend network coverage [1], increase channel capacity [2], and improve link reliability due to the spatial diversity offered by the relay nodes. When the source, relay, and destination devices are equipped with multiple antennas the communication system is referred to as a MIMO relay network. Due to the various advantages offered by MIMO relaying it is considered an integral component in the design of future generation wireless communication systems.

Relays can be classed as either decode and forward (DF) or amplify and forward (AF) also commonly known as regenerative and non-regenerative relaying respectively. In the DF protocol the relay decodes the received signal streams and then transmits the regenerated symbols to the next node in the network. For the AF scheme, which is the simpler of the two protocols, the relay simply amplifies the data received from the source and then transmits to the destination device.

Most of the works in the area of transceiver design for MIMO relay networks have focussed on the design of linear processors to enhance system performance under the assumption that each node in the network has the required CSI of the source to relay and relay to destination channels. In [1] the optimal relay precoder is designed to maximise the mutual information (MI) between the source and destination where the source precoder is designed to be a scaled identity matrix. The authors in [3] also focus on maximising the MI but introduce linear equalisation at the destination using the Wiener filter.

Different objective functions other than the maximisation of MI have also been well investigated such as the MMSE design criterion. In [4] the optimal relay precoder is derived that minimises the MSE where similar to [1] and [3] the power is allocated uniformly over the source antennas.

A unified framework for the design of two-hop MIMO relay systems based on majorisation theory is presented in [5] where the optimal source and relay precoders are derived for a broad range of different design criteria. It is shown that for Schur concave objective functions the optimal source and relay precoders jointly diagonalise the overall communication system and convert the MIMO relay channel into independent single input single output (SISO) subchannels. For Schur convex functions the communication process is only diagonalised up to a very specific rotation of the data symbols. This work was later extended to the case of multi-hop MIMO relaying in [6] where it is shown that the source and relay precoders have the same optimal structures as in [5].

In this work we focus on the joint design of linear processors for a two-hop network with THP employed at the source. THP transceiver designs have been well investigated for the case of point to point MIMO networks in e.g. [7] and [8]. As in many works e.g. [4], [5], and [6] we assume that the direct link between the source and destination antennas is negligible and that each node in the network has full CSI of the source to relay and relay to destination channels.

The remainder of the paper is organised as follows: In section II we introduce the signal model for the THP system under consideration. Section III presents the optimal MMSE THP transceiver design and a numerical example is provided in section IV. Finally conclusions are drawn in V.

Notation: In our notation vectors and matrices are denoted by lower and upper case bold font respectively. The sets of real and complex numbers are \mathbb{R} and \mathbb{C} , which in the case of vector and matrix quantities indicate dimensions by means of a superscript. The operators $\xi\{\cdot\}$, $\text{tr}\{\cdot\}$, $(\cdot)^H$, $(\cdot)^{-1}$, and $|\cdot|$ denote the expectation, trace, hermitian transpose, inverse, and determinant respectively. \mathbf{I}_m is the $m \times m$ identity matrix. The element in the i_{th} row and j_{th} column of matrix \mathbf{A} is denoted as a_{ij} and the i_{th} element of vector \mathbf{a} is denoted as a_i . Matrix rank is noted by $\text{rank}(\cdot)$ and $\text{diag}\{a_{11}, a_{22}, \dots, a_{NN}\}$ denotes a diagonal matrix with diagonal entries $\{a_{11}, a_{22}, \dots, a_{NN}\}$. We define the operator $[x]^+ \triangleq \max(x, 0)$.

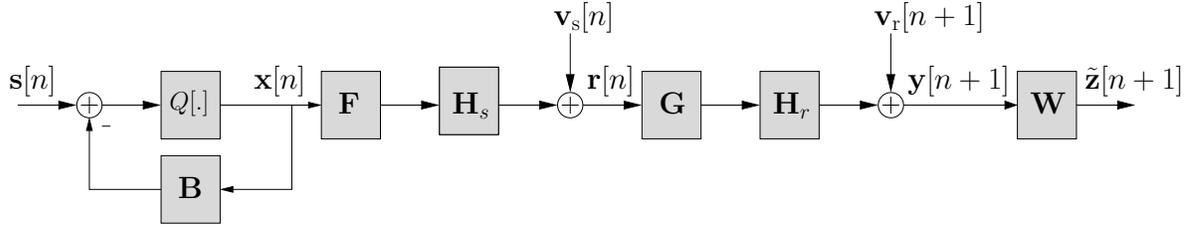


Fig. 1. MIMO relay system model with Tomlinson Harashima Precoding.

II. THP SYSTEM MODEL

In this section we introduce the signal model for a half-duplex two-hop MIMO relay network with THP shown in Fig. 1. We assume a system where the source, relay, and destination nodes have N_s , N_r , and N_d antennas respectively and there is no direct link between the source and destination. To ensure N_s independent data streams can be transmitted across the network we assume that $N_s \leq N_r$ and $N_s \leq N_d$. For convenience we also assume $N_d \leq N_r$.

The source symbols $\mathbf{s}[n] \in \mathbb{C}^{N_s}$ are drawn from M -ary quadrature amplitude modulated (QAM) constellations with covariance $\mathbf{R}_{ss} = \xi \{ \mathbf{s}[n] \mathbf{s}[n]^H \} = \sigma_s^2 \mathbf{I}_{N_s}$. The data symbols in $\mathbf{x}[n] \in \mathbb{C}^{N_s}$ are recursively computed according to

$$x_k[n] = Q \left[s_k[n] - \sum_{l=1}^{k-1} b_{kl} x_l[n] \right] \quad (1)$$

where $Q[\cdot]$ denotes the modulo operator and $\mathbf{B} \in \mathbb{C}^{N_s \times N_s}$ is a strictly lower left triangular matrix. It is shown in e.g. [7] and [8] that the operator in (1) is equivalent to the following operation

$$\mathbf{x}[n] = \mathbf{U}^{-1} \mathbf{z}[n] \quad (2)$$

where $\mathbf{U} \triangleq \mathbf{B} + \mathbf{I}_{N_s}$ is a lower left triangular matrix with unit diagonal elements and $\mathbf{z}[n] = \mathbf{s}[n] + \mathbf{d}[n]$ contains modified data symbols where $\mathbf{d}[n]$ is such that the real and imaginary components of $\mathbf{x}[n]$ are constrained to be within the region $(-\sqrt{M}, \sqrt{M})$. This leads to the symbols in $\mathbf{x}[n]$ having slightly higher energy than $\mathbf{s}[n]$. However for moderate to high M this energy increase can be neglected and $\xi \{ \mathbf{x}[n] \mathbf{x}[n]^H \} = \sigma_s^2 \mathbf{I}_{N_s}$ can be assumed [7], [8].

The symbols $\mathbf{x}[n]$ are then processed by the precoding matrix $\mathbf{F} \in \mathbb{C}^{N_s \times N_s}$ and transmitted across the source to relay channel $\mathbf{H}_s \in \mathbb{C}^{N_r \times N_s}$. The data vector $\mathbf{r}[n] \in \mathbb{C}^{N_r}$ received by the relay in the first time slot can thus be written as

$$\mathbf{r}[n] = \mathbf{H}_s \mathbf{F} \mathbf{x}[n] + \mathbf{v}_s[n] \quad (3)$$

where $\mathbf{v}_s[n] \in \mathbb{C}^{N_r}$ is an additive white Gaussian noise (AWGN) vector with covariance $\mathbf{R}_{v_s v_s} = \xi \{ \mathbf{v}_s[n] \mathbf{v}_s[n]^H \} = \sigma_{v_s}^2 \mathbf{I}_{N_r}$. In the second time slot the relay precodes the received data using $\mathbf{G} \in \mathbb{C}^{N_r \times N_r}$ and transmits over the second stage channel $\mathbf{H}_r \in \mathbb{C}^{N_d \times N_r}$ resulting in the received signal $\mathbf{y}[n+1] \in \mathbb{C}^{N_d}$ at the destination being

$$\mathbf{y}[n+1] = \mathbf{H}_r \mathbf{G} \mathbf{r}[n] + \mathbf{v}_r[n+1] \quad (4)$$

where $\mathbf{v}_r[n+1] \in \mathbb{C}^{N_d}$ contains AWGN samples with covariance $\mathbf{R}_{v_r v_r} = \xi \{ \mathbf{v}_r[n+1] \mathbf{v}_r[n+1]^H \} = \sigma_{v_r}^2 \mathbf{I}_{N_d}$. After processing by the equaliser matrix $\mathbf{W} \in \mathbb{C}^{N_s \times N_d}$ and using (3) and (4) we have

$$\tilde{\mathbf{z}}[n+1] = \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{x}[n] + \mathbf{W} \mathbf{v}[n+1] \quad (5)$$

where for convenience we define $\mathbf{H} \triangleq \mathbf{H}_r \mathbf{G} \mathbf{H}_s$ as the compound MIMO channel between the source and destination antennas and $\mathbf{v}[n+1] \triangleq \mathbf{H}_r \mathbf{G} \mathbf{v}_s[n] + \mathbf{v}_r[n+1]$ as the total noise term at the input to the equaliser with covariance $\mathbf{R}_{vv} = \xi \{ \mathbf{v}[n+1] \mathbf{v}[n+1]^H \} = \mathbf{H}_r \mathbf{G} \mathbf{R}_{v_s v_s} \mathbf{G}^H \mathbf{H}_r^H + \mathbf{R}_{v_r v_r}$. The modulo operator is then performed on the elements of $\tilde{\mathbf{z}}[n+1]$ to recover an estimate of the original transmit signals $\mathbf{s}[n]$. The error signal in terms of the modified symbols can be calculated as $\mathbf{e}[n+1] = \tilde{\mathbf{z}}[n+1] - \mathbf{z}[n]$ which using (2) and (5) results in

$$\mathbf{e}[n+1] = \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{x}[n] - \mathbf{U} \mathbf{x}[n] + \mathbf{W} \mathbf{v}[n+1] \quad (6)$$

where it is assumed that the modulo operator eliminates the effect of the periodic extension to the original symbol constellation [7].

III. MMSE TRANSCIVER DESIGN

In this section we derive the optimal THP processors that minimise the system MSE whilst abiding by average power constraints at both the source and relay nodes. We firstly formulate the constrained optimisation problem before presenting the optimal precoder structures for \mathbf{F} , \mathbf{G} , and \mathbf{U} . Finally an alternating algorithm is developed to allocate power at the source and relay terminals.

A. Optimal MMSE Receiver

Using the error in (6) we can write our error covariance matrix $\mathbf{R}_{ee} = \xi \{ \mathbf{e}[n+1] \mathbf{e}[n+1]^H \}$ as

$$\mathbf{R}_{ee} = (\mathbf{W} \mathbf{H} \mathbf{F} - \mathbf{U}) \mathbf{R}_{ss} (\mathbf{W} \mathbf{H} \mathbf{F} - \mathbf{U})^H + \mathbf{W} \mathbf{R}_{vv} \mathbf{W}^H. \quad (7)$$

The receiver matrix which minimises the MSE is the well known Wiener filter and is obtained by setting the derivative of (7) to zero and solving for \mathbf{W} giving the optimal equaliser solution as

$$\mathbf{W} = \mathbf{U} \mathbf{R}_{ss} \mathbf{F}^H \mathbf{H}^H (\mathbf{H} \mathbf{F} \mathbf{R}_{ss} \mathbf{F}^H \mathbf{H}^H + \mathbf{R}_{vv})^{-1}. \quad (8)$$

Substituting (8) in (7) we arrive at the concentrated MSE matrix

$$\mathbf{R}_{ee} = \mathbf{U} (\mathbf{R}_{ss}^{-1} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{vv}^{-1} \mathbf{H} \mathbf{F})^{-1} \mathbf{U}^H \quad (9)$$

where we have used the matrix inversion lemma [10]. We also note that \mathbf{R}_{ee} is now no longer a function of \mathbf{W} .

B. Constrained Optimisation Problem

The transceivers in this paper aim to minimise the arithmetic MSE. However as in [7] and [9] rather than directly minimising $\text{tr}\{\mathbf{R}_{ee}\}/N_s$ we shall minimise a lower bound on the MSE and then select processors such that the arithmetic MSE achieves the minimised lower bound. A lower bound can be obtained from the arithmetic-geometric mean inequality [10] which states that for a positive semi-definite matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ we have $|\mathbf{A}|^{1/N} \leq \text{tr}\{\mathbf{A}\}/N$ where equality is achieved when \mathbf{A} is a diagonal matrix with equal diagonal elements. This provides us with the following bounds on \mathbf{R}_{ee}

$$\begin{aligned} & |(\mathbf{R}_{ss}^{-1} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{vv}^{-1} \mathbf{H} \mathbf{F})|^{-1/N_s} \\ & \leq \text{tr}\{\mathbf{U}(\mathbf{R}_{ss}^{-1} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{vv}^{-1} \mathbf{H} \mathbf{F})^{-1} \mathbf{U}^H\} / N_s \end{aligned} \quad (10)$$

where we have used the facts that $|\mathbf{A}\mathbf{B}| = |\mathbf{B}\mathbf{A}|$, $|\mathbf{U}^H \mathbf{U}| = 1$ since \mathbf{U} is unit diagonal lower left triangular, and $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$. The lower and upper bounds are the geometric MSE and arithmetic MSE respectively. The arithmetic MSE can only achieve the lower bound in (10) when $\mathbf{R}_{ee} = \beta \mathbf{I}_{N_s}$ for a scalar value $\beta \geq 0$. Using the lower bound in (10) as our objective function and taking the source and relay average power constraints into consideration we arrive at the following constrained optimisation problem

$$\max_{\mathbf{F}, \mathbf{G}} |(\mathbf{R}_{ss}^{-1} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{vv}^{-1} \mathbf{H} \mathbf{F})| \quad (11)$$

$$\text{subject to} \quad \text{tr}\{\mathbf{F} \mathbf{R}_{ss} \mathbf{F}^H\} = P_s, \quad \text{and} \quad (12)$$

$$\text{tr}\{\mathbf{G}(\mathbf{H}_s \mathbf{F} \mathbf{R}_{ss} \mathbf{F}^H \mathbf{H}_s^H + \mathbf{R}_{v_s v_s}) \mathbf{G}^H\} = P_r \quad (13)$$

where the source and relay average power constraints are given by (12) and (13) respectively where P_s and P_r are the available power budgets. The minimisation problem in (10) has been converted to a maximisation problem since $\min|\mathbf{A}|^{-1}$ is equivalent to $\max|\mathbf{A}|$. It is worth noting that minimising the geometric MSE is equivalent to maximising the MI [11]. Our design thus not only simultaneously minimises the arithmetic and geometric MSE but also maximises the MI.

C. Optimal Precoder Structures

Having formulated our constrained optimisation problem we now focus on deriving the optimal structure for the precoders \mathbf{F} , \mathbf{G} , and \mathbf{U} . The matrix optimisation problem stated in (11), (12), and (13) can be significantly simplified if we consider channel matrices \mathbf{H}_s and \mathbf{H}_r in terms of their singular value decompositions (SVD)

$$\mathbf{H}_s = \mathbf{U}_s \mathbf{\Lambda} \mathbf{V}_s^H \quad \mathbf{H}_r = \mathbf{U}_r \mathbf{\Delta} \mathbf{V}_r^H \quad (14)$$

where matrices $\mathbf{U}_s \in \mathbb{C}^{N_r \times N_r}$, $\mathbf{V}_s \in \mathbb{C}^{N_s \times N_s}$, $\mathbf{U}_r \in \mathbb{C}^{N_a \times N_a}$, and $\mathbf{V}_r \in \mathbb{C}^{N_r \times N_r}$ are unitary matrices associated with the singular values of \mathbf{H}_s and \mathbf{H}_r . The diagonal matrices $\mathbf{\Lambda} \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{\Delta} \in \mathbb{C}^{N_a \times N_r}$ contain the singular values of channel matrices \mathbf{H}_s and \mathbf{H}_r . For later convenience we shall define $\tilde{\mathbf{V}}_r$, and $\tilde{\mathbf{U}}_s$ to contain the leftmost N_s columns of \mathbf{V}_r , and \mathbf{U}_s respectively. We shall

also define matrices $\tilde{\mathbf{\Lambda}} \triangleq \text{diag}\{\lambda_{11}, \lambda_{22}, \dots, \lambda_{N_s N_s}\}$ and $\tilde{\mathbf{\Delta}} \triangleq \text{diag}\{\delta_{11}, \delta_{22}, \dots, \delta_{N_s N_s}\}$ to contain the largest singular values of $\mathbf{\Lambda}$ and $\mathbf{\Delta}$ respectively. It is assumed here that $N_s \leq \text{rank}(\mathbf{H}_s)$ and $N_s \leq \text{rank}(\mathbf{H}_r)$.

We shall also find it convenient to express the source and relay precoders \mathbf{F} and \mathbf{G} in terms of the following decompositions

$$\mathbf{F} = \mathbf{\Theta} \mathbf{\Gamma} \mathbf{\Psi} \quad \mathbf{G} = \mathbf{\Xi} \mathbf{\Phi} \mathbf{\Upsilon} \quad (15)$$

where we have the unitary matrices $\mathbf{\Theta} \in \mathbb{C}^{N_s \times N_s}$, $\mathbf{\Psi} \in \mathbb{C}^{N_s \times N_s}$, $\mathbf{\Xi} \in \mathbb{C}^{N_r \times N_r}$, $\mathbf{\Upsilon} \in \mathbb{C}^{N_r \times N_r}$, and diagonal matrices $\mathbf{\Gamma} \in \mathbb{C}^{N_s \times N_s}$ and $\mathbf{\Phi} \in \mathbb{C}^{N_r \times N_r}$. As previously done for the channel decompositions we shall define $\tilde{\mathbf{\Xi}}$ and $\tilde{\mathbf{\Upsilon}}$ to contain the left N_s columns of $\mathbf{\Xi}$ and $\mathbf{\Upsilon}$ as well as diagonal matrices $\tilde{\mathbf{\Phi}} \triangleq \text{diag}\{\phi_{11}, \phi_{22}, \dots, \phi_{N_s N_s}\}$ and $\tilde{\mathbf{\Gamma}} \triangleq \text{diag}\{\gamma_{11}, \gamma_{22}, \dots, \gamma_{N_s N_s}\}$.

The optimal structures for \mathbf{F} and \mathbf{G} can be derived from the Hadamard determinant inequality which states that for a positive semi-definite matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ we have the inequality $|\mathbf{A}| \leq \prod_{i=1}^N a_{ii}$. Applying the inequality to our objective function in (11) and using the channel and precoder decompositions we can state that

$$|(\mathbf{R}_{ss}^{-1} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{vv}^{-1} \mathbf{H} \mathbf{F})| \leq \prod_{i=1}^{N_s} \left(\sigma_s^{-2} + \frac{\gamma_{ii}^2 \lambda_{ii}^2 \phi_{ii}^2 \delta_{ii}^2}{\phi_{ii}^2 \delta_{ii}^2 \sigma_{v_s}^2 + \sigma_{v_r}^2} \right) \quad (16)$$

where equality holds when we have $\mathbf{\Theta} = \mathbf{V}_s$, $\tilde{\mathbf{\Xi}} = \tilde{\mathbf{V}}_r$, and $\tilde{\mathbf{\Upsilon}} = \tilde{\mathbf{U}}_s^H$. We have thus established the following sets of optimal source and relay precoders

$$\mathbf{F} = \mathbf{V}_s \mathbf{\Gamma} \mathbf{\Psi} \quad \mathbf{G} = \tilde{\mathbf{V}}_r \tilde{\mathbf{\Phi}} \tilde{\mathbf{U}}_s^H \quad (17)$$

where $\mathbf{\Psi}$ is an arbitrary unitary matrix yet to be defined. We note that the bound in (16) would also hold with equality with source precoders of the form $\mathbf{F} = \mathbf{V}_s \mathbf{\Gamma}$. However the unitary matrix $\mathbf{\Psi}$ provides us with a degree of freedom that shall be exploited later on in the design. Interestingly the precoders in (17) have the same structure as the optimal processors derived in e.g. [5] and [6] for Schur convex objective functions.

Having established the optimal source and relay precoder structures we now focus on calculating the unitary matrix $\mathbf{\Psi}$ and the feedback matrix \mathbf{U} . Substituting (17) into the lower bound in (10) we can calculate the lower MSE bound that the transceiver may achieve to be

$$\bar{\sigma}^2 = \prod_{i=1}^{N_s} \left(\sigma_s^{-2} + \frac{\gamma_{ii}^2 \lambda_{ii}^2 \phi_{ii}^2 \delta_{ii}^2}{\phi_{ii}^2 \delta_{ii}^2 \sigma_{v_s}^2 + \sigma_{v_r}^2} \right)^{-1/N_s} \quad (18)$$

As noted previously, due to the arithmetic-geometric mean inequality, the MMSE transceiver can only achieve this lower bound if the error covariance matrix in (9) is a diagonal matrix with equal diagonal elements given by $\bar{\sigma}^2$. We can thus state that the following condition must be met

$$\mathbf{U}(\mathbf{R}_{ss}^{-1} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{vv}^{-1} \mathbf{H} \mathbf{F})^{-1} \mathbf{U}^H = \bar{\sigma}^2 \mathbf{I}_{N_s} \quad (19)$$

Substituting the optimal precoders (17) in (19) and using the channel decompositions we can re-write (19) as

$$\mathbf{U} \mathbf{\Psi}^H \mathbf{\Sigma}^{-1/2} \mathbf{\Sigma}^{-1/2} \mathbf{\Psi} \mathbf{U}^H = \bar{\sigma}^2 \mathbf{Q}^H \mathbf{Q} \quad (20)$$

where $\Sigma \triangleq (\sigma_s^{-2} \mathbf{I}_{N_s} + \Gamma^2 \tilde{\Lambda}^2 \tilde{\Phi}^2 \tilde{\Delta}^2 (\tilde{\Phi}^2 \tilde{\Delta}^2 \sigma_{v_1}^2 + \sigma_{v_2}^2 \mathbf{I}_{N_s}))^{-1}$ and $\mathbf{Q} \in \mathbb{C}^{N_s \times N_s}$ is an arbitrary unitary matrix. Clearly both sides of (20) are hermitian and we can state that $\Sigma^{-1/2} \Psi \mathbf{U}^H = \bar{\sigma} \mathbf{Q}$ which upon re-arranging leads us to the following matrix decomposition

$$\Sigma^{-1/2} = \mathbf{Q} \tilde{\mathbf{U}} \Psi^H \quad (21)$$

where we define $\tilde{\mathbf{U}} \triangleq \bar{\sigma} \mathbf{U}^{-H}$. The decomposition in (21) is referred to as the geometric mean decomposition [12] otherwise known as the equal diagonal QR decomposition [13]. Since $\tilde{\mathbf{U}}$ is an upper right triangular matrix with equal diagonal elements given by $\bar{\sigma}$ we can calculate the required unit diagonal lower left triangular matrix as $\mathbf{U} = \bar{\sigma} \tilde{\mathbf{U}}^{-H}$ from which we get the strictly lower left triangular matrix $\mathbf{B} = \mathbf{U} - \mathbf{I}_{N_s}$.

D. Source and Relay Power Allocation

Having derived the optimal precoder structures for \mathbf{F} , \mathbf{G} , and \mathbf{U} as well as the equaliser matrix \mathbf{W} we now focus on calculating the source and relay power allocation matrices Γ and $\tilde{\Phi}$. Substituting (17) and (14) into (11), (12), and (13) we arrive at the following scalar optimisation problem

$$\max_{\phi_{ii}, \gamma_{ii}} \prod_{i=1}^{N_s} \left(\sigma_s^{-2} + \frac{\gamma_{ii}^2 \lambda_{ii}^2 \phi_{ii}^2 \delta_{ii}^2}{\phi_{ii}^2 \delta_{ii}^2 \sigma_{v_s}^2 + \sigma_{v_r}^2} \right) \quad (22)$$

$$\text{subject to } \sum_{i=1}^{N_s} \gamma_{ii}^2 \sigma_s^2 = P_s, \quad \gamma_{ii}^2 \geq 0, \quad (23)$$

$$\text{and } \sum_{i=1}^{N_s} \phi_{ii}^2 (\gamma_{ii}^2 \lambda_{ii}^2 \sigma_s^2 + \sigma_{v_s}^2) = P_r, \quad \phi_{ii}^2 \geq 0. \quad (24)$$

A closed form solution to this optimisation problem is intractable since the problem is highly non convex. However for a set γ_{ii} the problem is convex with respect to ϕ_{ii} and similarly for a set ϕ_{ii} the problem is convex with respect to γ_{ii} . With these observations we propose to update γ_{ii} and ϕ_{ii} in an alternating fashion as in [5]. Let us first define

$$\varphi_{ii} \triangleq \gamma_{ii}^2 \sigma_s^2 \quad (25)$$

$$\text{and } \rho_{ii} \triangleq \phi_{ii}^2 (\gamma_{ii}^2 \lambda_{ii}^2 \sigma_s^2 + \sigma_{v_s}^2). \quad (26)$$

Substituting (25) and (26) into (22), (23), and (24) we can re-formulate the original optimisation problem as

$$\max_{\varphi_{ii}, \rho_{ii}} \sum_{i=1}^{N_s} \log \frac{\rho_{ii} \delta_{ii}^2 \varphi_{ii} \lambda_{ii}^2 + \rho_{ii} \delta_{ii}^2 \sigma_{v_s}^2 + \varphi_{ii} \lambda_{ii}^2 \sigma_{v_r}^2 + \sigma_{v_s}^2 \sigma_{v_r}^2}{\rho_{ii} \delta_{ii}^2 \sigma_{v_s}^2 + \varphi_{ii} \lambda_{ii}^2 \sigma_{v_r}^2 + \sigma_{v_s}^2 \sigma_{v_r}^2} \quad (27)$$

$$\text{subject to } \sum_{i=1}^{N_s} \varphi_{ii} = P_s, \quad \varphi_{ii} \geq 0, \quad (28)$$

$$\text{and } \sum_{i=1}^{N_s} \rho_{ii} = P_r, \quad \rho_{ii} \geq 0. \quad (29)$$

We note that the optimisation problem is symmetric in φ_{ii} and ρ_{ii} and moreover the source and relay power constraints are now independent of each other. For a given φ_{ii} that satisfies

the power constraint in (28) we can calculate the optimal ρ_{ii} by solving

$$\max_{\rho_{ii}} \sum_{i=1}^{N_s} \log \frac{\rho_{ii} \delta_{ii}^2 \varphi_{ii} \lambda_{ii}^2 + \rho_{ii} \delta_{ii}^2 \sigma_{v_s}^2 + \varphi_{ii} \lambda_{ii}^2 \sigma_{v_r}^2 + \sigma_{v_s}^2 \sigma_{v_r}^2}{\rho_{ii} \delta_{ii}^2 \sigma_{v_s}^2 + \varphi_{ii} \lambda_{ii}^2 \sigma_{v_r}^2 + \sigma_{v_s}^2 \sigma_{v_r}^2} \quad (30)$$

$$\text{subject to } \sum_{i=1}^{N_s} \rho_{ii} = P_r, \quad \rho_{ii} \geq 0. \quad (31)$$

Since the objective function and inequality constraint are both convex and the equality constraint is affine with respect to our design parameter ρ_{ii} we can solve this problem very efficiently by the Karush Kuhn Tucker (KKT) conditions of optimality [14]. The solution can be obtained to be

$$\rho_{ii} = \frac{\sigma_{v_r}^2}{2\delta_{ii}^2} \left[\sqrt{\frac{\lambda_{ii}^4 \varphi_{ii}^2}{\sigma_{v_s}^4} + \frac{4\lambda_{ii}^2 \varphi_{ii} \delta_{ii}^2 \mu_r}{\sigma_{v_s}^2 \sigma_{v_r}^2}} - \frac{\lambda_{ii}^2 \varphi_{ii}}{\sigma_{v_s}^2} - 2 \right]^+ \quad (32)$$

where the variable μ_r must be calculated to ensure that the power constraint in (31) is met with equality. In a similar fashion for a set ρ_{ii} the optimal φ_{ii} can be calculated by solving (27) and (28). The solution to φ_{ii} can be calculated to be

$$\varphi_{ii} = \frac{\sigma_{v_s}^2}{2\lambda_{ii}^2} \left[\sqrt{\frac{\delta_{ii}^4 \rho_{ii}^2}{\sigma_{v_r}^4} + \frac{4\lambda_{ii}^2 \rho_{ii} \delta_{ii}^2 \mu_s}{\sigma_{v_s}^2 \sigma_{v_r}^2}} - \frac{\delta_{ii}^2 \rho_{ii}}{\sigma_{v_r}^2} - 2 \right]^+ \quad (33)$$

where μ_s must be calculated to ensure that (28) holds with equality. The alternating algorithm is conducted in the following manner. After selecting an appropriate initial choice for φ_{ii} that satisfies (28) the algorithm updates ρ_{ii} according to (32) and φ_{ii} according to (33) in an alternating fashion. Since the updates of φ_{ii} and ρ_{ii} can never decrease the objective function [5] the algorithm is guaranteed to converge. Once the algorithm has converged we can finally calculate the power allocation parameters γ_{ii} and ϕ_{ii} at the source and relay using $\gamma_{ii} = \sqrt{\varphi_{ii}/\sigma_s^2}$ and $\phi_{ii} = \sqrt{\rho_{ii}/(\gamma_{ii}^2 \lambda_{ii}^2 \sigma_s^2 + \sigma_{v_s}^2)}$

IV. SIMULATION RESULTS

In this section we evaluate the performance of the proposed THP design compared to the linear maximum MI (MMI) and MMSE designs in [3] and [4] as well as the linear naive AF (NAF) scheme. In the NAF design the Wiener filter is utilised at the destination with the source and relay precoders being selected as $\mathbf{F} = \sqrt{P_s/N_s} \sigma_s^2 \mathbf{I}_{N_s}$ and $\mathbf{G} = \sqrt{P_r/\text{tr}\{\mathbf{H}_s \mathbf{F} \mathbf{F}^H \mathbf{H}_s^H \sigma_s^2 + \sigma_{v_s}^2 \mathbf{I}_{N_r}\}} \mathbf{I}_{N_r}$. We also compare the proposed method to a suboptimal THP transceiver where power is allocated uniformly across the source and relay antennas. The suboptimal design is the same as the proposed solution but with the source and relay power allocation matrices selected as $\Gamma = \sqrt{P_s/N_s} \sigma_s^2 \mathbf{I}_{N_s}$ and $\Phi = \sqrt{P_r/\text{tr}\{\Gamma^2 \Lambda_s^2 \sigma_s^2 + \sigma_{v_s}^2 \mathbf{I}_{N_s}\}} \mathbf{I}_{N_s}$.

We simulate a two-hop MIMO relay system with $N_s = N_r = N_d = 4$ antennas in each stage of the network with MIMO channels \mathbf{H}_s and \mathbf{H}_r having complex Gaussian entries with zero mean and unit variance. The symbols from the

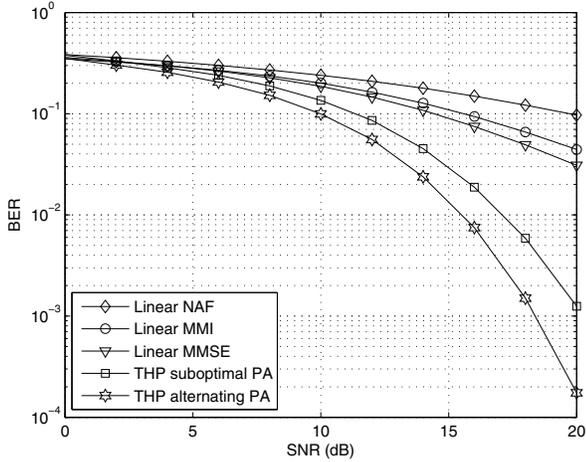


Fig. 2. BER versus SNR. $N_s = N_r = N_d = 4$

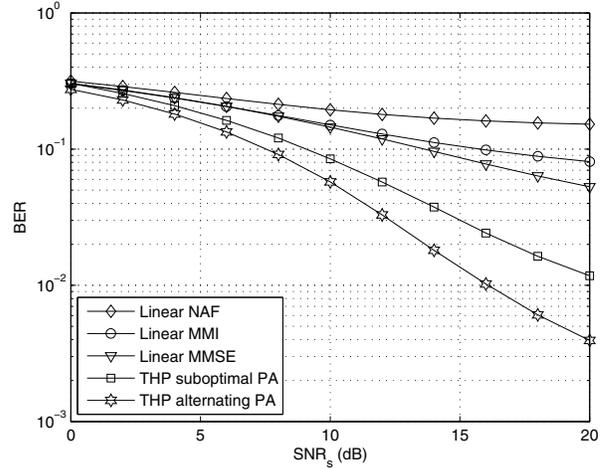


Fig. 3. BER versus SNR_s with $SNR_r = 15$ dB. $N_s = N_r = N_d = 4$

source antennas are selected from 16 QAM constellations and the SNR for the source to relay and relay to destination stages are defined as $SNR_s = P_s/N_s\sigma_{v_s}^2$ and $SNR_r = P_r/N_r\sigma_{v_r}^2$, respectively. Fig. 2 shows the uncoded BER for the proposed and benchmark schemes against varying SNR where we set $SNR = SNR_s = SNR_r$. Fig. 3 shows simulation results for varying SNR_s with $SNR_r = 15$ dB. All results were averaged over 500 independent channel realisations. Clearly both THP designs offer improved performance in terms of BER compared to the linear systems particularly at mid to high SNR values. The THP design that utilises the alternating power allocation algorithm provides the best performance over all SNR values.

V. CONCLUSIONS

In this paper we derived the optimal MMSE THP transceiver for MIMO relay networks in the absence of a direct link between the source and destination antennas. Under the assumption of full CSI the processors were jointly optimised to minimise the arithmetic MSE whilst abiding by average power constraints at the source and relay terminals. Simulation results show that the proposed schemes outperform existing transceiver designs in terms of BER.

ACKNOWLEDGEMENT

We gratefully acknowledge a Motorola University Research Grant and a University of Strathclyde Scholarship, which supported this research.

REFERENCES

- [1] X. Tang and Y. Hua. Optimal design of non-regenerative MIMO wireless relays. *IEEE Transactions on Wireless Communications*, 6(4):1398–1407, Apr. 2007.
- [2] B. Wang, J. Zhang, and A. Host-Madsen. On the capacity of MIMO relay channels. *IEEE Transactions on Information Theory*, 51(1):29–43, Jan. 2005.
- [3] O. Munoz-Medina, J. Vidal, and A. Agustin. Linear Transceiver Design in Nonregenerative Relays with Channel State Information. *IEEE Transactions on Signal Processing*, 55(6):2593–2604, Jun. 2007.

- [4] W. Guan and H. Luo. Joint MMSE transceiver design in non-regenerative MIMO relay systems. *IEEE Communications Letters*, 12(7):517–519, Jul. 2008.
- [5] Y. Rong, X. Tang, and Y. Ha. A Unified Framework for Optimizing Linear Non-Regenerative Multicarrier MIMO Relay Communication Systems. *IEEE Transactions on Signal Processing*, 57(12):4837–4851, Dec. 2009.
- [6] Y. Rong and Y. Hua. Optimality of Diagonalization of Multi-Hop MIMO Relays. *IEEE Transactions on Wireless Communications*, 8(12):6068–6077, Dec. 2009.
- [7] M. B. Shenoouda and T. N. Davidson. A Framework for Designing MIMO Systems with Decision Feedback Equalization or Tomlinson-Harashima Precoding. *IEEE Journal on Selected Areas in Communications*, 26(2):401–411, Feb. 2008.
- [8] A. A. D’Amico and M. Morelli. Joint Tx-Rx MMSE Design for MIMO Multicarrier Systems with Tomlinson-Harashima Precoding. *IEEE Transactions on Wireless Communications*, 7(8):3118–3127, Aug. 2008.
- [9] F. Xu, T. N. Davidson, J. K. Zhang, and K. M. Wong. Design of Block Transceivers with Decision Feedback Detection. *IEEE Transactions on Signal Processing*, 54(3):965–978, Mar. 2006.
- [10] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [11] 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 Transactions on Signal Processing*, 51(9):2381–2401, Sep. 2003.
- [12] Y. Jiang, J. Li, and W. W. Hager. Joint Transceiver Design for MIMO Communications Using Geometric Mean Decomposition. *IEEE Transactions on Signal Processing*, 53(10):3791–3801, Oct. 2005.
- [13] J. K. Zhang, A. Kavcic, and K. M. Wong. Equal Diagonal QR Decomposition and its Application to Precoder Design for Successive Cancellation Detection. *IEEE Transactions on Information Theory*, 51(1):154–172, Jan. 2005.
- [14] S. Boyd and L. Vandenberghe. *Convex Optimisation*. Cambridge University Press, 1985.