



Strathprints Institutional Repository

Millar, Andrew P. and Weiss, Stephan and Stewart, Robert W. (2015)
Training based channel estimation algorithms for dual hop MIMO OFDM
relay systems. IEEE Transactions on Communications, 63 (12). pp. 4711-
4726. ISSN 0090-6778 , <http://dx.doi.org/10.1109/TCOMM.2015.2481897>

This version is available at <http://strathprints.strath.ac.uk/54337/>

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: strathprints@strath.ac.uk

Training-Based Channel Estimation Algorithms for Dual Hop MIMO OFDM Relay Systems

Andrew P. Millar, Stephan Weiss, *Senior Member, IEEE*, and Robert W. Stewart

Abstract—In this paper, we consider minimum-mean-square error (MMSE) training-based channel estimation for two-hop multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) relaying systems. The channel estimation process is divided into two main phases. The relay-destination channel is estimated in the first phase and can be obtained using well-known point-to-point MIMO OFDM estimation methods. In the second phase, the source-relay channel is estimated at the destination with the use of a known training sequence that is transmitted from the source and forwarded to the destination by a nonregenerative relay. To obtain an estimate of the source-relay channel, the source training sequence, relay precoder, and destination processor, require to be optimized. To solve this problem, we first derive an iterative algorithm that involves sequentially solving a number of convex optimization problems to update the source, relay, and destination design variables. Since the iterative algorithm may be too computationally expensive for practical implementation, we then derive simplified solutions that have reduced computational complexity. Simulation results demonstrate the effectiveness of the proposed algorithms.

Index Terms—MIMO OFDM, relay networks, MMSE channel estimation, training sequences design.

I. INTRODUCTION

MIMO relaying has recently attracted significant interest due to the potential benefits of extended network coverage, increased data throughput, and robustness to channel impairments such as strong shadowing and multipath fading.

Transceiver designs for MIMO relaying have been extensively studied in e.g. [1]–[9], where it is mainly assumed that perfect channel state information (CSI) is available. The use of well-known conventional point-to-point channel estimation algorithms (such as those in [10]–[14]) for estimation in MIMO relay networks would require the source-relay and relay-destination channels to be estimated at the relay and destination, respectively. Such an approach is appropriate for regenerative relays (also known as decode-forward relay networks) where the relay has the signal processing functionality to perform the task of channel estimation [15], [16]. However,

non-regenerative relays (also known as amplify-forward relays) are expected to have simpler functionality and lower implementation cost than decode-forward relays. As such they have limited signal processing capabilities [17]–[23] and are therefore not expected to perform complex tasks such as channel estimation. In fact it is desirable to keep the complexity of non-regenerative relays to a minimum. For these reasons the estimation of the source-relay channel at the relay in such networks is considered impractical, and alternative estimation methods have therefore been suggested.

For channel estimation problems in non-regenerative MIMO relaying, one possible approach is to estimate the compound MIMO channel through observations at the destination device (as in e.g. [15], [24]). Although techniques that estimate the compound source-destination MIMO channel allow reliable detection at the destination device, they cannot fully exploit the potential benefits (such as improved MSE and BER performance etc.) that are made available by sophisticated transceiver designs such as those developed in [1]–[9], which require knowledge of both the source-relay and relay-destination channels. Alternative methods that are capable of estimating both these channels have therefore been developed [18], [19], [21]–[23], [25], [26], where all channel estimation is conducted at the destination.

Least squares (LS) and weighted least squares algorithms have been developed in [25] and [26], respectively, where the source-relay and relay-destination channels are estimated from the observed composite MIMO channel at the destination. Necessary and sufficient conditions for obtaining the source-relay and relay-destination channels from the composite channel observed at the destination are studied in [19].

Two-phase channel estimation procedures have been developed in [18], [21]–[23]. In these works, the relay-destination channel is estimated in the first phase using conventional techniques. In the second phase, the source sends training symbols to the relay, which precodes the received symbols and forwards them to the destination. The destination can then estimate the source-relay channel using the knowledge of the relay-destination channel obtained in the first phase of channel estimation. In [18] the authors derive the optimal source training symbols and relay precoder under the assumption that the relay-destination channel is perfectly estimated during the first phase. Imperfect estimation of the relay-destination channel adversely affects the source-relay estimate obtained by the algorithm in [18]. The works of [21]–[23] therefore develop robust estimation algorithms that account for such imperfect relay-destination CSI.

Manuscript received August 7, 2014; revised February 2, 2015 and July 14, 2015; accepted September 14, 2015. The associate editor coordinating the review of this paper and approving it for publication was Prof. A. Nallanathan. This work was in parts funded by the UK Engineering and Physical Sciences Research Council (EPSRC).

A. P. Millar is with Genesis Oil and Gas Consultants Ltd., Aberdeen AB10 1YH, Scotland (e-mail: andrew.paul.millar@gmail.com).

S. Weiss and R. W. Stewart are with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, Scotland (e-mail: stephan.weiss@strath.ac.uk; robert.stewart@strath.ac.uk).

Digital Object Identifier 10.1109/TCOMM.2015.2481897

In this paper we consider the channel estimation problem for non-regenerative MIMO OFDM relay networks for transmission over frequency selective channels. Similar to [18], [21]–[23] we adopt a two-phase channel estimation approach. In the proposed algorithms, the relay-destination channel is equivalent to a point-to-point MIMO OFDM channel estimation problem and the algorithms in e.g. [11]–[13] can be used to obtain the channel estimate at the destination. The main contribution of this paper lies in the estimation of the source-relay channel, which is conducted at the destination in the second phase of channel estimation. We consider the design of the source training matrix, relay precoder, and destination processor in order to estimate the source-relay channel at the destination. An iterative algorithm is firstly considered, before simplified approaches with reduced complexity are presented.

Notation: In our notation we denote scalars, vectors, and matrices by lower case normal font, lower case bold font, and upper case bold font respectively. The quantities \mathbf{I}_N and $\mathbf{0}_{N \times M}$ denote the $N \times N$ identity matrix and $N \times M$ zero matrix. The element in the i th row and j th column of matrix \mathbf{A} is denoted $[\mathbf{A}]_{ij}$. The operators $\mathbb{E}\{\cdot\}$, $\text{tr}\{\cdot\}$, $(\cdot)^T$, $(\cdot)^H$, $\{\cdot\}^\dagger$, and $\|\cdot\|_F$ denote expectation, trace, transpose, Hermitian transpose, pseudo inverse, and Frobenius norm, respectively. Matrix rank is denoted $\text{rank}\{\cdot\}$ and $\text{diag}\{\{\mathbf{A}_n\}_{n=1}^N\}$ produces a block diagonal matrix with the n th diagonal block given by \mathbf{A}_n . The operators $\min(\cdot)$ and $\max(\cdot)$ return the minimum and maximum, and we define $[x]^+ \triangleq \max(x, 0)$. The floor operator $\lfloor \cdot \rfloor$ returns the maximum integer not exceeding the argument. The Kronecker product is denoted by \otimes and the vectorisation operator is denoted $\text{vec}\{\cdot\}$. The notation $\mathbf{A} \succeq \mathbf{B}$ signifies that \mathbf{A} is positive semi-definite w.r.t. \mathbf{B} .

II. SIGNAL MODEL AND CHANNEL ESTIMATION

A. OFDM Channel Model

We consider a half duplex two-hop non-regenerative MIMO relaying system where the source, relay, and destination devices are equipped with N_s , N_r , and N_d antennas, respectively. The source-relay and relay-destination channels are considered to be frequency selective with $L + 1$ distinguishable delay paths between each transmit-receive antenna pair. We denote the l th MIMO taps of the source-relay and relay-destination channels by $\mathcal{H}_s[l] \in \mathbb{C}^{N_r \times N_s}$ and $\mathcal{H}_r[l] \in \mathbb{C}^{N_d \times N_r}$, respectively, and assume each delay path to be spatially correlated on both the transmit and receive sides. We adopt the Kronecker spatial correlation model and can thus decompose the source-relay and relay-destination channel taps as [27], [28]

$$\mathcal{H}_s[l] = \mathbf{Y}_s^{1/2}[l]\mathcal{H}_{sw}[l]\mathbf{\Theta}_s^{T/2}[l], \quad l = 0, \dots, L, \quad (1)$$

$$\mathcal{H}_r[l] = \mathbf{Y}_r^{1/2}[l]\mathcal{H}_{rw}[l]\mathbf{\Theta}_r^{T/2}[l], \quad l = 0, \dots, L, \quad (2)$$

where $\mathbf{Y}_s[l] \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{\Theta}_s[l] \in \mathbb{C}^{N_s \times N_s}$ are the receive side and transmit side spatial correlation matrices, respectively, for the l th delay path of the source-relay channel. Similarly, $\mathbf{Y}_r[l] \in \mathbb{C}^{N_d \times N_d}$ and $\mathbf{\Theta}_r[l] \in \mathbb{C}^{N_r \times N_r}$ are the spatial correlation matrices for the l th delay path of the relay-destination

channel. We assume that different delay paths are uncorrelated and the elements of $\mathcal{H}_s[l]$ and $\mathcal{H}_r[l]$ are circularly symmetric complex Gaussian random variables drawn from $\mathcal{CN}(0, \sigma_{h_s}^2[l])$ and $\mathcal{CN}(0, \sigma_{h_r}^2[l])$, respectively.

To deal with the frequency selective nature of the wireless channels, OFDM is employed with K subcarriers for both the source-relay and relay-destination transmission stages. Assuming that the OFDM cyclic prefix on each antenna branch is of length L , the frequency selective channels are decoupled into the K orthogonal narrowband subcarrier channel matrices

$$\mathbf{H}_{s,k} = \sum_{l=0}^L \mathcal{H}_s[l]W_{kl} \in \mathbb{C}^{N_r \times N_s}, \quad k = 1, \dots, K, \quad (3)$$

$$\mathbf{H}_{r,k} = \sum_{l=0}^L \mathcal{H}_r[l]W_{kl} \in \mathbb{C}^{N_d \times N_r}, \quad k = 1, \dots, K, \quad (4)$$

where $\mathbf{H}_{s,k}$ and $\mathbf{H}_{r,k}$ are the k th subcarriers for the source-relay and relay-destination channels, respectively, and we define $W_{kl} \triangleq e^{-j2\pi(k-1)l/K}$. Channel estimation can be conducted to either directly estimate the OFDM subcarrier channels in (3) and (4), or to estimate the underlying time domain channels characterised by (1) and (2). In this paper we develop channel estimation algorithms for the latter case since it generally requires fewer parameters to be estimated.

In our channel model we assume that both the source-relay and relay-destination channels remain constant over the channel estimation process. Furthermore, it is also assumed that the coherence time of both channels is long enough such that the subsequent transmission of data symbols may be conducted without the overhead spent on channel estimation and feedback becoming prohibitive and thereby limiting the system spectral efficiency.

B. Channel Estimation Procedure

Similar to the works in [18] and [21], we assume that the estimation of the source-relay and relay-destination channels is conducted in two training phases. In the first phase the source remains silent and the relay sends a known training sequence to the destination. The estimation of the relay-destination channel in this case is a standard point-to-point MIMO OFDM estimation problem and the algorithms developed in [11]–[13] can be used to estimate the channel at the destination. Since the estimation of the relay-destination channel is a standard MIMO OFDM channel estimation problem it shall not be discussed further. The main focus of this paper is on the estimation of the source-relay channel, which is obtained at the destination during the second phase of channel estimation.

The second phase of channel estimation is dedicated to estimating the source-relay channel matrices in (1). In this phase the source transmits a training sequence to the relay whilst the relay remains silent. For each OFDM subcarrier we can write the received signal $\mathbf{r}_k \in \mathbb{C}^{N_r}$ at the relay as

$$\mathbf{r}_k = \mathbf{H}_{s,k}\mathbf{s}_k + \mathbf{v}_{s,k}, \quad k = 1, \dots, K, \quad (5)$$

where $\mathbf{s}_k \in \mathbb{C}^{N_s}$ is the training signal for the k th subcarrier, and $\mathbf{v}_{s,k} \in \mathbb{C}^{N_r}$ is an additive white Gaussian noise (AWGN) vector, which contains independent identically distributed (i.i.d.) zero mean complex Gaussian random variables, and has covariance $\mathbb{E}\{\mathbf{v}_{s,k}\mathbf{v}_{s,k}^H\} = \sigma_{v_s}^2 \mathbf{I}_{N_r}$. Substituting (3) into (5), and using $\text{vec}[\mathbf{A}\mathbf{X}\mathbf{B}] = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}[\mathbf{X}]$, we have

$$\mathbf{r}_k = (\mathbf{m}_k \otimes \mathbf{I}_{N_r}) \mathbf{h}_s + \mathbf{v}_{s,k}, \quad k = 1, \dots, K, \quad (6)$$

where for notational convenience we define

$$\mathbf{m}_k \triangleq [\mathbf{s}_k^T \mathbf{W}_{k0}, \dots, \mathbf{s}_k^T \mathbf{W}_{kL}] \in \mathbb{C}^{N_s(L+1)}, \quad k = 1, \dots, K, \quad (7)$$

$$\mathbf{h}_s \triangleq \text{vec}[\mathcal{H}_s[0], \dots, \mathcal{H}_s[L]] \in \mathbb{C}^{Lh_s}. \quad (8)$$

In (8) we define $L_{h_s} \triangleq N_s N_r (L+1)$ as being the number of source-relay channel coefficients that require to be estimated. Collecting the received signals from (6) into a single column vector defined as $\mathbf{r} \triangleq \text{vec}[\mathbf{r}_1, \dots, \mathbf{r}_K] \in \mathbb{C}^{KN_r}$, it can be straightforwardly shown that

$$\mathbf{r} = (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{h}_s + \mathbf{v}_s, \quad (9)$$

where $\mathbf{v}_s \triangleq \text{vec}[\mathbf{v}_{s,1}, \dots, \mathbf{v}_{s,K}] \in \mathbb{C}^{KN_r}$ is the collection of source-relay AWGN vectors over all subcarriers, and we also define the matrices

$$\mathbf{M} \triangleq [\mathbf{F}_0 \mathbf{S}, \dots, \mathbf{F}_L \mathbf{S}] \in \mathbb{C}^{K \times N_s(L+1)} \quad (10)$$

$$\mathbf{F}_l \triangleq \text{diag}[\{W_{kl}\}_{k=1}^K] \in \mathbb{C}^{K \times K}, \quad l = 0, \dots, L, \quad (11)$$

$$\mathbf{S} \triangleq [\mathbf{s}_1, \dots, \mathbf{s}_K]^T \in \mathbb{C}^{K \times N_s}. \quad (12)$$

The matrix \mathbf{S} in (12) is the training matrix associated with the source-relay channel estimation problem, which requires to be optimised. It is worth noting here that if the relay device has the ability to perform channel estimation then, based on the received signal in (9), the estimate of \mathbf{h}_s can be computed at the relay using the point-to-point channel estimation algorithms in e.g. [12]. However, a non-regenerative relay as considered here may be limited in its processing capability and it is desirable to keep the computational expense at the relay as low as possible.

Similar to the works in [18] and [21], we consider that the relay only performs a simple precode-and-forward operation and channel estimation is then performed at the destination. The relay device thus precodes the received signal in (9) to produce the transmit signal

$$\hat{\mathbf{r}} = \mathbf{G}((\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{h}_s + \mathbf{v}_s) \in \mathbb{C}^{KN_r}, \quad (13)$$

where $\mathbf{G} \in \mathbb{C}^{KN_r \times KN_r}$ is the linear relay precoder that operates over all subcarriers and also requires to be optimised. After performing the precoding operation the relay transmits the signal

in (13) to the destination node, resulting in the signal $\mathbf{y} \in \mathbb{C}^{KN_d}$ received at the destination being given by

$$\mathbf{y} = \mathbf{H}_r \mathbf{G}(\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{h}_s + \mathbf{H}_r \mathbf{G} \mathbf{v}_s + \mathbf{v}_r, \quad (14)$$

where $\mathbf{v}_r \in \mathbb{C}^{KN_d}$ is the relay-destination AWGN vector that contains i.i.d. zero mean complex Gaussian random variables and has covariance $\mathbb{E}\{\mathbf{v}_r \mathbf{v}_r^H\} = \sigma_{v_r}^2 \mathbf{I}_{KN_d}$. In (14) the matrix $\mathbf{H}_r \in \mathbb{C}^{KN_d \times KN_r}$ describes the relay-destination OFDM channel over all subcarriers and is defined as

$$\mathbf{H}_r \triangleq \text{diag}[\{\mathbf{H}_{r,k}\}_{k=1}^K], \quad (15)$$

with the subcarrier channels $\mathbf{H}_{r,k}$ being given in (4).

From the received signal in (14) the task is to compute an estimate of the source-relay channel taps given in (1). This is equivalent to the estimation of the channel vector \mathbf{h}_s in (8). In the following we shall denote the estimate of \mathbf{h}_s by

$$\hat{\mathbf{h}}_s \triangleq \text{vec}[\hat{\mathcal{H}}_s[0], \dots, \hat{\mathcal{H}}_s[L]] \in \mathbb{C}^{Lh_s}, \quad (16)$$

with $\hat{\mathcal{H}}_s[l] \in \mathbb{C}^{N_r \times N_s}$ signifying the estimate of the l th source-relay MIMO channel tap in (1). In order to facilitate the optimisation of the channel estimate in (16) we introduce a linear processor $\mathbf{W} \in \mathbb{C}^{Lh_s \times KN_d}$ at the destination and let $\hat{\mathbf{h}}_s = \mathbf{W} \mathbf{y}$. Using the received signal in (14) the source-relay channel estimate is then given by

$$\hat{\mathbf{h}}_s = \mathbf{W}(\mathbf{H}_r \mathbf{G}(\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{h}_s + \mathbf{H}_r \mathbf{G} \mathbf{v}_s + \mathbf{v}_r). \quad (17)$$

To obtain a useful channel estimate in (17) we require to optimise the destination processor \mathbf{W} , the linear relay precoder \mathbf{G} , and the source training matrix \mathbf{S} (note that \mathbf{M} is a function of \mathbf{S} through (10)). Since both the source-relay and relay-destination channels are estimated at the destination device, it is convenient that the optimisation problem is solved at the destination. The source training matrix and relay precoder may then be communicated back to the source and relay, respectively, through low rate feedback channels.

III. PROBLEM FORMULATION

In this section we formulate the optimisation problem for deriving the source training matrix, the relay precoder and the destination processor to minimise the source-relay channel estimation MSE. We firstly denote the error between the estimated channel vector $\hat{\mathbf{h}}_s$ and the actual channel vector \mathbf{h}_s as $\mathbf{e} = \hat{\mathbf{h}}_s - \mathbf{h}_s \in \mathbb{C}^{Lh_s}$. Using (17) we can express the source-relay estimation error as

$$\mathbf{e} = (\mathbf{W} \mathbf{H}_r \mathbf{G}(\mathbf{M} \otimes \mathbf{I}_{N_r}) - \mathbf{I}_{Lh_s}) \mathbf{h}_s + \mathbf{W}(\mathbf{H}_r \mathbf{G} \mathbf{v}_s + \mathbf{v}_r). \quad (18)$$

The channel estimation MSE cost function is then given by $\Psi(\mathbf{S}, \mathbf{G}, \mathbf{W}) = \text{tr}\{\mathbb{E}\{\mathbf{e}\mathbf{e}^H\}\}$, where the expectation is taken w.r.t. the random noise components \mathbf{v}_s and \mathbf{v}_r , as well as

the unknown channel vector \mathbf{h}_s . Using (18) we can expand $\Psi(\mathbf{S}, \mathbf{G}, \mathbf{W})$ as

$$\begin{aligned} \Psi(\mathbf{S}, \mathbf{G}, \mathbf{W}) = & \text{tr} \left\{ \mathbf{W} \mathbf{H}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \right. \\ & \times \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \mathbf{G}^H \mathbf{H}_r^H \mathbf{W}^H \left. \right\} \\ & - \text{tr} \left\{ \mathbf{W} \mathbf{H}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} \right\} \\ & + \text{tr} \left\{ \mathbf{R}_{h_s} \right\} \\ & - \text{tr} \left\{ \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \mathbf{G}^H \mathbf{H}_r^H \mathbf{W}^H \right\} \\ & + \text{tr} \left\{ \mathbf{W} \left(\mathbf{H}_r \mathbf{G} \mathbf{G}^H \mathbf{H}_r^H \sigma_{v_s}^2 + \sigma_{v_r}^2 \mathbf{I}_{KN_d} \right) \mathbf{W}^H \right\}, \end{aligned} \quad (19)$$

where we define $\mathbf{R}_{h_s} \triangleq \mathbb{E}\{\mathbf{h}_s \mathbf{h}_s^H\} \in \mathbb{C}^{L_{h_s} \times L_{h_s}}$

$$\mathbf{R}_{h_s} = \text{diag} \left[\left\{ \sigma_{h_s}^2 [l] \boldsymbol{\Theta}_s [l] \otimes \boldsymbol{\Upsilon}_s [l] \right\}_{l=0}^L \right] \quad (20)$$

as the source-relay channel covariance matrix. Unfortunately, we find from (19) that minimising $\Psi(\mathbf{S}, \mathbf{G}, \mathbf{W})$ can only be conducted if the relay-destination channel matrix \mathbf{H}_r in (15) is precisely known.

In order to formulate an analytically tractable optimisation problem we make the simplifying assumption that the estimate of the relay-destination OFDM channel is sufficiently accurate. This is a valid assumption when the signal to noise ratio (SNR) during the first phase of channel estimation is sufficiently high such that relay-destination channel estimation errors are small enough to be negligible [21]. We therefore consider the cost function $\hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W})$ given by

$$\begin{aligned} \hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W}) = & \text{tr} \left\{ \mathbf{W} \hat{\mathbf{H}}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \right. \\ & \times \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \mathbf{G}^H \hat{\mathbf{H}}_r^H \mathbf{W}^H \left. \right\} \\ & - \text{tr} \left\{ \mathbf{W} \hat{\mathbf{H}}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} \right\} \\ & + \text{tr} \left\{ \mathbf{R}_{h_s} \right\} \\ & - \text{tr} \left\{ \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \mathbf{G}^H \hat{\mathbf{H}}_r^H \mathbf{W}^H \right\} \\ & + \text{tr} \left\{ \mathbf{W} \left(\hat{\mathbf{H}}_r \mathbf{G} \mathbf{G}^H \hat{\mathbf{H}}_r^H \sigma_{v_s}^2 + \sigma_{v_r}^2 \mathbf{I}_{KN_d} \right) \mathbf{W}^H \right\}, \end{aligned} \quad (21)$$

which is obtained from (19) simply by replacing \mathbf{H}_r with the corresponding channel estimate $\hat{\mathbf{H}}_r$.

As well as minimising the cost function $\hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W})$, the source and relay transmit powers should be constrained since these devices will have limited power budgets. The source transmit power constraint is given by

$$\text{tr} \left\{ \mathbf{S} \mathbf{S}^H \right\} \leq P_s, \quad (22)$$

with P_s being the power budget available to the source device. The relay transmit signal in (13) should also abide by a power constraint. Since (13) depends on the unknown noise signal \mathbf{v}_s as well as the channel vector \mathbf{h}_s we shall enforce an expected relay transmit power constraint given by $\text{tr}\{\mathbb{E}\{\hat{\mathbf{r}}\hat{\mathbf{r}}^H\}\} \leq P_r$, where the expectation is taken w.r.t. \mathbf{v}_s and \mathbf{h}_s , and P_r is the

power budget available to the relay. The relay power constraint can thus be written as

$$\text{tr} \left\{ \mathbf{G} \mathbf{T} \mathbf{G}^H \right\} \leq P_r, \quad (23)$$

where $\mathbf{T} \triangleq \mathbb{E}\{\mathbf{r} \mathbf{r}^H\} \in \mathbb{C}^{KN_r \times KN_r}$ and is given by

$$\mathbf{T} = (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H + \sigma_{v_s}^2 \mathbf{I}_{KN_r}. \quad (24)$$

We can now construct the optimisation problem for finding the source-relay channel estimate as

$$\min_{\mathbf{S}, \mathbf{G}, \mathbf{W}} \hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W}) \quad (25)$$

$$\text{s.t.} \quad \text{tr} \left\{ \mathbf{S} \mathbf{S}^H \right\} \leq P_s \quad (26)$$

$$\text{tr} \left\{ \mathbf{G} \mathbf{T} \mathbf{G}^H \right\} \leq P_r. \quad (27)$$

The optimisation problem in (25)–(27) is non-convex and obtaining the optimal solution in closed form is intractable. In the following sections we propose several algorithms for overcoming this problem.

IV. PROPOSED ITERATIVE ALGORITHM

It can be shown that with any two of the variables fixed the problem in (25)–(27) is convex w.r.t. the remaining variable. In fact, the individual design problems for the source training matrix, the relay precoder, and the destination processor can be formulated as (possibly) constrained quadratic matrix problems with matrix variables. With this observation we therefore suggest an iterative algorithm to solve (25)–(27) by sequentially updating each variable. Since each subproblem is a convex optimisation problem, iteratively updating each variable in a sequential fashion allows the algorithm to converge to (at least) a locally optimal solution. It is interesting to note that the proposed iterative algorithm is not too dissimilar in nature to iterative linear MMSE transceiver designs for MIMO relay systems (see [29]), and similar tools can be utilised to solve these problems.

A. Training Matrix Update

We focus firstly on solving (25)–(27) for updating the source training matrix \mathbf{S} when both \mathbf{G} and \mathbf{W} are fixed. To this end we note that $\hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W})$ can be equivalently written as

$$\begin{aligned} \hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W}) = & \left\| \left(\mathbf{W} \hat{\mathbf{H}}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) - \mathbf{I}_{L_{h_s}} \right) \mathbf{R}_{h_s}^{1/2} \right\|_F^2 \\ & + \left\| \mathbf{W} \left(\hat{\mathbf{H}}_r \mathbf{G} \mathbf{G}^H \hat{\mathbf{H}}_r^H \sigma_{v_s}^2 + \sigma_{v_r}^2 \mathbf{I}_{KN_d} \right)^{1/2} \right\|_F^2. \end{aligned} \quad (28)$$

Similarly, we can equivalently express the source and relay power constraints in (26) and (27) as

$$\|\mathbf{S}\|_F^2 \leq P_s \quad (29)$$

$$\left\| \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s}^{1/2} \right\|_F^2 \leq P_r - \|\mathbf{G} \sigma_{v_s}\|_F^2, \quad (30)$$

where to obtain (30) we have used the definition of \mathbf{T} in (24). Using (28)–(30) we can rewrite the optimisation problem in (25)–(27) for finding the training matrix \mathbf{S} as

$$\min_{\mathbf{S}} \left\| \left(\mathbf{W} \hat{\mathbf{H}}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) - \mathbf{I}_{L_{h_s}} \right) \mathbf{R}_{h_s}^{1/2} \right\|_F^2 \quad (31)$$

$$\text{s.t.} \quad \|\mathbf{S}\|_F^2 \leq P_s \quad (32)$$

$$\left\| \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s}^{1/2} \right\|_F^2 \leq P_r - \|\mathbf{G} \sigma_{v_s}\|_F^2. \quad (33)$$

We note that the contribution of the second term in (28) has been eliminated from the objective function in (31) since it is not a function of the training matrix \mathbf{S} . Introducing an auxiliary variable t that satisfies

$$t \geq \left\| \left(\mathbf{W} \hat{\mathbf{H}}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) - \mathbf{I}_{L_{h_s}} \right) \mathbf{R}_{h_s}^{1/2} \right\|_F, \quad (34)$$

we can solve (31)–(33) from the following second order conic program (SOCP) [30]

$$\min_{\mathbf{S}, t} t \quad (35)$$

$$\text{s.t.} \quad \|\mathbf{S}\|_F \leq \sqrt{P_s} \quad (36)$$

$$\left\| \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s}^{1/2} \right\|_F \leq \sqrt{P_r - \|\mathbf{G} \sigma_{v_s}\|_F^2} \quad (37)$$

$$\left\| \left(\mathbf{W} \hat{\mathbf{H}}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}) - \mathbf{I}_{L_{h_s}} \right) \mathbf{R}_{h_s}^{1/2} \right\|_F \leq t. \quad (38)$$

The SOCP in (35)–(38) is a standard convex optimisation problem and therefore the optimal solution can be efficiently found using interior point algorithms [30].

B. Relay Precoder Update

We now focus on updating the relay precoder by solving (25)–(27) for \mathbf{G} when both the source training matrix \mathbf{S} and the destination processor \mathbf{W} are fixed. Since the power constraint in (26) does not depend on \mathbf{G} we can find the relay precoder by solving

$$\min_{\mathbf{G}} \hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W}) \quad (39)$$

$$\text{s.t.} \quad \text{tr} \left\{ \mathbf{G} \mathbf{T} \mathbf{G}^H \right\} \leq P_r. \quad (40)$$

This optimisation problem is a quadratic matrix problem with a single constraint and is a standard convex optimisation problem. It is worthwhile mentioning that the optimisation problem for the relay precoder in (39)–(40) can be reformulated as a SOCP and solved using interior point methods. However, since there is only one constraint, a simpler solution can be derived from the KKT conditions [30]. The following KKT conditions are sufficient for optimality:

$$\left(\hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{H}}_r + \mu_r \mathbf{I}_{K_{N_r}} \right) \mathbf{G} \mathbf{T} = \hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \quad (41)$$

$$\text{tr} \left\{ \mathbf{G} \mathbf{T} \mathbf{G}^H \right\} - P_r \leq 0 \quad (42)$$

$$\mu_r \left(\text{tr} \left\{ \mathbf{G} \mathbf{T} \mathbf{G}^H \right\} - P_r \right) = 0 \quad (43)$$

$$\mu_r \geq 0, \quad (44)$$

where μ_r is the Lagrangian multiplier associated with the relay power constraint in (40), or equivalently that in (42). Solving (41) results in the optimal relay precoder given by

$$\mathbf{G} = \left(\hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{H}}_r + \mu_r \mathbf{I}_{K_{N_r}} \right)^{-1} \times \hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \mathbf{T}^{-1}. \quad (45)$$

The Lagrangian multiplier μ_r must now be calculated to ensure the power constraint in (40) is satisfied. Substituting (45) into (42) and (43) we require to find μ_r that satisfies

$$\text{tr} \left\{ \left(\hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{H}}_r + \mu_r \mathbf{I}_{K_{N_r}} \right)^{-2} \mathbf{X} \right\} \leq P_r, \quad (46)$$

$$\mu_r \left(\text{tr} \left\{ \left(\hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{H}}_r + \mu_r \mathbf{I}_{K_{N_r}} \right)^{-2} \mathbf{X} \right\} - P_r \right) = 0 \quad (47)$$

where for convenience we define

$$\mathbf{X} \triangleq \hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \mathbf{T}^{-1} \times (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} \mathbf{W} \hat{\mathbf{H}}_r. \quad (48)$$

The condition in (47) can be satisfied with either $\mu_r = 0$ or $\text{tr} \left\{ \left(\hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{H}}_r + \mu_r \mathbf{I}_{K_{N_r}} \right)^{-2} \mathbf{X} \right\} = P_r$. If $\mu_r = 0$ satisfies the condition in (46) then, since the left hand side of (46) is a monotonically decreasing function of μ_r , it is the only solution to satisfy both (46) and (47). On the other hand, if $\mu_r = 0$ does not satisfy (46) then we must compute a positive μ_r to satisfy both (46) and (47). To satisfy (46) and (47) in this case we require to compute μ_r such that $\text{tr} \left\{ \left(\hat{\mathbf{H}}_r^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{H}}_r + \mu_r \mathbf{I}_{K_{N_r}} \right)^{-2} \mathbf{X} \right\} = P_r$. Based on the fact that the left hand side of this expression is a monotonically decreasing function of μ_r , the method of bisection can be used to find μ_r in this case.

C. Destination Processor Update

We finally focus on updating the destination processor by solving (25)–(27) for \mathbf{W} when both the relay precoder \mathbf{G} and the source training matrix \mathbf{S} are fixed. Since the source and relay power constraints in (26) and (27) are independent of \mathbf{W} , the destination processor can be found by solving the unconstrained optimisation problem

$$\min_{\mathbf{W}} \hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W}). \quad (49)$$

From $\hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W})$ given in (21) it is straightforward to show that $\hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W})$ is convex in \mathbf{W} . The optimal solution to the unconstrained problem in (49) can therefore be found by setting the derivative of $\hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W})$ w.r.t. \mathbf{W}^* to zero and solving the result for \mathbf{W} . This leads to the optimal destination processor being given by

$$\mathbf{W} = \mathbf{R}_{h_s} (\mathbf{M} \otimes \mathbf{I}_{N_r})^H \mathbf{G}^H \hat{\mathbf{H}}_r^H \times \left(\hat{\mathbf{H}}_r \mathbf{G} \mathbf{T} \mathbf{G}^H \hat{\mathbf{H}}_r^H + \sigma_{v_r}^2 \mathbf{I}_{K_{N_d}} \right)^{-1}. \quad (50)$$

D. Summary of Iterative Algorithm

We now briefly summarise the proposed iterative algorithm. For the i th iteration, let us denote \mathbf{S}_i , \mathbf{G}_i , and \mathbf{W}_i , as the updates of the source training matrix, relay precoder, and destination processor, respectively. The main steps of the proposed iterative algorithm are shown in Algorithm 1:

Algorithm 1: Iterative algorithm for source-relay channel estimation.

Initialisation: Set $i = 0$. Initialise \mathbf{S}_0 and \mathbf{G}_0 to satisfy (26) and (27) and initialise \mathbf{W}_0 .

repeat

Set $i = i + 1$.

Update \mathbf{S}_i by solving the SOCP in (35)–(38).

Compute \mathbf{X} using (48).

if $\text{tr}\{(\hat{\mathbf{H}}_r^H \mathbf{W}_{i-1}^H \mathbf{W}_{i-1} \hat{\mathbf{H}}_r)^{-2} \mathbf{X}\} \leq P_r$ **then**

Set $\mu_{r,i} = 0$.

else

Solve $\text{tr}\{(\hat{\mathbf{H}}_r^H \mathbf{W}_{i-1}^H \mathbf{W}_{i-1} \hat{\mathbf{H}}_r + \mu_{r,i} \mathbf{I}_{KN_r})^{-2} \mathbf{X}\} = P_r$ for $\mu_{r,i}$ using the method of bisection.

end if

Update \mathbf{G}_i using (45).

Update \mathbf{W}_i using (50).

until $|\hat{\Psi}(\mathbf{S}_i, \mathbf{G}_i, \mathbf{W}_i) - \hat{\Psi}(\mathbf{S}_{i-1}, \mathbf{G}_{i-1}, \mathbf{W}_{i-1})| \leq \epsilon$ or $i = \text{maxiter}$.

Compute the source-relay channel estimate $\hat{\mathbf{h}}_s$ from (17).

As highlighted in Algorithm 1, the variables \mathbf{S}_i , \mathbf{G}_i , and \mathbf{W}_i are repeatedly updated until $|\hat{\Psi}(\mathbf{S}_i, \mathbf{G}_i, \mathbf{W}_i) - \hat{\Psi}(\mathbf{S}_{i-1}, \mathbf{G}_{i-1}, \mathbf{W}_{i-1})|$ falls below some threshold $\epsilon \in \mathbb{R}_+$, or until a maximum number of prespecified iterations are reached. Since the design variables are found by solving convex sub-problems, the channel estimation MSE can only decrease or maintain after each update and convergence is guaranteed.

V. PROPOSED SIMPLIFIED ALGORITHMS

As has been previously established, for any given \mathbf{S} and \mathbf{G} , the optimal destination processing matrix \mathbf{W} can be found by solving the unconstrained problem in (49) and the solution is given by (50). Substituting (50) into (21), and using the matrix inversion lemma, we can express the source-relay channel estimation MSE as in (51), shown at the bottom of the page. To obtain the matrix \mathbf{E}_1 in (51) we have also used the facts that $(\mathbf{A} \otimes \mathbf{B})^H = (\mathbf{A}^H \otimes \mathbf{B}^H)$ and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$, for matrices of commensurate dimensions. We

can now formulate the joint training matrix and relay precoder design problem as

$$\min_{\mathbf{S}, \mathbf{G}} \text{tr}\{\mathbf{E}_1\} + \text{tr}\{\mathbf{E}_2\} \quad (52)$$

$$\text{s.t.} \quad \text{tr}\{\mathbf{S}\mathbf{S}^H\} \leq P_s \quad (53)$$

$$\text{tr}\{\mathbf{G}\mathbf{T}\mathbf{G}^H\} \leq P_r. \quad (54)$$

The optimisation problem in (52)–(54) is not jointly convex in the design variables \mathbf{S} and \mathbf{G} and thus the optimal solution is difficult to obtain. In the following we therefore focus on simplified approaches to solving (52)–(54), which have reduced computational complexity compared to the iterative algorithm proposed in the previous section.

A. Optimal Relay Precoder Structure

We begin by deriving the optimal structure of the relay precoder \mathbf{G} as the solution to (52)–(54). Since \mathbf{E}_1 and the source power constraint in (53) are both independent of \mathbf{G} , we can calculate the optimal relay precoder structure by solving

$$\min_{\mathbf{G}} \text{tr}\{\mathbf{E}_2\} \quad (55)$$

$$\text{tr}\{\mathbf{G}\mathbf{T}\mathbf{G}^H\} \leq P_r. \quad (56)$$

Theorem 1: The optimal structure of the relay precoder \mathbf{G} as the solution to the problem in (55)–(56) is given by

$$\mathbf{G} = \mathbf{L}\mathbf{R}_{h_s} \left(\mathbf{M}^H \otimes \mathbf{I}_{N_r} \right) \mathbf{T}^{-1}, \quad (57)$$

where $\mathbf{L} \in \mathbb{C}^{KN_r \times L_{h_s}}$ is a matrix yet to be determined.

Proof: See Appendix A. \blacksquare

We see that the optimal relay precoder given in (57) can be decomposed into two main components. Specifically, it consists of a transmit/precoding matrix \mathbf{L} and a receiver matrix given by $\mathbf{R}_{h_s}(\mathbf{M}^H \otimes \mathbf{I}_{N_r})\mathbf{T}^{-1}$. Interestingly, given the relay received signal in (9), it can be straightforwardly shown that this receiver matrix is in fact the optimal linear matrix for producing the MMSE channel estimate at the relay. The action of the relay precoder is therefore to firstly produce a MMSE source-relay channel estimate, before precoding this estimate by \mathbf{L} and forwarding the result to the destination.

$$\hat{\Psi}(\mathbf{S}, \mathbf{G}, \mathbf{W}) = \text{tr} \left\{ \underbrace{\left(\mathbf{R}_{h_s}^{-1} + \sigma_{v_s}^{-2} \left(\mathbf{M}^H \mathbf{M} \otimes \mathbf{I}_{N_r} \right) \right)^{-1}}_{\mathbf{E}_1} \right\} + \text{tr} \left\{ \underbrace{\mathbf{T}^{-1/2} \left(\mathbf{M} \otimes \mathbf{I}_{N_r} \right) \mathbf{R}_{h_s} \mathbf{R}_{h_s} \left(\mathbf{M}^H \otimes \mathbf{I}_{N_r} \right) \mathbf{T}^{-1/2} \left(\sigma_{v_r}^{-2} \mathbf{T}^{1/2} \mathbf{G}^H \hat{\mathbf{H}}_r^H \hat{\mathbf{H}}_r \mathbf{G} \mathbf{T}^{1/2} + \mathbf{I}_{KN_r} \right)^{-1}}_{\mathbf{E}_2} \right\}. \quad (51)$$

Substituting the relay precoder structure of (57) into $\text{tr}\{\mathbf{E}_2\}$ in (51) and using the matrix inversion lemma we can write

$$\text{tr}\{\mathbf{E}_2\} = \text{tr} \left\{ \left(\sigma_{v_r}^{-2} \mathbf{L}^H \hat{\mathbf{H}}_r^H \hat{\mathbf{H}}_r \mathbf{L} + \mathbf{N}^{-1} \right)^{-1} \right\}, \quad (58)$$

where for convenience we define

$$\mathbf{N} \triangleq \mathbf{R}_{h_s} \left(\mathbf{M}^H \otimes \mathbf{I}_{N_r} \right) \mathbf{T}^{-1} \left(\mathbf{M} \otimes \mathbf{I}_{N_r} \right) \mathbf{R}_{h_s}. \quad (59)$$

For high SNR of the source-relay link, i.e. when we have $(\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} (\mathbf{M}^H \otimes \mathbf{I}_{N_r}) \gg \mathbf{I}_{KN_r}$, it is straightforward to show from (59) that \mathbf{N} approaches \mathbf{R}_{h_s} . In such a high SNR environment the source training matrix \mathbf{S} does not impact $\text{tr}\{\mathbf{E}_2\}$ in (58). The optimisation of \mathbf{S} in this case can then be computed independently of the relay precoder and we can therefore approximate and decompose the original optimisation problem in (52)–(54) into two separate problems. Specifically, using the optimal relay precoder structure in (57) as well as the matrix \mathbf{E}_1 defined in (51), we can approximate and decompose the problem (52)–(54) into the following source training matrix optimisation problem

$$\min_{\mathbf{S}} \text{tr} \left\{ \left(\mathbf{R}_{h_s}^{-1} + \sigma_{v_s}^{-2} \left(\mathbf{M}^H \mathbf{M} \otimes \mathbf{I}_{N_r} \right) \right)^{-1} \right\} \quad (60)$$

$$\text{s.t.} \quad \text{tr} \left\{ \mathbf{S} \mathbf{S}^H \right\} \leq P_s, \quad (61)$$

and relay precoder matrix optimisation problem

$$\min_{\mathbf{L}} \text{tr} \left\{ \left(\sigma_{v_r}^{-2} \mathbf{L}^H \hat{\mathbf{H}}_r^H \hat{\mathbf{H}}_r \mathbf{L} + \mathbf{N}^{-1} \right)^{-1} \right\} \quad (62)$$

$$\text{s.t.} \quad \text{tr} \left\{ \mathbf{L} \mathbf{N} \mathbf{L}^H \right\} \leq P_r. \quad (63)$$

The decomposition of the original problem in (52)–(54) using the high SNR approximation greatly simplifies the optimisation procedure since the training matrix \mathbf{S} can firstly be found by solving (60)–(61) and the matrix \mathbf{L} can subsequently be found by solving (62)–(63). Unlike the algorithm proposed in Section IV there is no need to iteratively update the variables in this case, which results in the proposed simplified algorithms having substantially reduced computational complexity compared to the iterative algorithm.

B. Proposed Simplified Algorithm 1

The source training matrix optimisation problem in (60)–(61) is equivalent to the point-to-point MIMO OFDM channel estimation problem considered in [12] and the optimal solution to (60)–(61) can be found using similar arguments to those made in [12]. Specifically, the optimal structure of the source training matrix is given by

$$\mathbf{S} = \mathbf{Q} \bar{\mathbf{S}}, \quad (64)$$

where $\bar{\mathbf{S}} \in \mathbb{C}^{N_s \times N_s}$ is a matrix yet to be determined and $\mathbf{Q} \in \mathbb{C}^{K \times N_s}$ is a semi-unitary matrix that satisfies

$$\mathbf{Q}^H \mathbf{F}_m^H \mathbf{F}_n \mathbf{Q} = \mathbf{0}_{N_s \times N_s} \quad \forall m \neq n \quad (65)$$

$$\mathbf{Q}^H \mathbf{F}_l^H \mathbf{F}_l \mathbf{Q} = \mathbf{I}_{N_s} \quad \forall l \quad (66)$$

$$\mathbf{Q}^H \mathbf{Q} = \mathbf{I}_{N_s}, \quad (67)$$

where \mathbf{F}_l were defined in (11). The semi-unitary matrix \mathbf{Q} that satisfies the properties in (65)–(67) can be constructed as follows: Partition the matrix \mathbf{Q} as $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_{N_s}]$, where $\mathbf{q}_i \in \mathbb{C}^K$ denotes the i th column of \mathbf{Q} . Let the first column of \mathbf{Q} be given by $\mathbf{q}_1 = \sqrt{1/K} \mathbf{1}_K$, where $\mathbf{1}_K$ is a K dimensional column vector with all elements equal to one, which satisfies $\|\mathbf{q}_1\| = 1$. The k th element of the remaining columns of \mathbf{Q} can then be constructed as $[\mathbf{q}_i]_k = e^{-j2\pi \lfloor K/N_r \rfloor (i-1)(k-1)/K}$.

Substituting (20) and (64) into (60)–(61), then after introducing the variable change

$$\hat{\mathbf{S}} = \bar{\mathbf{S}}^H \bar{\mathbf{S}}, \quad (68)$$

the optimisation problem in (60)–(61) is equivalent to the following problem in $\hat{\mathbf{S}}$

$$\min_{\hat{\mathbf{S}}} \sum_{l=0}^L \text{tr} \left\{ \left(\sigma_{h_s}^{-2} [l] \Theta_s^{-1} [l] \otimes \Upsilon_s^{-1} [l] + \sigma_{v_s}^{-2} \hat{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \quad (69)$$

$$\text{s.t.} \quad \text{tr} \left\{ \hat{\mathbf{S}} \right\} \leq P_s \quad (70)$$

$$\hat{\mathbf{S}} \geq \mathbf{0}_{N_s \times N_s}. \quad (71)$$

To obtain the objective function in (69) we have utilised the fact that the matrix \mathbf{Q} in (64) satisfies the conditions in (65)–(67), as well as the properties $(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})$ and $(\text{diag}[\{\mathbf{A}_n\}_{n=1}^N])^{-1} = \text{diag}[\{\mathbf{A}_n^{-1}\}_{n=1}^N]$. Introducing auxilliary matrices $\mathbf{K}[l] \in \mathbb{C}^{N_s N_r \times N_s N_r}$ satisfying

$$\mathbf{K}[l] \geq \left(\sigma_{h_s}^{-2} [l] \Theta_s^{-1} [l] \otimes \Upsilon_s^{-1} [l] + \sigma_{v_s}^{-2} \hat{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1}, \quad (72)$$

then by utilising the Schur complement lemma the problem in (69)–(71) can be solved by the following SDP [30]

$$\min_{\hat{\mathbf{S}}, \{\mathbf{K}[l]\}_{l=0}^L} \sum_{l=0}^L \text{tr} \{ \mathbf{K}[l] \} \quad (73)$$

$$\text{s.t.} \quad \text{tr} \left\{ \hat{\mathbf{S}} \right\} \leq P_s \quad (74)$$

$$\hat{\mathbf{S}} \geq \mathbf{0}_{N_s \times N_s}, \quad (75)$$

$$\begin{bmatrix} \mathbf{K}[l] & \mathbf{I}_{N_s N_r} \\ \mathbf{I}_{N_s N_r} & \mathbf{R}_{h_s}^{-1} [l] + \sigma_{v_s}^{-2} \hat{\mathbf{S}} \otimes \mathbf{I}_{N_r} \end{bmatrix} \geq \mathbf{0}_{M \times M}, \quad (76)$$

where $\mathbf{R}_{h_s} [l] \triangleq \sigma_{h_s}^2 [l] \Theta_s [l] \otimes \Upsilon_s [l]$ and $M \triangleq 2N_s N_r$. The SDP in (73)–(76) is a standard convex optimisation problem and thus the optimal solution can be efficiently found using interior point algorithms [30]. After solving (73)–(76) for $\hat{\mathbf{S}}$, the matrix $\bar{\mathbf{S}}$ can be computed from (68) as $\bar{\mathbf{S}} = \hat{\mathbf{S}}^{1/2}$, and finally the optimal training matrix \mathbf{S} is given by (64).

Having solved (60)–(61) for the optimal training matrix \mathbf{S} , we now turn our attention to deriving the optimal matrix \mathbf{L} as the solution to the optimisation problem in (62)–(63). We firstly restate the problem in (62)–(63) as

$$\min_{\mathbf{L}} \text{tr} \left\{ \mathbf{N} \left(\sigma_{v_r}^{-2} \mathbf{N}^{1/2} \mathbf{L}^H \hat{\mathbf{H}}_r^H \hat{\mathbf{H}}_r \mathbf{L} \mathbf{N}^{1/2} + \mathbf{I}_{L_{h_s}} \right)^{-1} \right\} \quad (77)$$

$$\text{s.t.} \quad \text{tr} \left\{ \mathbf{L} \mathbf{N} \mathbf{L}^H \right\} \leq P_r. \quad (78)$$

Let us now consider the singular value decomposition (SVD) of $\hat{\mathbf{H}}_r$ and eigenvalue decomposition (EVD) of N given by

$$\hat{\mathbf{H}}_r = \mathbf{U}_r \mathbf{\Delta} \mathbf{V}_r^H \quad (79)$$

$$N = \mathbf{U}_n \mathbf{\Psi} \mathbf{U}_n^H, \quad (80)$$

where $\mathbf{\Delta}$ is a $KN_d \times KN_r$ diagonal matrix containing the non-zero singular values $\{\delta_i\}_{i=1}^{R_r} \in \mathbb{R}_{++}$, and $\mathbf{\Psi}$ is an $L_{h_s} \times L_{h_s}$ diagonal matrix that contains the non-zero eigenvalues $\{\psi_i\}_{i=1}^{R_n} \in \mathbb{R}_{++}$. Here we define $R_r \triangleq \text{rank}\{\hat{\mathbf{H}}_r\}$ and $R_n \triangleq \text{rank}\{N\}$, and assume w.l.o.g. that the diagonal entries in $\mathbf{\Delta}$ and $\mathbf{\Psi}$ are arranged in descending order.

Theorem 2: The optimal relay precoding matrix \mathbf{L} as the solution to (77)–(78) is given by

$$\mathbf{L} = \mathbf{V}_r \mathbf{\Phi} \mathbf{U}_n^H, \quad (81)$$

where $\mathbf{\Phi}$ is a $KN_r \times L_{h_s}$ diagonal matrix with non-negative diagonal elements $\{\phi_i\}_{i=1}^R \in \mathbb{R}_+$. Here we define the variable $R \triangleq \min(R_r, R_n)$.

Proof: See Appendix B. \blacksquare

With the matrix \mathbf{L} given in (81) the optimal relay precoder is finally given from (57) by

$$\mathbf{G} = \mathbf{V}_r \mathbf{\Phi} \mathbf{U}_n^H \mathbf{R}_{h_s} \left(\mathbf{M}^H \otimes \mathbf{I}_{N_r} \right) \mathbf{T}^{-1}. \quad (82)$$

Substituting (79)–(81) into (62)–(63), the original matrix valued optimisation problem reduces to the scalar problem

$$\min_{\{\phi_i\}_{i=1}^R} \sum_{i=1}^R \frac{\psi_i \sigma_{v_r}^2}{\phi_i^2 \delta_i^2 \psi_i + \sigma_{v_r}^2} \quad (83)$$

$$\text{s.t.} \quad \sum_{i=1}^R \phi_i^2 \psi_i \leq P_r. \quad (84)$$

$$\phi_i \geq 0, \quad i = 1, \dots, R. \quad (85)$$

This problem has a standard waterfilling solution, which can be obtained from the KKT conditions [30], and is given by

$$\phi_i^2 = \frac{1}{\delta_i^2 \psi_i} \left[\sqrt{\frac{\delta_i^2 \psi_i \sigma_{v_r}^2}{\mu_r} - \sigma_{v_r}^2} \right]^+ \quad (86)$$

Substituting (86) into (84) we now require to calculate the waterlevel μ_r to satisfy the non-linear equation

$$\sum_{i=1}^R \frac{1}{\delta_i^2} \left[\sqrt{\frac{\delta_i^2 \psi_i \sigma_{v_r}^2}{\mu_r} - \sigma_{v_r}^2} \right]^+ \leq P_r, \quad (87)$$

which can be obtained using the waterfilling algorithm in [31].

We now briefly summarise the main steps for computing the source-relay channel estimate using the simplified algorithm. The matrix $\hat{\mathbf{S}}$ is firstly computed by solving the SDP (73)–(76) from which we then obtain $\bar{\mathbf{S}} = \hat{\mathbf{S}}^{1/2}$. The source training matrix \mathbf{S} is then given by (64). Having calculated \mathbf{S} , we then compute the relay precoding matrix \mathbf{G} from (82), where the elements of $\mathbf{\Phi}$ are computed according to (86). Finally, the destination processor \mathbf{W} is given by (50) and the source-relay channel estimate is obtained from (17).

C. Proposed Simplified Algorithm 2

Although the previously discussed simplified algorithm allows the source training matrix to be computed independently of the relay precoder, and is therefore more computationally efficient than the iterative algorithm in Section IV, it requires solving the SDP in (73)–(76). The channel estimation algorithm can be simplified further by deriving a suboptimal solution to the training matrix design problem in (60)–(61).

Substituting (64) into (60)–(61) the training matrix design problem can be stated as

$$\min_{\bar{\mathbf{S}}} \sum_{l=0}^L \text{tr} \left\{ \left(\mathbf{R}_{h_s}^{-1}[l] + \sigma_{v_s}^{-2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \quad (88)$$

$$\text{s.t.} \quad \text{tr} \left\{ \bar{\mathbf{S}} \bar{\mathbf{S}}^H \right\} \leq P_s, \quad (89)$$

where $\mathbf{R}_{h_s}[l] \triangleq \sigma_{h_s}^2[l] \mathbf{\Theta}_s[l] \otimes \mathbf{\Upsilon}_s[l]$.

We now consider a suboptimal solution to (88)–(89) that can be obtained in closed form. To this end it is shown in Appendix C that an upper bound to the training matrix objective function in (88) is

$$\sum_{l=0}^L \text{tr} \left\{ \left(\mathbf{R}_{h_s}^{-1}[l] + \sigma_{v_s}^{-2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \leq \text{tr} \left\{ \left(\bar{\mathbf{\Theta}}_s^{-1} \otimes \bar{\mathbf{\Upsilon}}_s^{-1} + \sigma_{v_s}^{-2} (L+1)^{-1} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\}, \quad (90)$$

where we define the matrices

$$\bar{\mathbf{\Theta}}_s \triangleq \sum_{l=0}^L \sigma_{h_s}^2[l] \mathbf{\Theta}_s[l] \quad (91)$$

$$\bar{\mathbf{\Upsilon}}_s \triangleq \text{diag} \left[\left\{ \max \left(\{v_{r,i}[l]\}_{l=0}^L \right) \right\}_{i=1}^{N_r} \right]. \quad (92)$$

Replacing the objective function in (88) with the upper bound in (90), we can obtain a suboptimal solution to the optimisation problem in (88)–(89) by solving

$$\min_{\bar{\mathbf{S}}} \text{tr} \left\{ \left(\bar{\mathbf{\Theta}}_s^{-1} \otimes \bar{\mathbf{\Upsilon}}_s^{-1} + \sigma_{v_s}^{-2} (L+1)^{-1} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \quad (93)$$

$$\text{s.t.} \quad \text{tr} \left\{ \bar{\mathbf{S}} \bar{\mathbf{S}}^H \right\} \leq P_s. \quad (94)$$

Before solving (93)–(94), let us introduce the EVD

$$\bar{\mathbf{\Theta}}_s = \mathbf{V}_{\bar{\Theta}_s} \mathbf{\Xi} \mathbf{V}_{\bar{\Theta}_s}^H, \quad (95)$$

where $\mathbf{\Xi}$ is an $N_s \times N_s$ diagonal matrix that contains the positive eigenvalues $\{\xi_i\}_{i=1}^{N_s} \in \mathbb{R}_{++}$.

Theorem 3: The optimal solution to the optimisation problem in (93)–(94) is given by

$$\bar{\mathbf{S}} = \mathbf{\Gamma} \mathbf{V}_{\bar{\Theta}_s}^H, \quad (96)$$

where $\mathbf{\Gamma}$ is a diagonal matrix of dimension $N_s \times N_s$ with non-negative diagonal entries $\{\gamma_i\}_{i=1}^{N_s} \in \mathbb{R}_+$.

Proof: See Appendix D. \blacksquare

Substituting (96) into (64) the suboptimal structure for the training matrix \mathbf{S} as the solution to the original problem in (60)–(61) is given by

$$\mathbf{S} = \mathbf{Q}\mathbf{G}\mathbf{V}\mathbf{\Theta}_s^H. \quad (97)$$

We would like to mention that this result coincides with the suboptimal training matrix solution derived in [12], although a different method of derivation was employed in [12].

Substituting (96) and (95) into (93)–(94) we can restate the matrix valued optimisation problem as the following scalar problem

$$\min_{\{\gamma_i\}_{i=1}^{N_s}} \sum_{i=1}^{N_s} \sum_{j=1}^{N_r} \frac{\xi_i \sigma_{v_s}^2 (L+1)}{\xi_i \gamma_i^2 + \nu_j^{-1} \sigma_{v_s}^2 (L+1)} \quad (98)$$

$$\text{s.t.} \quad \sum_{i=1}^{N_s} \gamma_i^2 \leq P_s. \quad (99)$$

$$\gamma_i \geq 0, \quad i = 1, \dots, N_s. \quad (100)$$

The optimal solution to (98)–(100) is still difficult to obtain in closed form and we therefore consider a suboptimal solution to this problem. It is straightforward to show that the objective function in (98) is concave in ν_j . We can therefore obtain an upper bound to the objective function by using Jensen's inequality, and a suboptimal solution to (98)–(100) can be obtained by minimising this upper bound. This suboptimal solution to (98)–(100) can be found by solving

$$\min_{\{\gamma_i\}_{i=1}^{N_s}} \sum_{i=1}^{N_s} \frac{N_r \xi_i \sigma_{v_s}^2 (L+1)}{\xi_i \gamma_i^2 + N_r \sigma_{v_s}^2 (L+1) \sum_{j=1}^{N_r} \nu_j^{-1}} \quad (101)$$

$$\text{s.t.} \quad \sum_{i=1}^{N_s} \gamma_i^2 \leq P_s. \quad (102)$$

$$\gamma_i \geq 0, \quad i = 1, \dots, N_s. \quad (103)$$

The problem (101)–(103) can be solved from the KKT conditions of optimality [30] and is given by

$$\gamma_i^2 = \frac{1}{\xi_i} \left[\sqrt{\frac{N_r \xi_i^2 \sigma_{v_s}^2 (L+1)}{\mu_s} - N_r \sigma_{v_s}^2 (L+1) \sum_{j=1}^{N_r} \nu_j^{-1}} \right]^+, \quad (104)$$

where μ_s is the KKT multiplier required to be computed to satisfy the constraint in (102) and can be found using the waterfilling algorithm in [31].

In summary, the proposed simplified source-relay channel estimate can be obtained as follows. The source training matrix \mathbf{S} is firstly given by (97), after which the relay precoder \mathbf{G} can then be computed according to (82). Finally, the destination processor \mathbf{W} is calculated from (50) and the source-relay channel estimate is then obtained from (17).

VI. SIMULATION RESULTS

A. Simulation Setup

We consider a two-hop MIMO OFDM relaying system equipped with N_s , N_r , and N_d antennas at the source, relay,

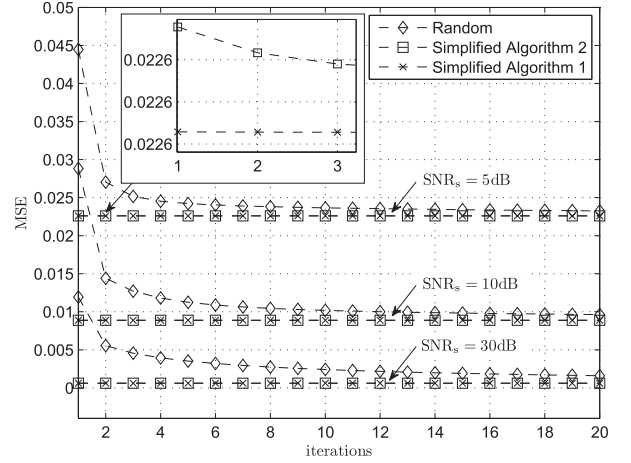


Fig. 1. Convergence of the proposed iterative channel estimation algorithm with different initialisations for $N_s = N_r = N_d = 3$, $\text{SNR}_s = \{5, 10, 30\}$ dB, $\text{SNR}_r = 20$ dB, $\rho_s[l] = \rho_r[l] = \varrho_s[l] = \varrho_r[l] = 0.5, \forall l$, and $\hat{\mathbf{H}}_r = \mathbf{H}_r$.

and destination devices. The frequency selective paths between each transmit and receive antenna pair are considered to be of length $L = 4$ with the l th MIMO channel taps being modelled according to (1) and (2). The transmit spatial correlation matrices $\mathbf{\Theta}_s[l]$ and $\mathbf{\Theta}_r[l]$, and the receive spatial correlation matrices $\mathbf{\Upsilon}_s[l]$ and $\mathbf{\Upsilon}_r[l]$, are modelled using the exponential model (see e.g. [32], [33]) and are given by $[\mathbf{\Theta}_s[l]]_{mn} = \rho_s[l] |m-n|$, $[\mathbf{\Theta}_r[l]]_{mn} = \rho_r[l] |m-n|$, $[\mathbf{\Upsilon}_s[l]]_{mn} = \varrho_s[l] |m-n|$, and $[\mathbf{\Upsilon}_r[l]]_{mn} = \varrho_r[l] |m-n|$, where the correlation co-efficients $\rho_s[l]$, $\rho_r[l]$, $\varrho_s[l]$, and $\varrho_r[l]$, define the level of spatial correlation. The elements of $\mathcal{H}_{sw}[l]$ and $\mathcal{H}_{rw}[l]$ in (1) and (2) are drawn from i.i.d. complex Gaussian distributions with zero mean and variances $\sigma_{h_s}^2[l]$ and $\sigma_{h_r}^2[l]$, and in all simulations we set $\sigma_{h_s}^2[l] = \sigma_{h_r}^2[l] = 1/(L+1), \forall l$.

We consider an OFDM system that utilises $K = 32$ sub-carriers. For channel estimation, the relay-destination channel is estimated in the first phase where the SNR of the relay-destination channel during the first phase of channel estimation is $\text{SNR}_r = \bar{P}_r / (K \sigma_{v_r}^2)$, where \bar{P}_r is the power afforded to the relay for estimation in the first phase. The source-relay channel is estimated during the second phase. During the second phase of channel estimation, the SNRs of the source-relay and relay-destination channels are given by $\text{SNR}_s = P_s / (K \sigma_{v_s}^2)$ and $\text{SNR}_r = P_r / (K \sigma_{v_r}^2)$, respectively.

B. Performance of the Proposed Algorithms

In our first set of simulation examples we assume a system with $N_s = N_r = N_d = 3$ and that the relay-destination channel estimate obtained during the first phase of channel estimation is accurate such that $\hat{\mathbf{H}}_r = \mathbf{H}_r$ can be assumed. This is a practical assumption when the relay-destination SNR during the first training phase is sufficiently high.

The convergence of the iterative algorithm proposed in Section IV is firstly investigated when the source training matrix and relay precoder are initialised as random matrices scaled to satisfy the power constraints, as well as when it is initialised using the solutions for the simplified algorithms presented in Section V-B and Section V-C. Figure 1 shows

the convergence of the algorithm for $\text{SNR}_s = \{5, 10, 30\}$ dB, $\text{SNR}_r = 20$ dB, and $\rho_s[l] = \rho_r[l] = \varrho_s[l] = \varrho_r[l] = 0.5, \forall l$. It can be seen that the algorithm converges after around 5-6 iterations when initialised with random matrices. Interestingly, when the iterative algorithm is initialised using the simplified solutions presented in sections V-B and V-C, only a very small performance improvement can be seen from the initial estimates. This result suggests that the simplified solutions must be close to a local optima (if not the global optima). It is evident that a judicious initialisation of the iterative algorithm is to initialise it using one of the proposed simplified solutions. In all simulation examples henceforth, the iterative algorithm is initialised using the proposed simplified algorithm in Section V-B. Furthermore, in all subsequent simulations, the iterative algorithm is terminated after the difference in MSE between consecutive iterations falls below a threshold of 1×10^{-6} or a maximum of 5 iterations is reached.

We now compare the performance of the proposed solutions to various benchmark LS and MMSE source-relay channel estimation algorithms. The benchmark LS algorithms utilise the optimal least squares destination processor $\mathbf{W} = (\mathbf{H}_r \mathbf{G} (\mathbf{M} \otimes \mathbf{I}_{N_r}))^\dagger$ whilst the MMSE benchmark algorithms employ the optimal MMSE destination processor in (50). In these algorithms the relay precoder is selected as a naive amplify forward (NAF) matrix given by $\mathbf{G} = \alpha \mathbf{I}_{KN_r}$, where $\alpha = \sqrt{P_r / \text{tr}\{(\mathbf{M}_s \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} (\mathbf{M}_s \otimes \mathbf{I}_{N_r})^H + \sigma_{v_s}^2 \mathbf{I}_{KN_r}\}}$ ensures the relay power constraint is met with equality. The benchmark algorithms either utilise random training symbols (RTS) or an equal power allocation (EPA) matrix. The RTS matrix \mathbf{S} contains randomly generated QPSK symbols, whereas the EPA matrix is given by $\mathbf{S} = \sqrt{\text{tr}\{P_s/N_s\}} \mathbf{Q}$ where \mathbf{Q} is the semi-unitary matrix satisfying (65)–(67). The proposed algorithms are also compared to the optimal point-to-point MIMO OFDM channel estimation algorithm developed in [12]. It is important to note that the utilisation of this algorithm assumes the source-relay channel can be estimated at the relay device. In cases where it is practical to perform channel estimation at the relay, the algorithm in [12] provides the optimal solution. However, for cases where the processing cost and power consumption at the relay should be minimised, the proposed algorithms are more appropriate.

Figure 2 shows the results of the proposed and benchmark algorithms against varying SNR_s with the spatial correlation coefficients $\rho_s[l] = \rho_r[l] = \varrho_s[l] = \varrho_r[l] = 0.2, \forall l$. It is observed that the proposed algorithms show improved performance compared to the benchmark LS NAF and MMSE NAF designs across the whole SNR region. Interestingly, the proposed simplified solutions have very similar performance to the proposed iterative algorithm, which is guaranteed to converge to at least a locally optimal solution. The close performance of the simplified solutions to the iterative approach (which was initialised using Simplified Algorithm 1) suggests that the simplified solutions are close to a locally optimal (if not the globally optimal) solution. We also note that, in the high SNR_s region, the proposed algorithms suffer a slight loss in performance compared to the optimal MMSE algorithm in [12]. This is due to the fact that proposed algorithms suffer from noise

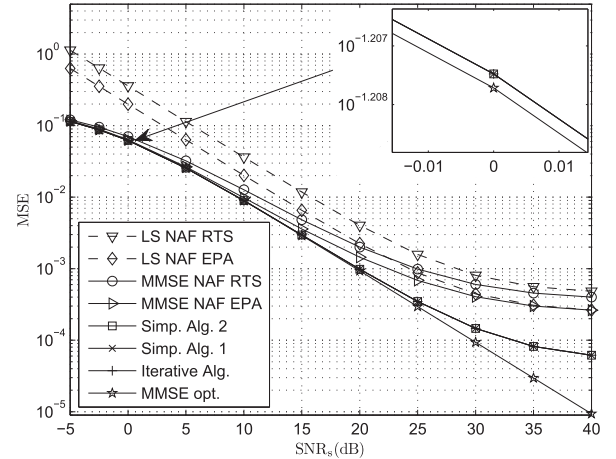


Fig. 2. Source-relay channel estimation MSE against varying SNR_s for the proposed and benchmark algorithms with $N_s = N_r = N_d = 3$, $\text{SNR}_r = 30$ dB, $\rho_s[l] = \rho_r[l] = \varrho_s[l] = \varrho_r[l] = 0.2, \forall l$, and $\hat{\mathbf{H}}_r = \mathbf{H}_r$.

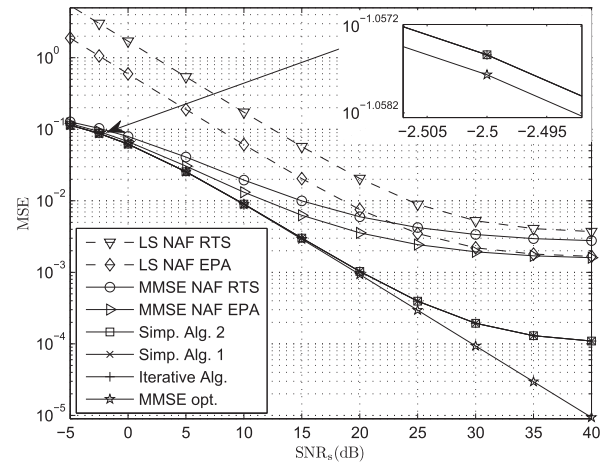


Fig. 3. Source-relay channel estimation MSE against varying SNR_s for the proposed and benchmark algorithms with $N_s = N_r = N_d = 3$, $\text{SNR}_r = 30$ dB, $\rho_s[l] = \varrho_s[l] = 0.2, \rho_r[l] = \varrho_r[l] = 0.8, \forall l$, and $\hat{\mathbf{H}}_r = \mathbf{H}_r$.

added by the relay-destination channel whereas the algorithm in [12] is not affected by such noise.

Figure 3 show the performance of the various algorithms, with all simulation parameters set as in the previous algorithm but with the relay-destination spatial correlation co-efficients now set as $\rho_r[l] = \varrho_r[l] = 0.8, \forall l$. It is observed that the performance of the proposed algorithms, as well as the LS NAF and MMSE NAF algorithms suffer a loss in performance when the relay-destination channel is highly correlated. The algorithm in [12] does not suffer such a performance since as previously noted it is independent of any relay-destination channel parameters. Figure 4 shows the performance of the different algorithms with the source-relay channel spatial coefficients now increased to $\rho_s[l] = \varrho_s[l] = 0.8, \forall l$, and the relay-destination channel spatial co-efficients set as $\rho_r[l] = \varrho_r[l] = 0.2, \forall l$. Comparing the results in Figure 4 to those in Figure 2 it is seen that the proposed algorithms have improved MSE when the source-relay channel is highly correlated.

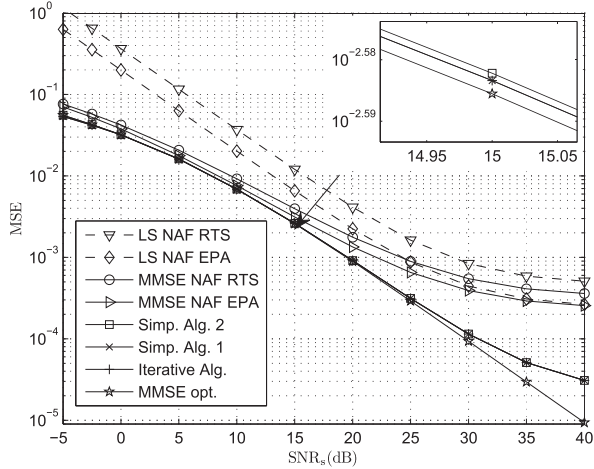


Fig. 4. Source-relay channel estimation MSE against varying SNR_s for the proposed and benchmark algorithms with $N_s = N_r = N_d = 3$, $\text{SNR}_r = 30$ dB, $\rho_s[l] = \rho_r[l] = \rho_d[l] = 0.8$, $\rho_r[l] = \rho_r[l] = 0.2 \forall l$, and $\hat{\mathbf{H}}_r = \mathbf{H}_r$.

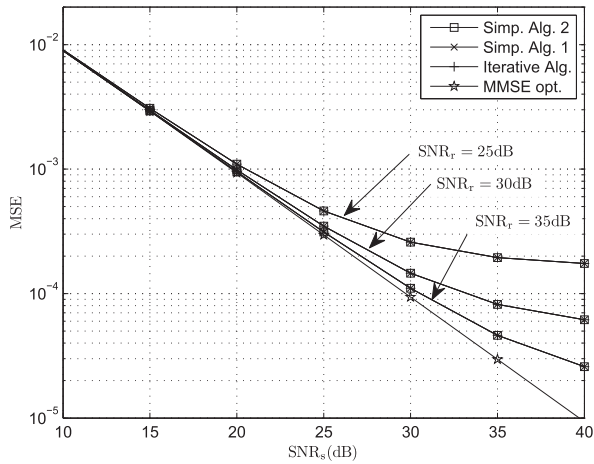


Fig. 5. Source-relay channel estimation MSE against varying SNR_s for the proposed and benchmark algorithms with $N_s = N_r = N_d = 3$, $\text{SNR}_r = \{25, 30, 35\}$ dB, $\rho_s[l] = \rho_r[l] = \rho_d[l] = 0.2 \forall l$, and $\hat{\mathbf{H}}_r = \mathbf{H}_r$.

The effect that the relay-destination channel noise has on the performance of the proposed algorithms is highlighted in Figure 5, where the channel estimation MSE is assessed for different SNR_r values. It is clear the all proposed algorithms have poorer performance for lower SNR_r of the relay-destination link, whereas if the point-to-point estimation algorithm of [12] is used to estimate the source-relay channel at the relay device then the resulting channel estimate is independent of SNR_r .

Figure 6 demonstrates the performance of the proposed source-relay channel estimation algorithms for various antenna configurations. Specifically, the performance of the proposed algorithms are assessed with $N_s = N_d = \{1, 2, 4\}$ and with $N_r = \{3, 6\}$. It is observed that, for all source and destination antenna configurations, the performance of the proposed algorithms decreases with an increasing number of relay antennas. This results due to the fact that an increased number of relay antennas results in an increased number of channel co-efficients to estimate and a poorer channel estimation MSE consequently results.

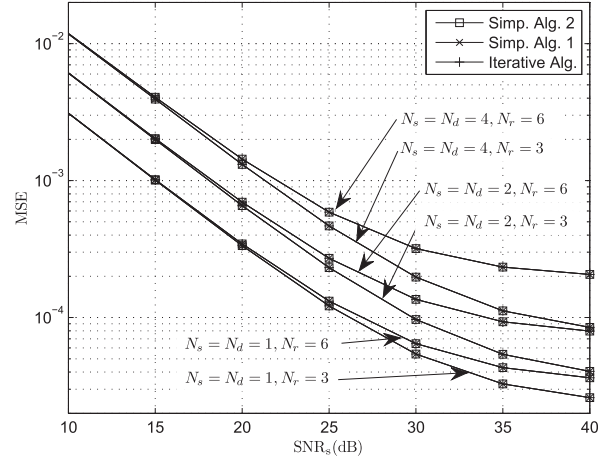


Fig. 6. Source-relay channel estimation MSE against varying SNR_s of the proposed algorithms for various antenna configurations with $\text{SNR}_r = 30$ dB, $\rho_s[l] = \rho_r[l] = \rho_d[l] = 0.2 \forall l$, and $\hat{\mathbf{H}}_r = \mathbf{H}_r$.

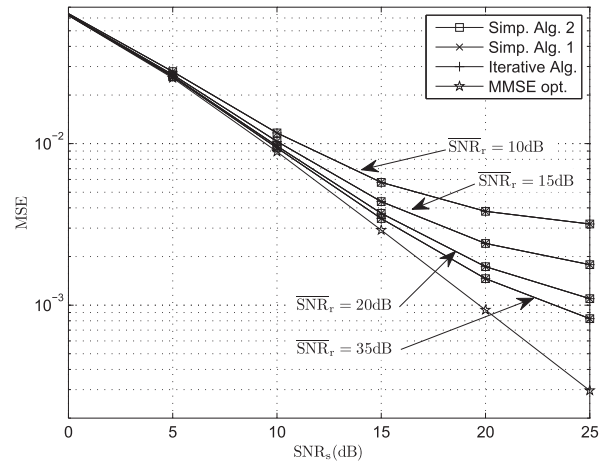


Fig. 7. Source-relay channel estimation MSE against varying SNR_s of the proposed algorithms for $N_s = N_r = N_d = 3$, $\text{SNR}_r = 20$ dB, $\rho_s[l] = \rho_r[l] = \rho_d[l] = 0.2 \forall l$, and $\overline{\text{SNR}}_r = \{10, 15, 20, 35\}$ dB.

We now investigate the performance of the proposed algorithms when the relay-destination channel is estimated imperfectly during the first phase of channel estimation i.e. when $\hat{\mathbf{H}}_r \neq \mathbf{H}_r$. We consider that the relay-destination channel is estimated using the optimal algorithm in [12], with the SNR during this channel estimation phase being $\overline{\text{SNR}}_r$. The quality of the relay-destination channel is obviously highly dependent on $\overline{\text{SNR}}_r$, with low $\overline{\text{SNR}}_r$ signifying a poor relay-destination channel estimate and vice versa. As has been noted in the previous sections, the performance of the proposed source-relay channel estimation algorithms are dependent on the quality of the relay-destination channel estimate.

Figure 7 shows the performance of the proposed algorithms compared to that achieved by estimating the source-relay channel at the relay using the optimal MMSE algorithm in [12]. Results are shown for a system with $N_s = N_r = N_d = 3$ and with the spatial correlation co-efficients set as $\rho_s[l] = \rho_r[l] = \rho_d[l] = 0.2$. The effect of any relay-destination channel estimation error from the first phase of channel estimation is highlighted by different $\overline{\text{SNR}}_r$ values. It can clearly be seen

from these results that the proposed designs (as indeed do any designs where the source-relay channel is estimated at the destination) suffer a degradation in performance when the relay-destination channel is not sufficiently accurate. Estimation of the source-relay channel at the relay node does not suffer this drawback since it does not require knowledge of the relay-destination channel. However, it is again stressed that channel estimation at the relay may be impractical in many cases since the relay device may be a low cost, low power unit, and may not be capable of performing the task of channel estimation.

VII. CONCLUSIONS

In this paper we have studied the problem of channel estimation of spatially correlated frequency selective channels in dual hop MIMO OFDM relay networks. The estimation of the source-relay and relay-destination channels was conducted in two phases. In the first phase the relay-destination channel was estimated using convention point-to-point MIMO OFDM techniques. The source-relay channel was then estimated at the destination in the second phase of channel estimation.

To obtain the MMSE source-relay channel estimate, algorithms were developed to design the source training matrix, the relay precoder, and the destination processor. An iterative algorithm, which has guaranteed convergence, was firstly proposed where each variable was iteratively updated through convex programming. Due to its iterative nature, the proposed iterative algorithm would be too computationally expensive for practical implementation. Two suboptimal algorithms were therefore derived using a high SNR approximation. The suboptimal algorithms have comparable performance to the iterative algorithm but at reduced computational cost, making them more suitable for practical implementation.

Simulation results demonstrated that the proposed algorithms could achieve a better source-relay channel estimate than various benchmarks that also estimated the source-relay channel at the destination. The proposed algorithms were also compared to the optimal point-to-point MIMO OFDM channel estimation design where it was assumed that the source-relay channel could be estimated at the relay device (instead of at the destination). It was shown that if the relay-destination channel was sufficiently accurate and the relay-destination SNR was also sufficiently high, then the proposed algorithms had similar performance to this optimal algorithm. However, the performance of the proposed algorithms was degraded with poorer quality relay-destination channel estimates and/or lower relay-destination channel noise. It is concluded that if the relay device has the capability of performing channel estimation then it can achieve better performance than estimating the channel at the destination. However, when the relay is unable to perform this task (e.g due to computational cost and power constraints) then the proposed algorithms should be used to estimate the source-relay channel at the destination.

APPENDIX A

In this appendix we prove the optimal relay precoder structure given in (57) of Theorem 1. To do so we shall require the following lemma:

Lemma 1: [34] For Hermitian positive semi-definite matrices $\mathbf{A} \in \mathbb{C}^{N \times N}$ and $\mathbf{B} \in \mathbb{C}^{N \times N}$, with eigenvalues $\{\lambda_{a,i}\}_{i=1}^N \in \mathbb{R}_+$ and $\{\lambda_{b,i}\}_{i=1}^N \in \mathbb{R}_+$ arranged in descending order, we have the inequality

$$\text{tr}\{\mathbf{A}\mathbf{B}\} \geq \sum_{i=1}^N \lambda_{a,i} \lambda_{b,N-i+1}, \quad (105)$$

where equality holds when \mathbf{A} is diagonal with diagonal elements arranged in descending order and \mathbf{B} is diagonal with elements arranged in ascending order.

To derive the optimal relay precoder let us introduce the following singular value decompositions (SVD's)

$$\mathbf{Y} \triangleq \mathbf{T}^{-1/2} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} = \mathbf{U}_y \mathbf{\Lambda} \mathbf{V}_y^H \quad (106)$$

$$\mathbf{Z} = \hat{\mathbf{H}}_r \mathbf{G} \mathbf{T}^{1/2} = \mathbf{U}_z \mathbf{\Omega} \mathbf{V}_z^H \quad (107)$$

$$\hat{\mathbf{H}}_r = \mathbf{U}_r \mathbf{\Delta} \mathbf{V}_r^H, \quad (108)$$

where $\mathbf{\Lambda}$, $\mathbf{\Omega}$, and $\mathbf{\Delta}$ are $KN_r \times L_{h_s}$, $KN_d \times KN_r$, and $KN_d \times KN_r$ diagonal matrices, respectively, and contain the positive singular values $\{\lambda_i\}_{i=1}^{R_t} \in \mathbb{R}_{++}$, $\{\omega_i\}_{i=1}^{R_z} \in \mathbb{R}_{++}$, $\{\delta_i\}_{i=1}^{R_r} \in \mathbb{R}_{++}$, respectively. Here we define $R_y \triangleq \text{rank}\{\mathbf{Y}\}$, $R_z \triangleq \text{rank}\{\mathbf{Z}\}$, and $R_r \triangleq \text{rank}\{\hat{\mathbf{H}}_r\}$. The singular values in $\mathbf{\Lambda}$, $\mathbf{\Omega}$, and $\mathbf{\Delta}$ are assumed w.l.o.g. to be arranged in descending order. Substituting (108) into (107) and solving the resulting equation for \mathbf{G} we can parameterise the relay precoder as

$$\mathbf{G} = \mathbf{V}_r \mathbf{\Delta}^\dagger \mathbf{U}_r^H \mathbf{U}_z \mathbf{\Omega} \mathbf{V}_z^H \mathbf{T}^{-1/2}. \quad (109)$$

We wish to find the specific relay precoder from the parameterised set (109) that minimises the objective function in (55) and satisfies the power constraint in (56). We focus firstly on identifying the relay precoder structure that minimises (55). To this end we note that by substituting (106) and (107) into $\text{tr}\{\mathbf{E}_2\}$ in (51), we can write

$$\begin{aligned} \text{tr}\{\mathbf{E}_2\} &= \text{tr}\left\{ \mathbf{Y} \mathbf{Y}^H \left(\sigma_{v_r}^{-2} \mathbf{Z}^H \mathbf{Z} + \mathbf{I}_{KN_r} \right)^{-1} \right\} \quad (110) \\ &= \text{tr}\left\{ \mathbf{U}_y \mathbf{\Lambda} \mathbf{\Lambda}^T \mathbf{U}_y^H \left(\sigma_{v_r}^{-2} \mathbf{V}_z \mathbf{\Omega}^T \mathbf{\Omega} \mathbf{V}_z^H + \mathbf{I}_{KN_r} \right)^{-1} \right\} \quad (111) \end{aligned}$$

$$\geq \sum_{i=1}^{KN_r} \frac{\lambda_i^2 \sigma_{v_r}^2}{\omega_i^2 + \sigma_{v_r}^2}, \quad (112)$$

where the lower bound in (112) is obtained by applying Lemma 1 to (111). It is clear that the lower bound in (112) holds with equality when $\mathbf{V}_z = \mathbf{U}_y$. We therefore find that any precoder given by (109) achieves the lower bound in (112), and therefore minimises (55), provided that $\mathbf{V}_z = \mathbf{U}_y$. To identify a unique precoder we shall select the one that consumes the least transmit power. Substituting (109) into $\text{tr}\{\mathbf{G} \mathbf{T} \mathbf{G}^H\}$ the power consumed by the relay is

$$\text{tr}\{\mathbf{G} \mathbf{T} \mathbf{G}^H\} = \text{tr}\left\{ \mathbf{U}_z \mathbf{\Omega} \mathbf{\Omega}^T \mathbf{U}_z^H \mathbf{U}_r \left(\mathbf{\Delta}^T \right)^\dagger \mathbf{\Delta}^\dagger \mathbf{U}_r^H \right\} \quad (113)$$

$$\geq \sum_{i=1}^{KN_d} \frac{\omega_i^2}{\delta_i^2}, \quad (114)$$

where the lower bound in (114) is obtained by applying Lemma 1 to (113) and holds with equality when $\mathbf{U}_z = \mathbf{U}_r$. Substituting $\mathbf{V}_z = \mathbf{U}_y$ and $\mathbf{U}_z = \mathbf{U}_r$ into (109) we therefore find that the optimal relay precoder that minimises the objective function in (55) whilst consuming the least transmit power is given by

$$\mathbf{G} = \mathbf{V}_r \mathbf{\Delta}^\dagger \mathbf{\Omega} \mathbf{U}_y^H \mathbf{T}^{-1/2}. \quad (115)$$

We now note that from the decomposition in (106) we have

$$\mathbf{U}_y = \mathbf{T}^{-1/2} (\mathbf{M} \otimes \mathbf{I}_{N_r}) \mathbf{R}_{h_s} \mathbf{V}_y \mathbf{\Delta}^\dagger, \quad (116)$$

which upon substituting into (115) results in

$$\mathbf{G} = \mathbf{V}_r \mathbf{\Delta}^\dagger \mathbf{\Omega} (\mathbf{\Lambda}^T)^\dagger \mathbf{V}_y^H \mathbf{R}_{h_s} (\mathbf{M}^H \otimes \mathbf{I}_{N_r}) \mathbf{T}^{-1}. \quad (117)$$

By defining $\mathbf{L} \triangleq \mathbf{V}_r \mathbf{\Delta}^\dagger \mathbf{\Omega} (\mathbf{\Lambda}^T)^\dagger \mathbf{V}_y^H$ we prove the optimal relay structure as stated in (57) of Theorem 1.

APPENDIX B

The proof of the optimal structure of \mathbf{L} in (81) follows similar arguments to those made in Appendix A. Let us firstly introduce

$$\mathbf{M} \triangleq \hat{\mathbf{H}}_r \mathbf{L} \mathbf{N}^{1/2} = \mathbf{U}_m \mathbf{D} \mathbf{V}_m^H, \quad (118)$$

where \mathbf{D} is a $KN_d \times L_{h_s}$ diagonal matrix and contains the singular values $\{d_i\}_{i=1}^{R_m} \in \mathbb{R}_{++}$, with $R_m \triangleq \text{rank}\{\mathbf{M}\}$. Substituting (79) and (80) into (118) we parameterise the set of matrices \mathbf{L} by

$$\mathbf{L} = \mathbf{V}_r \mathbf{\Delta}^\dagger \mathbf{U}_r^H \mathbf{U}_m \mathbf{D} \mathbf{V}_m^H \mathbf{N}^{-1/2}. \quad (119)$$

To find the specific matrix from (119) as the optimal solution to (77)–(78) we firstly consider the objective function. Substituting (79) and (80) into (77) we can show that

$$\text{tr} \left\{ \mathbf{N} \left(\sigma_{v_r}^{-2} \mathbf{N}^{1/2} \mathbf{L}^H \hat{\mathbf{H}}_r^H \hat{\mathbf{H}}_r \mathbf{L} \mathbf{N}^{1/2} + \mathbf{I}_{L_{h_s}} \right)^{-1} \right\} \quad (120)$$

$$= \text{tr} \left\{ \mathbf{U}_n \mathbf{\Psi} \mathbf{U}_n^H \left(\sigma_{v_r}^{-2} \mathbf{V}_m \mathbf{D}^T \mathbf{D} \mathbf{V}_m^H + \mathbf{I}_{L_{h_s}} \right)^{-1} \right\} \quad (121)$$

$$\geq \sum_{i=1}^{L_{h_s}} \frac{\psi_i \sigma_{v_r}^2}{d_i^2 + \sigma_{v_r}^2}, \quad (122)$$

where the lower bound is obtained from Lemma 1 in Appendix A and holds with equality when $\mathbf{V}_m = \mathbf{U}_n$. Matrices \mathbf{L} of the form in (119) therefore minimise the objective function in (77) when $\mathbf{V}_m = \mathbf{U}_n$. To identify the matrix \mathbf{U}_m in (119) we consider the relay power constraint. Substituting (119) into (78) and making use of Lemma 1 we find that

$$\text{tr} \left\{ \mathbf{L} \mathbf{N} \mathbf{L}^H \right\} = \text{tr} \left\{ \mathbf{U}_m \mathbf{D} \mathbf{D}^T \mathbf{U}_m^H \mathbf{U}_r (\mathbf{\Lambda}^T)^\dagger \mathbf{\Delta}^\dagger \mathbf{U}_r^H \right\} \quad (123)$$

$$\geq \sum_{i=1}^{KN_d} \frac{d_i^2}{\delta_i^2}, \quad (124)$$

where the lower bound is achieved when $\mathbf{U}_m = \mathbf{U}_r$. We therefore find that the matrices in (119) are optimal solutions

of the problem in (77)–(78) when $\mathbf{V}_m = \mathbf{U}_n$ and $\mathbf{U}_m = \mathbf{U}_r$. Substituting $\mathbf{V}_m = \mathbf{U}_n$, $\mathbf{U}_m = \mathbf{U}_r$, and the decomposition of (80) into (119) we have

$$\mathbf{L} = \mathbf{V}_r \mathbf{\Delta}^\dagger \mathbf{D} \mathbf{\Psi}^{-1/2} \mathbf{U}_n^H, \quad (125)$$

which proves the optimal \mathbf{L} in (81) with the definition of the diagonal matrix $\mathbf{\Phi} \triangleq \mathbf{\Delta}^\dagger \mathbf{D} \mathbf{\Psi}^{-1/2}$. Substituting $\mathbf{\Phi} = \mathbf{\Delta}^\dagger \mathbf{D} \mathbf{\Psi}^{-1/2}$ into the lower bound in (122) we can write

$$\sum_{i=1}^{L_{h_s}} \frac{\psi_i \sigma_{v_r}^2}{d_i^2 + \sigma_{v_r}^2} = \sum_{i=1}^{L_{h_s}} \frac{\psi_i \sigma_{v_r}^2}{\delta_i^2 \phi_i^2 \psi_i + \sigma_{v_r}^2}. \quad (126)$$

Since $R_r = \text{rank}\{\hat{\mathbf{H}}_r\} = \text{rank}\{\mathbf{\Delta}\}$ and $R_n = \text{rank}\{\mathbf{N}\} = \text{rank}\{\mathbf{\Psi}\}$ it is straightforward to see that (126) will only depend on $R \triangleq \min(R_r, R_n)$ diagonal entries of $\mathbf{\Phi}$. In other words we should have $\text{rank}\{\mathbf{\Phi}\} \leq R$ since setting any $\{\phi_i\}_{R+1}^{\min(KN_r, L_{h_s})}$ will not decrease the objective function in (126) but will lead to an increased transmission power. This concludes the proof of Theorem 2.

APPENDIX C

In this appendix we prove the upper bound in (90). To do so let us firstly prove the following lemma:

Lemma 2: For positive semi-definite matrices $\mathbf{A} \in \mathbb{C}^{N \times N}$ and $\mathbf{B} \in \mathbb{C}^{N \times N}$, the function $q(\mathbf{A}) = \text{tr}\{(\mathbf{A}^{-1} + \mathbf{B})^{-1}\}$ is a concave function of \mathbf{A} .

Proof: Using the matrix inverse identity $(\mathbf{I}_N + \mathbf{P})^{-1} = \mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1} \mathbf{P}$ it is straightforward to show that

$$q(\mathbf{A}) = \text{tr} \left\{ \left(\mathbf{A}^{-1} + \mathbf{B} \right)^{-1} \right\} \quad (127)$$

$$= \text{tr} \left\{ \mathbf{B}^{-1} \right\} - \text{tr} \left\{ (\mathbf{B} \mathbf{A} \mathbf{B} + \mathbf{B})^{-1} \right\}, \quad (128)$$

from which we see that proving the concavity of $q(\mathbf{A})$ is equivalent to showing that $f(\mathbf{A}) \triangleq \text{tr}\{(\mathbf{B} \mathbf{A} \mathbf{B} + \mathbf{B})^{-1}\}$ is convex w.r.t. \mathbf{A} . Let $\mathbf{A} = \mathbf{U}_a + t \mathbf{V}_a$ where $\mathbf{U}_a \in \mathbb{C}^{N \times N}$ and $\mathbf{V}_a \in \mathbb{C}^{N \times N}$ are positive semi-definite and $t \geq 0$, then define

$$f(t) \triangleq f(\mathbf{U}_a + t \mathbf{V}_a) \quad (129)$$

$$= \text{tr} \left\{ (\mathbf{B} \mathbf{U}_a \mathbf{B} + \mathbf{B} + t \mathbf{B} \mathbf{V}_a \mathbf{B})^{-1} \right\}. \quad (130)$$

Proving that $f(\mathbf{A})$ is convex w.r.t. \mathbf{A} is equivalent to proving that $f(t)$ is convex w.r.t. t , which can be proven by showing that the second derivative of $f(t)$ is non-negative [30]. To this end let $\mathbf{X} \triangleq \mathbf{B} \mathbf{U}_a \mathbf{B} + \mathbf{B}$ and $\mathbf{Y} \triangleq \mathbf{B} \mathbf{V}_a \mathbf{B}$ and write

$$f(t) = \text{tr} \left\{ (\mathbf{X} + t \mathbf{Y})^{-1} \right\} \quad (131)$$

$$= \text{tr} \left\{ \mathbf{X}^{-1} \left(\mathbf{I}_N + t \mathbf{X}^{-1/2} \mathbf{Y} \mathbf{X}^{-1/2} \right)^{-1} \right\}. \quad (132)$$

Considering the EVD $\mathbf{Z} \triangleq \mathbf{X}^{-1/2} \mathbf{Y} \mathbf{X}^{-1/2} = \mathbf{U}_z \mathbf{\Lambda}_z \mathbf{U}_z^H$, we can further write (132) as

$$f(t) = \text{tr} \left\{ \mathbf{U}_z^H \mathbf{X}^{-1} \mathbf{U}_z (\mathbf{I}_N + t \mathbf{\Lambda}_z)^{-1} \right\} \quad (133)$$

$$= \sum_{i=1}^N \frac{a_i}{1 + t \lambda_{z,i}}, \quad (134)$$

where $a_i \triangleq [\mathbf{U}_z^H \mathbf{X}^{-1} \mathbf{U}_z]_{ii} \in \mathbb{R}_+$ and $\{\lambda_{z,i}\}_{i=1}^N \in \mathbb{R}_+$ are the eigenvalues in $\mathbf{\Lambda}_z$. We note here that the matrices \mathbf{Z} and $\mathbf{U}_z^H \mathbf{X}^{-1} \mathbf{U}_z$ are Hermitian positive semi-definite and we therefore have $\{a_i\}_{i=1}^N \geq 0$ and $\{\lambda_{z,i}\}_{i=1}^N \geq 0$. With these observations it is straightforward to show from (134) that

$$\frac{\partial f^2(t)}{\partial t^2} = \sum_{i=1}^N \frac{\lambda_{z,i} a_i}{(1 + t\lambda_{z,i})^4} \geq 0, \quad (135)$$

which proves the convexity of $f(t)$. Since $f(t)$ is convex w.r.t. t we find that $q(\mathbf{A})$ in (127) is concave w.r.t. \mathbf{A} . ■

Using Lemma 2 we can now show that

$$\text{tr} \left\{ \left(\mathbf{R}_{h_s}^{-1}[l] + \sigma_{v_s}^{-2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\}, \quad l = 0, \dots, L, \quad (136)$$

are convex functions of $\mathbf{R}_{h_s}[l]$. Therefore, by applying Jensen's inequality to (136), we can obtain

$$\begin{aligned} & \frac{1}{L+1} \sum_{l=0}^L \text{tr} \left\{ \left(\mathbf{R}_{h_s}^{-1}[l] + \sigma_{v_s}^{-2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \\ & \leq \sum_{l=0}^L \text{tr} \left\{ \left((L+1) \bar{\mathbf{R}}_{h_s}^{-1} + \sigma_{v_s}^{-2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\}, \quad (137) \end{aligned}$$

where $\bar{\mathbf{R}}_{h_s} \triangleq \sum_{l=0}^L \mathbf{R}_{h_s}[l]$ and is given by

$$\bar{\mathbf{R}}_{h_s} = \sum_{l=0}^L \sigma_{h_s}^2[l] \bar{\Theta}_s[l] \otimes \Upsilon_s[l]. \quad (138)$$

From the definition of $\bar{\Upsilon}_s$ in (92) it is straightforward to see that $\bar{\Upsilon}_s \succeq \Upsilon_s[l]$, $\forall l$, and consequently from (138) we have $\bar{\mathbf{R}}_{h_s} \preceq \bar{\Theta}_s \otimes \bar{\Upsilon}_s$ (and equivalently $\bar{\mathbf{R}}_{h_s}^{-1} \succeq \bar{\Theta}_s^{-1} \otimes \bar{\Upsilon}_s^{-1}$), where $\bar{\Theta}_s$ was defined in (91). With this result it can then be shown from (137) that

$$\begin{aligned} & \sum_{l=0}^L \text{tr} \left\{ \left((L+1) \bar{\mathbf{R}}_{h_s}^{-1} + \sigma_{v_s}^{-2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \\ & \leq \sum_{l=0}^L \text{tr} \left\{ \left((L+1) \bar{\Theta}_s^{-1} \otimes \bar{\Upsilon}_s^{-1} + \sigma_{v_s}^{-2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\}. \quad (139) \end{aligned}$$

Finally, combining the inequalities in (137) and (139) yields the desired result in (90) after multiplication by $L+1$.

APPENDIX D

In this appendix we prove the structure of $\bar{\mathbf{S}}$ given in (96) of Theorem 3. Let us firstly introduce the SVD

$$\tilde{\mathbf{S}} \triangleq \bar{\mathbf{S}} \bar{\Theta}_s^{-1/2} = \mathbf{U}_{\tilde{\mathbf{S}}} \mathbf{\Pi} \mathbf{V}_{\tilde{\mathbf{S}}}^H, \quad (140)$$

where $\mathbf{\Pi}$ is a square $N_s \times N_s$ diagonal matrix and contains the non-negative singular values $\{\pi_i\}_{i=1}^{N_s} \in \mathbb{R}_+$. Substituting (95)

into (140) and solving for $\bar{\mathbf{S}}$ we find the general family of matrices $\bar{\mathbf{S}}$ given by

$$\bar{\mathbf{S}} = \mathbf{U}_{\tilde{\mathbf{S}}} \mathbf{\Pi} \mathbf{V}_{\tilde{\mathbf{S}}}^H \bar{\Theta}_s^{-1/2}. \quad (141)$$

We now note that, through some straightforward deductions, we can write the objective function in (93) as

$$\begin{aligned} & \text{tr} \left\{ \left(\bar{\Theta}_s^{-1} \otimes \bar{\Upsilon}_s^{-1} + \sigma_{v_s}^{-2} (L+1)^{-1} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \\ & = \text{tr} \left\{ \left(\bar{\Theta}_s \otimes \bar{\Upsilon}_s \right) \left(\mathbf{I}_{L h_s} + \sigma_{v_s}^{-2} (L+1)^{-1} \right. \right. \\ & \quad \left. \left. \times \left(\bar{\Theta}_s^{1/2} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \bar{\Theta}_s^{1/2} \otimes \bar{\Upsilon}_s \right) \right)^{-1} \right\}, \quad (142) \end{aligned}$$

where we have used the fact that, for matrices of commensurate dimensions, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D})$. Substituting (95) and (140) into the right hand side of (142), and again making use of the previous Kronecker product rule, we can write

$$\begin{aligned} & \text{tr} \left\{ \left(\bar{\Theta}_s^{-1} \otimes \bar{\Upsilon}_s^{-1} + \sigma_{v_s}^{-2} (L+1)^{-1} \bar{\mathbf{S}}^H \bar{\mathbf{S}} \otimes \mathbf{I}_{N_r} \right)^{-1} \right\} \\ & = \text{tr} \left\{ \left(\mathbf{\Xi} \otimes \bar{\Upsilon}_s \right) \left(\mathbf{I}_{L h_s} + \sigma_{v_s}^{-2} (L+1)^{-1} \right. \right. \\ & \quad \left. \left. \times \left(\mathbf{V}_{\bar{\Theta}_s}^H \mathbf{V}_{\tilde{\mathbf{S}}} \mathbf{\Pi}^T \mathbf{\Pi} \mathbf{V}_{\tilde{\mathbf{S}}}^H \mathbf{V}_{\bar{\Theta}_s} \otimes \bar{\Upsilon}_s \right) \right)^{-1} \right\}, \quad (143) \end{aligned}$$

$$\geq \sum_{i=1}^{N_s} \sum_{j=1}^{N_r} \frac{\xi_i \nu_j \sigma_{v_s}^2 (L+1)}{\sigma_{v_s}^2 (L+1) + \pi_i^2 \nu_j}, \quad (144)$$

where the lower bound is obtained by applying Lemma 1 from Appendix A to (143). We note that (143) is invariant to the unitary matrix $\mathbf{U}_{\tilde{\mathbf{S}}}$ and that the lower bound in (144) holds with equality when $\mathbf{V}_{\tilde{\mathbf{S}}} = \mathbf{V}_{\bar{\Theta}_s}$. Therefore, matrices given by (141) minimise the objective function in (93) when $\mathbf{V}_{\tilde{\mathbf{S}}} = \mathbf{V}_{\bar{\Theta}_s}$. We now note that by substituting (95) into (141), and the resulting equation into $\text{tr}\{\bar{\mathbf{S}}\bar{\mathbf{S}}^H\}$, we have

$$\text{tr} \left\{ \bar{\mathbf{S}}\bar{\mathbf{S}}^H \right\} = \text{tr} \left\{ \mathbf{V}_{\tilde{\mathbf{S}}} \mathbf{\Pi}^T \mathbf{\Pi} \mathbf{V}_{\tilde{\mathbf{S}}}^H \mathbf{V}_{\bar{\Theta}_s} \mathbf{\Xi}^{-1} \mathbf{V}_{\bar{\Theta}_s}^H \right\} \quad (145)$$

$$\geq \sum_{i=1}^{N_s} \frac{\pi_i^2}{\xi_i}, \quad (146)$$

where the lower bound results from the use of Lemma 1 and holds with equality when $\mathbf{V}_{\tilde{\mathbf{S}}} = \mathbf{V}_{\bar{\Theta}_s}$. Since both (143) and (145) are invariant to $\mathbf{U}_{\tilde{\mathbf{S}}}$ we can w.l.o.g. select $\mathbf{U}_{\tilde{\mathbf{S}}} = \mathbf{I}_{N_s}$. We therefore find that, by substituting $\mathbf{U}_{\tilde{\mathbf{S}}} = \mathbf{I}_{N_s}$ and $\mathbf{V}_{\tilde{\mathbf{S}}} = \mathbf{V}_{\bar{\Theta}_s}$ into (141), the optimal structure of $\bar{\mathbf{S}}$ is

$$\bar{\mathbf{S}} = \mathbf{\Pi} \mathbf{\Xi}^{-1/2} \mathbf{V}_{\bar{\Theta}_s}^H, \quad (147)$$

which, after defining the diagonal matrix $\mathbf{\Gamma} \triangleq \mathbf{\Pi} \mathbf{\Xi}^{-1/2}$, concludes the proof of Theorem 3.

ACKNOWLEDGMENT

The authors would like to thank the editor and anonymous reviewers for their valuable comments and suggestions that helped improve the quality of this manuscript.

REFERENCES

- [1] 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.
- [2] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, Jun. 2007.
- [3] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.
- [4] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068–6077, Dec. 2009.
- [5] Y. Rong, "Optimal linear non-regenerative multi-hop MIMO relays with MMSE-DFE receiver at the destination," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2268–2279, Jul. 2010.
- [6] A. P. Millar, S. Weiss, and R. W. Stewart, "Precoder design for MIMO relay networks with direct link and decision feedback equalisation," *IEEE Commun. Lett.*, vol. 15, no. 10, pp. 1044–1046, Oct. 2011.
- [7] A. P. Millar, S. Weiss, and R. W. Stewart, "Tomlinson harashima precoding design for non-regenerative MIMO relay networks," in *Proc. 73rd IEEE Veh. Technol. Conf. (VTC Spring)*, May 2011, pp. 1–5.
- [8] A. P. Millar, S. Weiss, and R. W. Stewart, "THP transceiver design for MIMO relaying with direct link and partial CSI," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1204–1207, Jun. 2013.
- [9] F.-S. Tseng, M.-Y. Chang, and W.-R. Wu, "Joint Tomlinson-Harashima source and linear relay precoder design in amplify-and-forward MIMO relay systems via MMSE criterion," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1687–1698, May 2011.
- [10] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.
- [11] H. Zhang, Y. G. Li, A. Reid, and J. Terry, "Optimum training symbol design for MIMO OFDM in correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2343–2347, Sep. 2006.
- [12] H. D. Tuan, H. H. Kha, H. H. Nguyen, and V.-J. Luong, "Optimized training sequences for spatially correlated MIMO-OFDM," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2768–2778, Sep. 2010.
- [13] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67–75, Jan. 2002.
- [14] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May. 2006.
- [15] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May. 2008.
- [16] S. Su and Y. Jing, "Channel training design in amplify-and-forward MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3380–3391, Oct. 2011.
- [17] P. Lioliou, M. Viberg, and M. Matthaiou, "Bayesian approach to channel estimation for AF MIMO relaying systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1440–1451, Sep. 2012.
- [18] T. Kong and Y. Hua, "Optimal design of source and relay pilots for MIMO relay channel estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4438–4446, Sep. 2011.
- [19] J. Ma, P. Orlik, J. Zhang, and G. Y. Li, "Pilot matrix design for interim channel estimation in two hop MIMO AF relay systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2009, pp. 1–5.
- [20] J. Ma, P. Orlik, J. Zhang, and G. Y. Li, "Pilot matrix design for estimating cascaded channels in two-hop mimo amplify-and-forward relay systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1956–1965, Jun. 2011.
- [21] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Robust channel estimation algorithm for dual-hop MIMO relay channels," in *Proc. IEEE 23rd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 2347–2352.
- [22] F.-S. Tseng, W.-J. Huang, and W.-R. Wu, "Robust far-end channel estimation in three-node amplify-and-forward MIMO relay systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 3752–3766, Oct. 2013.
- [23] F.-S. Tseng, W.-J. Huang, and W.-R. Wu, "Robust far-end channel estimation design for MIMO relay systems," in *Proc. Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Dec. 2013, pp. 1–6.
- [24] C. S. Patel and G. L. Stuber, "Channel estimation for amplify and forward relay based cooperation diversity systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2348–2356, Jun. 2007.
- [25] P. Lioliou and M. Viberg, "Least square based channel estimation for MIMO relays," in *Proc. Int. ITG Workshop Smart Antennas (WSA)*, Feb. 2008, pp. 90–95.
- [26] P. Lioliou, M. Viberg, and M. Coldrey, "Performance analysis of relay channel estimation," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput.*, Nov. 2009, pp. 1533–1537.
- [27] H. Bolcskei, M. Borgmann, and A. J. Paulraj, "Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 427–439, Apr. 2003.
- [28] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.
- [29] 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.
- [30] S. Boyd and L. Vandenberghe, *Convex Optimisation*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [31] D. P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686–695, Feb. 2005.
- [32] 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.
- [33] C. Xing, S. Ma, Y. C. Wu, and T. S. Ng, "Transceiver design for dual-hop nonregenerative MIMO-OFDM relay systems under channel uncertainties," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6325–6339, Dec. 2010.
- [34] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York, NY, USA: Academic, 1979.



Andrew P. Millar received the M.Eng. and Ph.D. degrees from the University of Strathclyde, Glasgow, Scotland, in 2008 and 2014, respectively, both in electronic and electrical engineering. During his Ph.D. degree, he studied in the Signal and Image Processing Group, Department of Electronic and Electrical Engineering. He currently works for Genesis Oil and Gas Consultants Ltd. in the environmental modeling group. His research interests include transceiver design and channel estimation for multiple-input multiple-output systems, relaying networks, statistical and adaptive signal processing, matrix theory, convex optimisation, oil spill modeling, chemical discharge modelling, and the development of underwater noise propagation algorithms and software.



Stephan Weiss (S'97–A'98–M'99–SM'05) received the Dipl.-Ing. degree from the University of Erlangen-Nürnberg, Erlangen, Germany, in 1995, and the Ph.D. degree from the University of Strathclyde, Glasgow, Scotland, in 1998, both in electronic and electrical engineering. He is a Reader with the Department of EEE, University of Strathclyde. From 1999 to 2006, he was a Lecturer and Senior Lecturer with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K. Prior to this, he was a Visiting Lecturer at the University of Strathclyde from 1998 to 1999, and a Visiting Scholar at the University of Southern California, Los Angeles, CA, USA, from 1996 to 1997. He has authored more than 250 technical papers. His research interests include adaptive, multirate, and array signal processing with applications in communications, audio, and biomedical signal processing. He is a member of EURASIP. He was the Technical Co-Chair for EUSIPCO 2009, and General Chair for ISPLC 2014 in Glasgow. For his work in biomedical signal processing, he was the corecipient of the 2001 Research Award of the German Society on hearing aids, and the VTC-Spring Best Paper Award in the MIMO systems track in 2011.



Robert W. Stewart is a Professor and currently the Head of the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, Scotland. He has more than 25 years experience working on FPGA-based design for communications, adaptive DSP, software defined radio, and most recently wireless white space communications. From 1997 to 2011, he was also a Visiting Professor at the University of California, Los Angeles (UCLA), Los Angeles, CA, USA, and from 2004 to 2012, he was the CEO with the technology company Steepest Ascent Ltd. working in the area of DSP and communications systems simulation and design. He has also consulted extensively to industry in Europe and the United States. In recent years, he has been a coauthor for two technology books, one on using the Zynq FPGA, and another on practical and low-cost software-defined radio systems.