
Precoding and Relaying Algorithms for Multiuser MIMO Downlink Channels

Idoia Jimenez Ramirez

Supervisor:

Mikel Mendicute Errasti



MONDRAGON
UNIBERTSITATEA

A thesis submitted for the degree of
Doctor por Mondragon Unibertsitatea

Department of Electronics and Computer Science
Mondragon Goi Eskola Politeknikoa
Mondragon Unibertsitatea

April 2013

Familixandako,

Ta Ionendako.

Learn from yesterday,

live for today,

hope for tomorrow.

The important thing is not to stop questioning.

Albert Einstein

Agradecimientos

Algo más de cuatro años han pasado desde que empezó esta aventura. Durante este periodo he tenido la oportunidad de conocer a muchos investigadores que han ayudado a que esta tesis llegue hoy a su fin. A todos los que han estado en los buenos pero sobre todo en los malos momentos gracias.

En primer lugar, me gustaría agradecer a mi director de tesis **Mikel Mendicute** su inestimable ayuda. Gracias por su disponibilidad y por los consejos dados que han hecho posible el buen fin de esta tesis. Por otro lado, me gustaría agradecer a **Mondragon Goi Eskola Politeknikoa** la beca de investigación que me otorgó hace ya cuatro largos años.

Quisiera agradecer muy especialmente a **Stephan Weiss** de **Strathclyde University** su cálida acogida así como su aportación. Gracias también a mis compañeros del **Centre for excellence in Signal and Image Processing (CeSIP)**, especialmente a **Andrew Millar** por hacerme sentir una más y hacer más fácil la adaptación cuando estas lejos de casa.

Gracias a mis compañeros de departamento por brindarme su ayuda así como por esos largos cafés que ayudan a la desconexión. Me gustaría agradecer especialmente a **Iker Sobron**, **Maitane Barrenechea**, **Lorena Martinez** y **Pello Otxandiano** su inestimable ayuda. No me gustaría olvidarme de **Aritz Legarda**, **Maite Beamurgia**, **Iñaki Garitano**, **Lorea Belategi** y **Joxe Aizpurua** así como de los recién llegados **Aitor Lizeaga**, **Alain Perez** y **Enaitz Ezpeleta**. Gracias por la ayuda prestada y por los buenos momentos vividos pero sobre todo por hacer más fácil el día a día.

Finalmente me gustaría expresar mi más sincera gratitud a **mis padres** y a mis hermanos **Txapu**, **Inma** y **Fran**. Gracias por interesarse por esta tesis y por los buenos consejos. No me gustaría olvidarme de mis sobrinos **Danel**, **Elene**, **Beñat**, **Luken** y como no, del gran **Markelxo**. No menos importante ha sido la ayuda prestada por mis **amig@s**. Gracias por haberme dado ánimos en momentos de flaqueza. Pero sobre todo gracias en mayúsculas a **Zubi**, por estar siempre ahí, por animarme a seguir aprendiendo y sobre todo por obligarme a parar cuando más bloqueada estaba.

Acknowledgments

More than four years have passed since I started this adventure. During this period, I have had the opportunity to meet many researchers who have helped making this thesis comes to an end today. To all those who have been in good times but especially in bad times thank you.

First of all, I would like to thank my thesis supervisor **Mikel Mendicute** his invaluable help. Thank you for your availability and all the advice given that made possible the success of this thesis. Furthermore, I would like to thank to **Mondragon Goi Eskola Politeknikoa** the research grant that awarded me four years ago.

I would like to extend special thanks to **Stephan Weiss** of **Strathclyde University** for his warm welcome as well as for his contribution to this PhD dissertation. Thanks also to my workmates at **Centre for Excellence in Signal and Image Processing (CESIP)** and specially thanks to **Andrew Millar** for making me feel one more and for making easier the adaptation when you are away from home.

I would also like to express my gratitude to my fellow colleagues, for the enjoyable coffee breaks and all the great moments of laughter and jokes we shared together. I would like to show special appreciation to **Iker Sobrón**, **Maitane Barrenechea**, **Lorena Martinez** and **Pello Otxandiano** for their irreplaceable help. I would not like to forget **Aritz Legarda**, **Maite Beamurgia**, **Iñaki Garitano**, **Lorea Belategi** and **Joxe Aizpurua** as well as the newcomers **Aitor Lizeaga**, **Alain Perez** and **Enaitz Ezpeleta**. Thanks for the cheering me up and for the good moments lived that makes easier the day-to-day.

Finally I would like to express my sincere gratitude to my parents and my siblings **Txapu**, **Inma** and **Fran**. Thanks for your interest and for your good advice. I would not like to forget my nephews **Danel**, **Beñat**, **Elene**, **Luken** and of course, the great **Markeltxo**. Not less important was the help given by **my friends**. Thanks for bucking me up in moments of weakness. But mostly thanks to **Zubi**, for always being there, for encouraging me to keep learning and specially for forced me to stop when I was locked.

Abstract

In the last years, research has focused on multiple-input multiple-output (MIMO) wireless technology due to the capacity and performance improvement it provides, offering a higher spectral efficiency. In addition, when multiple users take part in the network, the scenario becomes much more complex, since resources like bandwidth, time or transmission power must be shared. Furthermore, the performance of the system is degraded as a consequence of the noise and multiuser interference (MUI). When the transmission is conducted from a base station (BS) to multiple users, a pre-equalization stage called precoding is applied. By means of this, each user will be able to interpret the signal independently, without the knowledge of the channel.

Precoding techniques are classified into linear and non-linear. In fact, the non-linear Tomlinson-Harashima precoding (THP) and vector precoding (VP) techniques have been shown to achieve very good results in multiuser broadcast channels, outperforming the performance of linear systems considerably. Nevertheless, the computational complexity also increases as a consequence of successive interference cancellation in THP or the search for the closest point in an infinite lattice for the case of VP.

In order to improve the coverage and range of the network in fading environments, as well as for MUI mitigation, a relaying terminal can be introduced between the BS and the users, being the signal modified either regeneratively or not to reach the destination. Hence, apart from incorporating a precoding stage at the BS for the separation of user streams, multiuser MIMO relaying systems can include a processing stage at the relay. By means of this, the signal will be provided by the characteristic necessary to reach the destination in a more efficient way.

Throughout this thesis, the design of transceivers for non-regenerative multiuser MIMO relay systems is addressed, where the main challenge resides on the joint design of the precoder at the base station and the relaying matrix. We propose the use of linear and non-linear precoding techniques such as THP and VP at the downlink relayed transmission scenario. Apart from the joint design of the linear and non-linear transceiver design, block diagonalization (BD) is proposed in combination with the mentioned precoding techniques.

BD, which maps each user signal at the null space created by the interfering users, is applied at the relay for interference cancellation whilst linear processing and VP are set at the BS for users signal separation and overall error minimization.

The main contributions of this research work are the novel optimal proposals based on non-linear precoding techniques and the consideration of BD for the joint design. Furthermore, simple and computationally efficient suboptimal approaches based on linear and non-linear precoding techniques are first here proposed. In addition, block diagonal geometric mean decomposition is presented in combination with VP for the optimal and suboptimal design of the transmitter. To conclude, a complete complexity analysis is given in order to show the effectiveness of the proposed schemes.

Resumen

En los últimos años, la investigación se ha centrado en la tecnología inalámbrica MIMO gracias a la mejora en cuanto a capacidad y rendimiento, ofreciendo una mayor eficiencia espectral. Además, cuando múltiples usuarios forman la red, el escenario es mucho más complejo, ya que se han de compartir recursos como ancho de banda, tiempo o potencia de transmisión. Aun así, el rendimiento se ve degradado como consecuencia del ruido y de las interferencias multiusuario. Cuando la transmisión se establece desde la estación base (BS) a múltiples usuarios, ha de llevarse a cabo un pre-procesamiento de la señal conocido como precodificación o *precoding*. A través de este último cada usuario será capaz de interpretar la señal de manera independiente, sin necesidad de conocer el canal.

Las técnicas de precodificación se clasifican en lineales y no lineales. De hecho, la precodificación Tomlinson-Harashima (THP) y la precodificación vectorial (VP) corresponden a técnicas no lineales para la adquisición no cooperativa de la señal en canales *broadcast*. Estas técnicas, mejoran el rendimiento de las lineales de manera considerable. Sin embargo, la complejidad computacional también se ve incrementada debido a la cancelación sucesiva de interferencias en THP y la búsqueda del punto más cercano en un *lattice* o reticulado infinito en VP.

Con el objetivo de aumentar el rango y cobertura de la red en entornos hostiles, además de para mitigar el efecto de las interferencias, un retransmisor o *relay* puede ser introducido entre el transmisor y el receptor, pudiendo ser la señal modificada de manera regenerativa o no para llegar al destino. Por consiguiente, aparte de incorporar una etapa de *precoding* en la estación base para adquisición no cooperativa de la señal, las redes MIMO mutiusuario con retransmisores pueden incorporar una etapa de procesamiento en el relay. Esta etapa, conocida como *relaying*, proporcionará a la señal las características necesarias para alcanzar al destino.

A lo largo de esta tesis se realiza el diseño de transceptores o *transceivers* para redes MIMO multiusuario con retransmisores no regenerativos, donde el mayor reto reside en el diseño óptimo o subóptimo conjunto del precodificador en la estación base y el procesador en el relay. Aparte del diseño conjunto lineal y no lineal, la diagonalización por bloques,

del inglés *block diagonalization* (BD), es propuesto junto a estas técnicas de precodificación. BD, que mapea la señal de cada usuario en el espacio nulo creado por los usuarios interferentes, es aplicado en el retransmisor para la cancelación de interferencias, mientras que el procesamiento lineal o VP se establecen en BS para la separación de usuarios así como para la minimización del error.

Las contribuciones principales de este trabajo de investigación son las propuestas óptimas novedales basadas en precodificación no lineal así como la consideración de BD para el diseño conjunto. Además, varios sistemas subóptimos basados en técnicas lineales y no lineales son propuestos aquí por primera vez, siendo diferenciados por su simplicidad y por ser computacionalmente eficientes. Además, la descomposición geométrica del canal por bloques (BD-GMD) es presentado junto a VP para el diseño óptimo y subóptimo de los transceptores. Para concluir, un análisis completo es proporcionado con el firme objetivo de mostrar la eficacia de los sistemas propuestos.

Laburpena

Azkenengo urteetan, ikerketa haririk gabeko MIMO teknologian zentratu da, kapazitate eta errendimenduaren hobekuntzaren ondorioz, eraginkortasun espektrala handiagotuz. Gainera, erabiltzaile bat baino gehiagok parte hartzen dutenean, egoera konplexuagoa bihurtzen da, batez ere banda zabalera, denbora edota potentzia bezalako errekurtsok partekatu behar dituztelako. Aldi berean, errendimendua gutxitzen da beste erabiltzaileek sortutako interferentzien eta zarataren ondorioz. Transmizioa terminal batetik (BS) terminal anitzetara ezartzen denean, erabiltzaileen arteko kooperazio eza dela eta, ekualizazio prozesu bat ezarri beharra dago transmisorean. Azkenengo honen bitartez, prekodifikazio edo *precoding* bezala ezagutzen dena, erabiltzaile bakoitzak bere seinalea interpreta dezake kanala ezagutu gabe.

Prekodifikazio tekniken artean linealak eta ez linealak topa ditzazkegu. Izan ere, THP eta VP teknika ez linealek seinalearen interpretazio ez kooperatiboa ahalbidetzen dute *broadcast* kanaletan. Teknika hauek linealen errendimendua hobetzen dute nabarmenki. Hala ere, zailtasun konputazionala areagotzen da, batez ere THP-k erabiltzen duen interferentziak ekiditzeko prozesuagatik eta VP-k erretikulatu infinito baten burutzen duen bilaketaren ondorioz.

Hutsaltze inguruneetan, haririk gabeko sarearen estaldura eta barrutia handitzeko, eta aldi berean erabiltzaileen artean sortutako interferentziak gutxitzeko, transmisorearen eta jasotzaileen artean errepikatzaile baten ezarpena proposatzen da. Honen bitartez, erabiltzaileetara heldu ahal izango da, seinalearen modifikazioaren edo modifikazio ezaren bitartez. Hortaz, predifikazio etapari errepikatzaile prozesua edo relaying prozesua gehitu beharko litzaioke. Etapa honetan, *relaying* bezala ezagutzen dena, seinaleak hartzailera heltzeko behar dituen ezaugarriak ematen dizkio.

Tesi honetan zehar erabiltzaile anitzeko errepikatzaile ez erregeneratiboetan oinarritutako MIMO sareak aztertuko dira, ikerketa filtroen diseinu optimo eta suboptimoan zentratuz. Hortaz, erronka handiena transmisorean eta errepikatzailean kokatutako filtroen diseinu bateratuan egongo litzateke. Diseinua *precoding* teknika lineala eta ez linealendako burutuko da. Gainera blokeka burututako diagonalizazioa (BD) proposatzen da *precoding* lineala eta

ez linealarekin batera. BD-k erabiltzaile bakoitzaren seinalea interferentziarik gabeko espazio nuluan mapeatzen du. Teknika hau errepikatzaile terminalean aplikatzen da, interferentziak ekidituz. Prekodifikazio lineala eta VP aldiz transmisorean aplikatuko dira sistemaren errore totala gutxitu ahal izateko.

Aldi berean, BD-GMD deskonposizioa, ingelesez block diagonal geometric mean decomposition dena, VP prekoding teknikarekin konbinatzen da filtroen diseinu optimo eta suboptimoa gauzatzeko. Azkenik, konplexutasuna aztertzen da, proposatutako sistemen hobekuntza azpimarratzeko.

Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself were developed entirely by myself at the Signal Theory and Communications Area, Department of Electronics and Computer Science, at the University of Mondragon.

The software used to perform the simulations was developed entirely by myself with the following exceptions: the Matlab implementation of the sphere encoder algorithm that has been developed by Maitane Barrenechea.

Idoia Jimenez Ramirez
Department of Electronics and Computer Science
Mondragon Goi Eskola Politeknikoa
Mondragon Unibertsitatea
April, 2013

Contents

| | |
|--|----------|
| Acknowledgments | iii |
| Abstract | v |
| Declaration of Originality | xi |
| Contents | xii |
| List of Figures | xvi |
| List of Tables | xx |
| List of Algorithms | xxii |
| List of Symbols | xxv |
| 1 Introduction | 1 |
| 1.1 Motivation and Objectives | 1 |
| 1.1.1 Main Objectives | 2 |
| 1.2 Thesis Contributions | 2 |
| 1.3 Organization of the Thesis | 3 |
| 2 Background and Related Work | 6 |
| 2.1 Introduction | 6 |
| 2.2 MIMO | 6 |
| 2.2.1 MIMO Channel Capacity | 7 |
| 2.2.2 Parallel Decomposition of the MIMO Channel | 8 |
| 2.3 Multiuser MIMO Communications | 10 |
| 2.3.1 Multiple Access Channel | 10 |
| 2.3.2 Broadcast Channel | 12 |
| 2.3.3 MU-MIMO Channel Capacity Region | 13 |
| 2.3.3.1 MIMO-MAC Channel Capacity | 13 |
| 2.3.3.2 MIMO-BC Channel Capacity | 13 |
| 2.4 Multiuser MIMO Relaying Communications | 14 |
| 2.4.1 Relaying Systems | 16 |
| 2.4.1.1 Amplify-and-Forward | 17 |

| | | |
|----------|--|-----------|
| 2.4.1.2 | Decode-and-Forward | 17 |
| 2.4.2 | Multiuser MIMO Relaying Model | 18 |
| 2.4.2.1 | Multiuser MIMO AF Downlink Relaying | 19 |
| 2.5 | Precoding techniques | 21 |
| 2.5.1 | Linear Precoding | 22 |
| 2.5.2 | Non-Linear Precoding Techniques | 22 |
| 2.5.2.1 | Tomlinson-Harashima Precoding | 22 |
| 2.5.2.2 | Vector Precoding | 24 |
| 2.6 | Chapter summary | 24 |
| 3 | Linear Precoding in Multiuser MIMO AF Relaying Systems | 26 |
| 3.1 | System Model | 26 |
| 3.2 | Joint Local Optimal Linear Transceiver Design | 28 |
| 3.2.1 | Zero-Forcing Criterion | 28 |
| 3.2.2 | Minimum Mean Square Error Criterion | 30 |
| 3.2.3 | Receive Equalization Based Design | 32 |
| 3.2.3.1 | Zero-Forcing Equalizer | 32 |
| 3.2.3.2 | Minimum Mean Square Error Equalizer | 32 |
| 3.2.4 | Simulation Results | 33 |
| 3.2.4.1 | BER Performance | 33 |
| 3.2.4.2 | Achievable Sum-Rate | 36 |
| 3.2.4.3 | Convergence | 37 |
| 3.3 | Block Diagonalization Based Designs | 38 |
| 3.3.1 | Block Diagonalization System Model | 39 |
| 3.3.2 | Design of Precoding and Relaying Matrices | 41 |
| 3.3.3 | Partial Block Diagonalization Linear Approach | 43 |
| 3.3.3.1 | Zero-Forcing Partial BD | 45 |
| 3.3.3.2 | Minimum Mean Square Error Partial BD | 45 |
| 3.3.3.3 | Receiver Equalization | 47 |
| 3.3.4 | Simulation Results | 47 |
| 3.3.4.1 | BER Performance | 47 |
| 3.3.4.2 | Achievable Sum-Rate | 48 |
| 3.3.4.3 | Convergence | 50 |
| 3.4 | Computational Complexity | 52 |
| 3.4.1 | Complexity analysis for the joint linear approaches | 54 |
| 3.4.2 | Complexity analysis for block diagonalization based approaches | 56 |
| 3.5 | Chapter summary | 58 |

| | | |
|----------|---|-----------|
| 4 | Suboptimal Linear Precoding in Multiuser MIMO AF Relaying Systems | 60 |
| 4.1 | Suboptimal Linear Block Diagonalization with All-Pass Relaying | 61 |
| 4.1.1 | Zero Forcing | 62 |
| 4.1.2 | Minimum Mean Square Error | 63 |
| 4.1.3 | Simulation Results | 63 |
| 4.1.3.1 | BER Performance | 64 |
| 4.1.3.2 | Achievable Sum-Rate | 65 |
| 4.1.3.3 | Error Measurement | 66 |
| 4.2 | Independent Hop-by-Hop MMSE Approach | 66 |
| 4.2.1 | Simulation Results | 68 |
| 4.2.1.1 | BER Performance | 68 |
| 4.2.1.2 | Achievable Sum-Rate | 69 |
| 4.2.1.3 | Error Measurement | 69 |
| 4.3 | System Diagonalization Based Approaches | 71 |
| 4.3.1 | System Diagonalization Model | 72 |
| 4.3.2 | Optimal Precoder Structure | 73 |
| 4.3.3 | Sum-MSE Minimization | 74 |
| 4.3.3.1 | Equally Distributed Power Allocation | 78 |
| 4.3.3.2 | Sum-MSE Minimization with User Selection | 79 |
| 4.3.4 | Sum-Rate Maximization Based Approach | 80 |
| 4.3.4.1 | Equally Distributed Power Allocation Approach | 84 |
| 4.3.4.2 | Sum-Rate Maximization Based User Selection Algorithm | 84 |
| 4.3.5 | Simulation Results | 84 |
| 4.3.5.1 | BER Performance | 86 |
| 4.3.5.2 | Achievable Sum-Rate | 87 |
| 4.4 | Computational Complexity | 88 |
| 4.4.1 | Complexity Analysis for the BD All-Pass Filter Design | 88 |
| 4.4.2 | Complexity Analysis for Independent Hop-by-Hop Approach | 90 |
| 4.4.3 | Complexity Analysis for System Diagonalization Based Approaches | 90 |
| 4.4.4 | Complexity Performance Trade-Off | 92 |
| 4.5 | Chapter Summary | 93 |
| 5 | Non-Linear Precoding in Multiuser MIMO AF Relaying Systems | 95 |
| 5.1 | Local Optimal Design with Tomlinson-Harashima Precoding | 96 |
| 5.1.1 | System Model | 96 |
| 5.1.2 | Joint Filter Design with Tomlinson-Harashima Precoding | 99 |
| 5.2 | Joint Linear Design with VP Transmission | 102 |
| 5.2.1 | System Model | 103 |

| | | |
|----------|---|------------|
| 5.2.2 | Joint Design of the Filters | 104 |
| 5.2.3 | Vector Precoding with Block Diagonalization | 106 |
| 5.2.3.1 | Joint Design of Precoder and Relay under BD-VP Criterion | 107 |
| 5.3 | Simulation Results | 109 |
| 5.3.1 | BER Performance | 110 |
| 5.3.2 | Achievable Sum-Rate | 112 |
| 5.3.3 | Convergence | 114 |
| 5.4 | Complexity Analysis | 114 |
| 5.4.1 | Number of FLOPs and Complexity Order Analysis | 116 |
| 5.4.2 | Running-Time Analysis | 117 |
| 5.5 | Chapter Summary | 118 |
| 6 | Suboptimal Non-Linear Approaches | 120 |
| 6.1 | Suboptimal Design for Tomlinson-Harashima Precoding | 121 |
| 6.1.1 | Simulation Results | 123 |
| 6.1.1.1 | BER Performance | 124 |
| 6.1.1.2 | Achievable Sum-Rate | 125 |
| 6.1.1.3 | Mean Square Error | 126 |
| 6.2 | Suboptimal Vector Precoding Designs | 126 |
| 6.2.1 | Hop-by-Hop MMSE with VP Transmission | 127 |
| 6.2.2 | Problem Decomposition based Suboptimal Approach | 129 |
| 6.2.3 | Simulation Results | 132 |
| 6.2.3.1 | BER Performance | 133 |
| 6.2.3.2 | Achievable Sum-Rate | 134 |
| 6.2.3.3 | Mean Square Error | 135 |
| 6.3 | Computational Complexity | 135 |
| 6.3.1 | FLOP Operations and Complexity Order Analysis | 136 |
| 6.3.2 | Run-Time Analysis | 137 |
| 6.4 | Chapter Summary | 138 |
| 7 | Block Diagonal Geometric Mean Decomposition for Multiuser MIMO Re- | |
| | laying Systems | 140 |
| 7.1 | BD-GMD Algorithm Analysis | 141 |
| 7.2 | VP Transmission with Optimal BD-GMD | 144 |
| 7.2.1 | System Model | 144 |
| 7.2.2 | Joint Filter Design and BD-GMD Application | 145 |
| 7.3 | VP Transmission with Suboptimal BD-GMD | 149 |
| 7.4 | Simulation Results | 153 |
| 7.4.1 | BER Performance | 154 |

| | | |
|----------|--|------------|
| 7.4.2 | Achievable Sum-Rate | 155 |
| 7.4.3 | MSE Measurement | 155 |
| 7.5 | Complexity Analysis | 157 |
| 7.5.1 | Number of FLOPs and Computational Complexity Analysis | 157 |
| 7.5.2 | Running-Time Analysis | 159 |
| 7.5.3 | Complexity-Performance Trade-Off | 160 |
| 7.6 | Chapter summary | 161 |
| 8 | Summary and Conclusions | 162 |
| 8.1 | Thesis Contributions | 164 |
| 8.2 | Suggestions for Further Research | 165 |
| A | Publications | 167 |
| B | Joint Filter Design Analysis | 169 |
| B.1 | Zero Forcing Based Design | 169 |
| B.2 | Minimum Mean Square Error Based Design | 171 |
| C | Derivation of Power Allocation Matrices for Sum-MSE Minimization | 174 |
| D | Derivation of Power Allocation Matrices for Sum-Rate Maximization | 176 |
| E | Suboptimal THP Approach Analysis | 179 |
| F | Decomposition Based Approach Analysis | 183 |
| | References | 184 |

List of Figures

| | | |
|------|--|----|
| 2.1 | Narrowband single user MIMO system with M transmit and N receive antennas. | 7 |
| 2.2 | Block diagram of the parallel decomposition of the MIMO channel (a) and parallel independent SISO channels (b). | 9 |
| 2.3 | MIMO-MAC system diagram (a) and block diagram (b) with K users each one with M_k antennas and an N -antenna BS. | 10 |
| 2.4 | MIMO-BC system (a) and simplified block diagram (b) of a system with K N_k -antenna users and an M -antenna BS. | 12 |
| 2.5 | Capacity region of a MU-MIMO MAC system with $K = 2$ users. | 14 |
| 2.6 | Capacity region for a 2-user channel with $M = 2$ and $N_k = 1$ computed by exploiting duality. | 15 |
| 2.7 | Three terminal relaying-based network. | 15 |
| 2.8 | Amplify-and-forward relaying strategy. | 17 |
| 2.9 | Decode-and-forward relaying strategy. | 18 |
| 2.10 | MIMO-MAC relaying system diagram with K users with M_k antennas each, a relay equipped with R antennas and an N -antenna BS. | 18 |
| 2.11 | MIMO-BC relaying system diagram with K users, each one with N_k antennas, a relay equipped with R antennas and an M -antenna BS. | 19 |
| 2.12 | Multiuser MIMO AF downlink relaying block diagram. | 20 |
| 2.13 | Tomlinson-Harashima precoding structure. | 23 |
| 2.14 | Vector precoding structure. | 24 |
| 3.1 | Block diagram of multiuser MIMO downlink AF linear relaying system. | 27 |
| 3.2 | BER performance of the ZF and MMSE linear precoding techniques with and without equalizer in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation. | 34 |
| 3.3 | BER performance of the ZF and MMSE linear precoding techniques with and without equalizer in the $6 \times 6 \times \{6 \times 1\}$ set-up (B) with QPSK modulation. | 35 |
| 3.4 | BER performance of the ZF and MMSE linear precoding techniques with and without equalizer in the $6 \times 6 \times \{2 \times 3\}$ set-up (C) with QPSK modulation. | 35 |

| | | |
|------|--|----|
| 3.5 | Achievable sum-rate of the ZF and MMSE linear precoding techniques with and without equalizer in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 36 |
| 3.6 | BER versus ρ_1 results for different iteration numbers of <i>Lin-MMSE</i> and <i>Lin-MMSE-EQ</i> linear approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation and $\rho_2 = 15$ dB. | 37 |
| 3.7 | MSE convergence of <i>Lin-MMSE</i> and <i>Lin-MMSE-EQ</i> linear approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 38 |
| 3.8 | Partial block diagonalized system composed by a BS transmitting through a relay to K multi-antenna users. | 43 |
| 3.9 | BER performance obtained by MMSE-BD approaches for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation. | 48 |
| 3.10 | BER performance obtained by MMSE-BD approaches for $8 \times 8 \times \{4 \times 2\}$ set-up (D) with QPSK modulation. | 49 |
| 3.11 | BER performance obtained by MMSE-BD approaches for $8 \times 8 \times \{2 \times 4\}$ set-up (E) with QPSK modulation. | 49 |
| 3.12 | Sum-rate achieved by MMSE-BD approaches in $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 50 |
| 3.13 | Sum-rate achieved by MMSE-BD approaches in $8 \times 8 \times \{4 \times 2\}$ set-up (D). | 51 |
| 3.14 | Achievable sum-rate of the MMSE block diagonalization technique with and without equalizer in the $8 \times 8 \times \{2 \times 4\}$ set-up (E). | 51 |
| 3.15 | BER performance obtained by MMSE-BD approaches for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_2=15$ dB for a fixed number of iterations. | 52 |
| 3.16 | Convergence of the MMSE block diagonalization technique with and without equalizer. | 53 |
| 3.17 | Running-time for <i>Lin-MMSE</i> and <i>Lin-MMSE-EQ</i> for $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ set-ups (A and B) when $\rho_2 = 15$ dB. | 55 |
| 3.18 | Running-time for <i>Lin-MMSE-BD</i> and <i>Lin-MMSE-EQ-BD</i> for $4 \times 4 \times \{4 \times 1\}$, $8 \times 8 \times \{4 \times 2\}$ and $8 \times 8 \times \{2 \times 4\}$ set-ups (A, D and E) when $\rho_2 = 15$ dB. | 57 |
| 4.1 | BER curves for <i>Lin-MMSE-BD-All-Pass-Filter</i> approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation. | 64 |
| 4.2 | Achievable sum-rate of <i>Lin-MMSE-BD-All-Pass-Filter</i> approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 65 |
| 4.3 | MSE error of <i>Lin-MMSE-BD-All-Pass-Filter</i> approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 66 |
| 4.4 | BER performance of the proposed <i>Lin-MMSE-HH</i> approach for the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation. | 69 |
| 4.5 | Comparison of sum-rate versus SNR for <i>Lin-MMSE-HH</i> approach for the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 70 |

| | | |
|------|--|-----|
| 4.6 | Convergence analysis by means of MSE error versus numbers of iterations for <i>Lin-MMSE-HH</i> approach for the $4 \times 4 \times \{4 \times 1\}$ set-up (A) when $\rho_1 = 10$ dB and $\rho_2 = 15$ dB. | 70 |
| 4.7 | Percentage of negative values obtained for the sum-MSE diagonalized system for $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ set-ups (A and B). | 78 |
| 4.8 | Percentage of negative values obtained for the sum-rate maximization system for $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ set-ups (A and B). | 83 |
| 4.9 | BER curves of sum-MSE diagonalized approaches for QPSK modulation and $4 \times 4 \times \{4 \times 1\}$ set-up (A) for different sets of users. | 85 |
| 4.10 | Sum-rate versus ρ_1 of diagonalization based approach for $4 \times 4 \times \{4 \times 1\}$ set-up (A) and $\rho_2 = 15$ dB for different sets of users. | 86 |
| 4.11 | BER performance comparison for diagonalization based approaches with QPSK modulation for $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 87 |
| 4.12 | Achievable sum-rate for diagonalization based approaches for $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 87 |
| 4.13 | Run-time for <i>Lin-MMSE-All-Pass-Filter</i> for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with $\rho_2 = 15$ dB. | 89 |
| 4.14 | Execution-time for <i>Lin-MMSE-HH</i> for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with $\rho_2 = 15$ dB. | 91 |
| 4.15 | Running-time for system diagonalization based algorithms for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with $\rho_2 = 15$ dB. | 92 |
| 5.1 | Multiuser MIMO downlink AF relaying block diagram for THP precoding. | 97 |
| 5.2 | Multiuser MIMO downlink AF relaying block diagram for VP precoding. | 103 |
| 5.3 | Incidence of the lattice values in the optimum solution vector for different SNR regions [Barrenechea12]. | 110 |
| 5.4 | BER performance for non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_2 = 15$ dB. | 111 |
| 5.5 | BER performance for non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_1 = 15$ dB. | 111 |
| 5.6 | BER performance curves for non-linear precoding techniques in the $8 \times 8 \times \{2 \times 4\}$ set-up (E) with QPSK modulation. | 112 |
| 5.7 | Sum-rate achieved by non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 113 |
| 5.8 | Comparison of BER vs. SNR for non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when they are forced to stop after a certain number of iterations. | 115 |
| 5.9 | Error obtained by non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 116 |

| | | |
|------|--|-----|
| 5.10 | Simulation time required for <i>THP-MMSE</i> , <i>VP-MMSE</i> and <i>VP-MMSE-BD</i> computation for $4 \times 4 \times \{4 \times 1\}$ set-up when $\rho_2 = 15$ dB. | 118 |
| 6.1 | BER performance versus SNR for MMSE non-linear suboptimal THP precoding technique in the $4 \times 4 \times \{4 \times 1\}$ with QPSK modulation. | 125 |
| 6.2 | Sum-rate versus SNR of <i>THP-MMSE-Sub</i> approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 125 |
| 6.3 | Mean square error of MMSE non-linear suboptimal THP precoding technique in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 126 |
| 6.4 | BER performance for MMSE non-linear VP suboptimal precoding techniques in $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation. | 133 |
| 6.5 | Sum-rate achieved by MMSE non-linear VP suboptimal precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 134 |
| 6.6 | Mean square error measurement for <i>VP-MMSE-HH</i> and <i>VP-MMSE-Sub</i> approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 135 |
| 6.7 | Run-time needed for <i>THP-MMSE-Sub</i> , <i>VP-MMSE-HH</i> and <i>VP-MMSE-Sub</i> algorithm for $4 \times 4 \times \{4 \times 1\}$ set-up when $\rho_2 = 15$ dB. | 138 |
| 7.1 | Columns and rows modification [Jiang05b] | 143 |
| 7.2 | BER performance for MMSE BD-GMD systems in the $4 \times 4 \times \{4 \times 1\}$ with QPSK modulation. | 154 |
| 7.3 | Achievable sum-rate for MMSE BD-GMD approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 155 |
| 7.4 | BER curves of MMSE BD-GMD in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_2 = 15$ dB and the number of iterations is fixed. | 156 |
| 7.5 | Error measurement for MMSE BD-GMD approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A). | 157 |
| 7.6 | Running-time for MMSE BD-GMD approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) when $\rho_2=15$ dB. | 159 |

List of Tables

| | | |
|-----|---|-----|
| 3.1 | Number of FLOPs for linear ZF and MMSE approaches. | 54 |
| 3.2 | Complexity order and number of FLOPs for linear ZF and MMSE approaches. | 55 |
| 3.3 | Number of FLOPs for <i>Lin-MMSE-BD</i> and <i>Lin-MMSE-EQ-BD</i> | 56 |
| 3.4 | Complexity order and number of FLOPs for <i>Lin-MMSE-BD</i> and <i>Lin-MMSE-EQ-BD</i> | 57 |
| 4.1 | Number of FLOPs for <i>Lin-MMSE-BD-All-Pass-Filter</i> | 88 |
| 4.2 | Complexity order and number of FLOPs when $n = 4$ for <i>Lin-MMSE-BD-All-Pass-Filter</i> | 89 |
| 4.3 | Number of FLOPs for <i>Lin-MMSE-HH</i> | 90 |
| 4.4 | Number of FLOPs when $n = 4$ and complexity order for <i>Lin-MMSE-HH</i> approach. | 90 |
| 4.5 | Complexity order for <i>Lin-min-MSE-Naive</i> , <i>Lin-min-MSE K=12</i> , <i>Lin-max-SR-Naive</i> and <i>Lin-max-SR K=12</i> approaches. | 91 |
| 5.1 | Number of FLOPs necessary for matrix computation by <i>THP-MMSE</i> , <i>VP-MMSE</i> and <i>VP-MMSE-BD</i> approaches. | 116 |
| 5.2 | Number of FLOPs for sphere encoder and modulo operation. | 117 |
| 5.3 | Complexity order of <i>THP-MMSE</i> , <i>VP-MMSE</i> and <i>VP-MMSE-BD</i> | 117 |
| 6.1 | Number of FLOPs required by <i>THP-MMSE-Sub</i> , <i>VP-MMSE-HH</i> and <i>VP-MMSE-Sub</i> approaches. | 136 |
| 6.2 | Complexity order for <i>THP-MMSE-Sub</i> , <i>VP-MMSE-HH</i> and <i>VP-MMSE-Sub</i> approaches when $N_B = 100$ and $T = \boldsymbol{\nu} = 25$ | 137 |
| 7.1 | Number of FLOPs required by BD-GMD approaches for the computation of \mathbf{F} and \mathbf{W} | 158 |
| 7.2 | Number of FLOPs for sphere encoder, modulo operation and GMD algorithm. | 158 |
| 7.3 | Complexity order for <i>VP-MMSE-BD-GMD</i> and <i>VP-MMSE-GMD-Sub</i> | 158 |

List of Algorithms

| | | |
|----|--|-----|
| 1 | Iterative algorithm for the computation of the linear precoding matrix \mathbf{F} and the relaying matrix \mathbf{W} under ZF criterion. | 30 |
| 2 | Search of \mathbf{F} and \mathbf{W}_1 under the MMSE criterion for the partial BD approach. | 46 |
| 3 | $x_{k,i}$ and $y_{k,i}$ derivation for the MSE minimization. | 77 |
| 4 | Computation of $x_{k,i}$ and $y_{k,i}$ under the MMSE criterion including the equally distribution of the power, when not all users can be served with the optimal approach. | 79 |
| 5 | User selection algorithm for sum-MSE minimization. | 81 |
| 6 | Estimation of $\bar{\mathbf{F}}$, $\bar{\mathbf{W}}$ and \mathbf{B} for THP MMSE approach. | 101 |
| 7 | Computation of \mathbf{B} under MMSE criterion. | 101 |
| 8 | Feedback loop. | 101 |
| 9 | Computation of $\bar{\mathbf{F}}$, $\bar{\mathbf{W}}$ and \mathbf{a} under the MMSE criterion. | 106 |
| 10 | Computation of $\bar{\mathbf{F}}$, $\bar{\mathbf{W}}$ and \mathbf{B} under MMSE criterion for suboptimal THP approaches. | 124 |
| 11 | Calculation of \mathbf{F} , \mathbf{W} and \mathbf{a} under the MMSE criterion for problem decomposition based approach. | 132 |
| 12 | Permutation process. | 143 |
| 13 | Geometric mean decomposition based on singular value decomposition approach. | 144 |
| 14 | Computation of \mathbf{F} , \mathbf{W}_1 and the perturbation vector \mathbf{a} under the MMSE criterion for BD-GMD based approach. | 149 |
| 15 | Search of \mathbf{F} and the perturbation vector \mathbf{a} under the MMSE criterion for suboptimal GMD approach. | 153 |
| 16 | Bisection method for the search of the water-level factor. | 178 |

Acronyms

| | |
|----------------|---|
| AF | amplify-and-forward |
| AWGN | additive white Gaussian noise |
| BER | bit error rate |
| BD | block diagonalization |
| BD-GMD | block diagonal geometric mean decomposition |
| BS | base station |
| BC | broadcast channel |
| CSI | channel state information |
| DF | decode-and-forward |
| DFE | decision feedback equalizer |
| DPC | dirty paper coding |
| DS-CDMA | direct spread-CDMA |
| FD | full-duplex |
| FLOP | floating-point operation |
| GMD | geometric mean decomposition |
| i.i.d | independent and identically distributed |
| ISI | intersymbol interference |
| HD | half-duplex |

| | |
|----------------|--|
| KKT | Karush-Kuhn-Tucker |
| MAC | multiple access channel |
| MMSE | minimum mean square error |
| MSE | mean square error |
| MIMO | multiple-input multiple-output |
| ML | maximum likelihood |
| MMSE | minimum mean square error |
| MRC | maximum ratio combining |
| MS | mobile stations |
| MU-MIMO | multiuser-MIMO |
| MUI | multiuser interference |
| NP-hard | non-deterministic polynomial-time hard |
| OCI | other-cell interference |
| QPSK | quadrature phase shift keying |
| RS | relay station |
| SE | sphere encoder |
| SISO | single-input single-output |
| SM | spatial multiplexing |
| SNR | signal-to-noise ratio |
| SR | sum-rate |
| SVD | singular value decomposition |
| THP | Tomlinson-Harashima precoding |
| VP | vector precoding |
| ZF | zero forcing |
| ZF-DPC | zero forcing-dirty paper coding |

List of Symbols

| | |
|------------------|---|
| \mathbf{a} | Perturbation vector |
| \arg | Argument |
| \mathbf{B} | Feedback filter |
| β_1 | First hop power scaling factor |
| β_2 | Second hop power scaling factor |
| C | Achievable sum-rate |
| \mathbb{C} | Complex value |
| \mathbf{D} | Equalization matrix |
| \mathbf{d} | Transmitted data vector by the BS |
| $ \mathbf{A} $ | Determinant of the matrix \mathbf{A} |
| \mathbf{D}_k | User k equalization matrix |
| \mathbf{d}_k | Transmitted data vector belonging to the user k |
| $E[\mathbf{A}]$ | Expectation of \mathbf{A} |
| E_{BS} | Total transmission power at the base station |
| ϵ_{min} | Minimum convergence error |
| E_R | Relay transmission power |
| \mathbf{F} | Precoding matrix |
| \mathbf{F}_k | Precoding matrix belonging to the user k |

| | |
|---------------------------|---|
| $\lfloor \bullet \rfloor$ | Floor operator |
| $\bar{\mathbf{F}}$ | Unscaled precoding matrix |
| \mathbf{H}_1 | First hop channel matrix |
| \mathbf{H}_2 | Second hop channel matrix |
| $\mathbf{H}_{2,k}$ | Second hop channel matrix created between the relay and user k |
| \mathbf{H}_e | Equivalent channel matrix |
| $\mathbf{H}_{e,k}$ | Equivalent channel matrix related to the user k |
| \mathbf{A}^H | Hermitian of the matrix \mathbf{A} |
| $\tilde{\mathbf{H}}_k$ | Channel matrix containing the channels of the interfering users |
| \mathbf{I}_N | N -dimensional identity matrix |
| $\Im(a)$ | Imaginary part of a |
| \mathbf{A}^{-1} | Inverse of the matrix \mathbf{A} |
| K | Total number of users |
| K_{max} | Maximum number of selected users |
| L | Total number of transmitted streams |
| λ_i | i^{th} singular value |
| $\lambda_{1,k,i}$ | Matrix $\bar{\mathbf{\Lambda}}_{1,k}$ i^{th} diagonal value |
| $\lambda_{e,k,i}$ | Matrix $\bar{\mathbf{\Lambda}}_{e,k}$ i^{th} diagonal value |
| L_k | Total number of transmitted streams per user |
| \log_2 | Base 2 logarithm |
| M | Transmit antennas |
| $\max(\bullet)$ | Maximum value |
| $\min(\bullet)$ | Minimum value |
| $\mathbf{M}(\bullet)$ | Modulo operation |
| μ_1 | Factor that satisfies the total power constraint at the first hop |

| | |
|-----------------------------|--|
| μ_2 | Factor that satisfies the total power constraint at the second hop |
| N | Total number of receiver antennas |
| \mathbf{n}_1 | First hop Gaussian noise vector |
| \mathbf{n}_2 | Second hop Gaussian noise vector |
| $\mathbf{n}_{2,k}$ | Second hop user k Gaussian noise vector |
| N_k | Number of receive antennas per user |
| N_{01} | Noise power spectral density at the first hop |
| N_{02} | Noise power spectral density at the second hop |
| ν_M | Voronoi fundamental region |
| ν_{QPSK} | QPSK constellation lattice |
| $\mathbf{\Omega}_k$ | User k power allocation matrix for the first hop |
| $\omega_{k,i}$ | i^{th} diagonal value of $\mathbf{\Omega}_k$ |
| $\mathbf{\Phi}_k$ | User k power allocation matrix for the second hop |
| $\phi_{k,i}$ | i^{th} diagonal value of $\mathbf{\Phi}_k$ |
| $Q(\bullet)$ | Detection process |
| R | Relaying antennas |
| $\text{rank}(\mathbf{A})$ | Matrix \mathbf{A} rank |
| \mathbf{R}_d | Input non-linear signal covariance matrix |
| \mathbb{R} | Real value |
| $\Re(a)$ | Real part of a |
| ρ_1 | First hop signal-to-noise ratio |
| ρ_2 | Second hop signal-to-noise ratio |
| R_k | Sum-rate achieved by user k |
| $\mathbf{R}_{\mathbf{n}_1}$ | First hop noise covariance matrix |
| $\mathbf{R}_{\mathbf{n}_2}$ | Second hop noise covariance matrix |

| | |
|--------------------------|--|
| \mathbf{R}_s | Input linear signal covariance matrix |
| $\mathbf{R}_{s,k}$ | Input linear signal covariance matrix belonging to the user k |
| \mathbf{R}_v | Feedback signal covariance matrix |
| \mathbf{R}_{y_1} | Relaying input signal covariance matrix |
| \mathbf{s} | Transmitted data vector |
| $\mathbf{\Lambda}_1$ | Matrix containing the first hop singular values |
| $\mathbf{\Lambda}_{e,k}$ | Matrix containing the equivalent channel singular values |
| \mathbf{s}_k | Transmitted data vector belonging to the user k |
| τ | Modulo constant |
| $\text{Tr}(\mathbf{A})$ | Trace of the matrix \mathbf{A} |
| \mathbf{U}_1 | Matrix containing the first hop left singular vectors |
| $\mathbf{U}_{e,k}$ | Matrix containing the equivalent channel left singular vectors |
| \mathbf{v} | Feedback symbol vector |
| \mathbf{V}_1 | Matrix containing the first hop right singular vectors |
| $\mathbf{V}_{e,k}$ | Matrix containing the equivalent channel right singular vectors |
| \mathbf{W}_1 | Relaying processing matrix for BD approach |
| $\mathbf{W}_{1,k}$ | Relaying processing matrix for BD approach related to the user k |
| $\mathbf{W}_{2,k}$ | Block diagonalization matrix related to the user k |
| \mathbf{W}_2 | Block diagonalization matrix |
| \mathbf{W} | R-dimensional relaying matrix |
| $\overline{\mathbf{W}}$ | Unscaled relaying matrix |
| ξ | Sum-MSE |
| ξ_1 | First hop sum-MSE |
| ξ_2 | Second hop sum-MSE |
| \mathbf{y}_1 | Received data vector at the relay |

$\mathbf{y}_{2,k}$ Received data vector at the k^{th} terminal

\mathbf{y}_{BS} Precoded symbols at the BS

Chapter 1

Introduction

The growing demand in coverage and data transmission rate entails a clear need to design and implement more efficient wireless networks that guarantee the high demands imposed by the emerging applications. Furthermore, considering the need of greater capacities and performance, the wireless technology MIMO has been very used for this purpose, increasing the spectral efficiency and the reliability of the network. As its acronym indicates, the terminals are equipped with multiple antennas. By means of these, the performance is improved with the use of diversity and multiplexing techniques.

Nowadays, it is more common to find wireless networks where multiple users share resources such as time, bandwidth or transmission power. Multiuser networks, where a base station sends information to multiple users and vice versa, outperform point-to-point systems, being more efficient and useful. Moreover, it should be underlined that multiuser MIMO systems can achieve spatial multiplexing gain at the base station without equipping the terminals with multiple antennas. By means of this, smaller area and lower cost terminals can be implemented, translating the design complexity to the base station.

The increased demand in coverage, speed, ubiquity, delay and/or power consumption makes clear the necessity of implement multihop or relaying-based networks which make possible the expansion of the network in fading environments, increasing the performance and reducing multiuser interference. In the last years, the research has focused on multiuser MIMO relaying systems as a consequence of the aforementioned benefits as well as the low power consumption that arouses. Nevertheless, the design complexity and the performance of multihop networks is determined by the relaying strategy, which can include or not the regeneration of the forwarded data.

1.1 Motivation and Objectives

The inclusion of relaying terminals in MIMO networks makes difficult the design of the terminals. Furthermore, if multiple users take part in the communication, the complexity increases considerably, supposing a challenge for the design. It is known that when multiple users share the wireless channel or resources, the non-cooperative acquisition of data is vital

to keep the scalability of the system. For this purpose, it is necessary the implementation of a processing stage at the base station. This processing stage, known as precoding, will make possible the non-cooperative acquisition of data. Furthermore, the design of the precoder in multiuser MIMO multihop frameworks is more complex. Indeed, filter design involves the design of two processors, one located at the base station and the other placed at the relaying terminal.

Up to now, linear processing [Li09] has been used for the joint design of the filters. Nevertheless, non-linear precoding techniques such as THP [Chae08b] are more-and-more employed in multihop networks. In [Schmidt05] is shown that vector precoding outperforms THP in multiuser MIMO frameworks. Until now, VP has not been considered for the joint design of the filters in multiuser MIMO relaying systems. Despite its better performance, this precoding technique presents several disadvantages. On one hand, the computation of the perturbation vector is known to be in the class of non-deterministic polynomial-time hard (NP-hard) problem, increasing significantly the computational complexity of the system. On the other hand, as a consequence of the modification of the Gaussian noise as it passes through the modulo operator, the distribution of the signal results unknown. Furthermore, due to the dependence of the processors, the design is more complex. Thus, in order to obtain a local optimal solution, it is necessary the application of an iterative algorithm, increasing in this way the complexity of the system.

1.1.1 Main Objectives

Taking into consideration the features previously exposed, the main objectives of the thesis are the following:

- Theoretical analysis of linear and non-linear precoding techniques in multiuser MIMO multihop systems.
- Design of suboptimal algorithms for complexity reduction.
- Application of channel decomposition techniques for user diagonalization employed for the optimal and suboptimal design of the filters.
- Complexity analysis and set-up simulation.

1.2 Thesis Contributions

This section describes the main contributions of this thesis, showing the publications associated with the main contributions:

- Suboptimal linear precoding design and analysis for multiuser MIMO relaying systems [Jimenez11] [Jimenez13a] (Submitted).
- Application of block diagonalization to multiuser multihop frameworks. Several designs based on BD have been proposed for linear precoding as well as for vector precoding [Jimenez13a] (Submitted).
- Linear suboptimal algorithms based on singular value decomposition have been proposed for the minimization of the mean square error and the maximization of the sum-capacity [Jimenez13a] (Submitted)
- Equally power distribution and user selection based algorithms have been implemented [Jimenez13a] (Submitted).
- Design and analysis of non-linear multiuser MIMO relaying schemes. The local optimal THP and VP solutions have been derived [Jimenez12a] [Jimenez12b].
- Design and implementation of suboptimal algorithms for the minimization of the computational complexity for non-linear precoding [Jimenez12b] [Jimenez11].
- Design and implementation of local optimal approaches for multiuser MIMO relaying set-ups based on geometric mean decomposition, employing linear and vector precoding techniques [Jimenez11] [Jimenez13b] (Submitted).
- Complexity analysis and simulation of the algorithms [Jimenez13a] [Jimenez13b] (Submitted).
- Several simulation results have been provided showing the bit error rate (BER) performance, the achievable sum-rate as well as the convergence of the iterative algorithms.

1.3 Organization of the Thesis

This dissertation is divided into eight chapters, being the first assigned for the introduction of the main research topic, the objectives and the motivation.

Chapter 2 sums up the main features of MIMO and multiuser MIMO. The capacity of MIMO point-to-point is introduced as well as downlink and uplink multiuser MIMO channels. Then, relaying-based systems are introduced, focusing on the main research topic of this dissertation. Taking into consideration the necessity of a processing stage at both transmitter and relaying terminals, linear and non-linear precoding techniques are introduced briefly.

Chapter 3 focuses on the joint linear design of the filters. On one hand, the local optimal expressions for the filters are derived from the joint optimization. Then, block diagonalization

is employed for interference cancellation, obtaining the processors jointly. The main difficulty of BD-based design is the lack of convergence. In order to give a solution, BD is applied at the relay for interference suppression while linear precoding is employed at the base station for user separation. As a consequence of the interdependence of the filters, an iterative process has to be applied to obtain the optimal solution. Finally Chapter 3 ends with a complete complexity analysis.

With the aim of minimizing the computational complexity achieved by the joint linear optimal approaches, suboptimal linear approaches are proposed in Chapter 4 for the design of the precoder and the relaying matrix. Three different approaches are given. The first one designs the relaying matrix to limit the power at this terminal without processing the incoming signal. By means of this approach, the optimization problem focuses on precoder design at the base station. The second approach is considered mainly for its simplicity. Throughout the minimization of the mean square error of each hop, the precoding and relaying matrices are obtained independently. Finally, singular value decomposition based approach is proposed. The channel created from the base station to the end users is diagonalized by means of this decomposition technique, creating multiple parallel subchannels for the minimization of the mean square error and the maximization of the sum-rate. This chapter concludes with the analysis of the computational cost, showing the effectiveness of these suboptimal approaches.

In the same way, Chapter 5 focuses on the analysis and the derivation of joint non-linear precoding and relaying matrices. The joint local optimal THP and VP solutions are proposed and VP is also presented in combination with BD for matrix derivation. Whilst BD is applied at the relaying terminal for interference suppression, VP is employed at the base station for mean square error minimization and to limit the power at both BS and relay station (RS) terminals. As pointed out before, due to the interconnection of the filters an iterative process has to be applied, increasing considerably the complexity of the system. Furthermore, for VP approaches the cost is specially high due to the search of the perturbation vector in each iteration as a consequence of the lack of knowledge of the statistics of the latter, whereas for THP based approaches the complexity increases owing to feedback loop executed for successive interference cancellation.

In order to reduced the complexity of the joint non-linear optimal approaches, Chapter 6 presents three suboptimal non-linear approaches. The first approach proposes the minimization of the mean square error of the second hop, obtaining the relaying matrix, whereas THP is applied at the base station for successive interference cancellation. Then, each hop mean square error is minimized independently, breaking with the interconnection of the precoding and relaying matrices. VP is added at the BS for total error minimization. Finally, the last proposal divides the optimization problem into master problem and sub-problem. While the latter optimizes directly the filters, the former obtains the optimal perturbation vector that

minimizes the overall error. This last approach still being iterative. Nevertheless, the search of the perturbation vector in each iteration is avoided. This chapter also finishes with a complete complexity analysis.

Chapter 7 analyses the application of geometric mean decomposition that divides the channel into multiple parallel subchannels with the same signal-to-noise ratio per user. VP is added at the base station, improving the performance of the system. This chapter concludes proposing suboptimal geometric mean decomposition (GMD) approaches for VP based design with the purpose of reducing the computational cost of the system.

Finally, Chapter 8 sums-up the results obtained in this research work and collects the main conclusions. Additionally, a complete list of publications is given in Appendix A.

Background and Related Work

2.1 Introduction

The inclusion of relays in fading environments with the aim of increasing the coverage of the networks has aroused a steady focus of research. Furthermore, due to the potential benefits of MIMO wireless technique, where multiple antennas are placed at the transmitter and receiver, the capacity and the performance of relay systems can be enhanced.

This dissertation focuses on the design of precoding algorithms for multiuser MIMO relaying systems, being the theoretical background related to the topic introduced in this second chapter. The chapter starts with a brief introduction to MIMO systems, where the channel capacity is analysed. Multiuser MIMO networks are then studied, giving a theoretical background of the uplink and downlink channels and their channel capacity. The chapter will proceed with a review of relaying or multihop networks, analysing the main relaying strategies. Finally, linear and non-linear multiuser precoding techniques will be described, since the extension of the latter will be the main topic of this dissertation.

2.2 MIMO

The growing demand in coverage or data transmission rate implies a clear need to design and implement more efficient wireless networks that guarantee the high demands imposed by the emerging applications.

Taking into account the necessity of getting higher capacities and performance, in the last years the research has focused on multi-antenna MIMO technology [Telatar99][Foschini98]. As it can be seen in Figure 2.1 and also as its acronym suggests, both the transmitter and receiver terminals are equipped with multiple antennas, increasing in this way the data transmission rate and the robustness of the system.

In single-input single-output (SISO) systems, due to the fading caused by the multipath propagation, the capacity and reliability are reduced. However, instead of suffering from multipath propagation, MIMO exploits it to increase the capacity through diversity or spatial

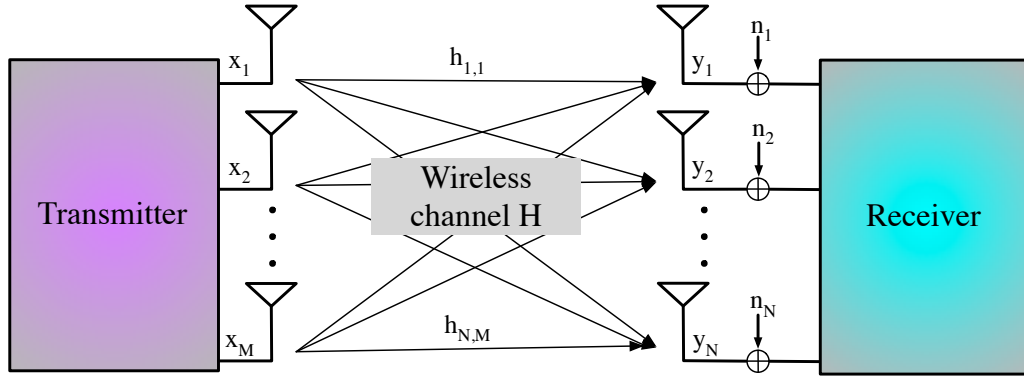


Figure 2.1: Narrowband single user MIMO system with M transmit and N receive antennas.

multiplexing (SM). On one hand, diversity gain can be achieved by implementing space-time coding, where signal redundancy is introduced in space and time, achieving in this way an improvement in both reliability and quality. On the other hand, SM multiplies the transmitted throughput with no penalty in bandwidth or transmission power. Furthermore, to get a higher performance, a trade-off between diversity and SM can be achieved.

A narrowband point-to-point communication system of M transmit and N receive antennas, depicted in Figure 2.1, can be described by the following equation:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,M} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N,1} & h_{N,2} & \cdots & h_{N,M} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{bmatrix},$$

or equivalently as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (2.1)$$

where \mathbf{x} represents the $M \times 1$ transmitted symbol vector, \mathbf{n} is the $N \times 1$ Gaussian noise vector with zero mean and covariance matrix $N_0\mathbf{I}_N$, and $\mathbf{H} \in \mathbb{C}^{N \times M}$ is the matrix of channel gains, whose elements $h_{n,m}$ represents the gain from transmit antenna m to receive antenna n . The entries of the channel matrix are assumed to have an independent and identically distributed (i.i.d) with zero-mean circularly symmetric Gaussian distribution with $E[|h_{n,m}|^2] = 1$.

2.2.1 MIMO Channel Capacity

The capacity of a channel represents the theoretical limit of the amount of data bits that can be transmitted without error over the communication link per second and bandwidth Hertz. In 1948, Claude Shannon [Shannon48] derived the capacity of SISO memoryless narrowband

channels, defined as:

$$C = \log_2(1 + \rho) \quad (\text{bits/s/Hz}),$$

where $\rho = E_{Tr}/N_0$ stands for the signal-to-noise ratio (SNR), being E_{Tr} and N_0 the transmit power and noise power spectral density, respectively.

As it is described in [Winters87] [Telatar99] [Foschini98], the capacity increases significantly with MIMO systems without increasing the transmission power or bandwidth. This is the main advantage of MIMO compared to SISO systems. In [Telatar99] capacity expressions for MIMO systems are proposed, which greatly outperform the expressions derived by Shannon for SISO set-ups. By means of these equations, it is shown that the spectral efficiency increases linearly with $\min(M, N)$, when the Rayleigh channel is unknown at the transmitter and the power is equally distributed among all the transmitted streams.

The capacity of MIMO flat deterministic channels for a given power constraint E_{Tr} stands for

$$C = \max_{\text{Tr}(\mathbf{R}_x) \leq E_{Tr}} \log_2(|\mathbf{I}_N + \mathbf{R}_n^{-1} \mathbf{H} \mathbf{R}_x \mathbf{H}^H|) \quad (\text{bits/s/Hz}),$$

where $\text{Tr}(\mathbf{A})$, $|\mathbf{A}|$ and \mathbf{A}^H are the trace, determinant and Hermitian of matrix \mathbf{A} , respectively. Additionally, $\mathbf{R}_n = \text{E}[\mathbf{nn}^H]$ and $\mathbf{R}_x = \text{E}[\mathbf{xx}^H]$ represent the noise and transmit symbol covariance matrices. Obviously, the transmit symbol covariance and the channel matrices play an important role at the derivation of the capacity. If there is no information about the channel at the transmitter, Telatar proposes to distribute the available power equally among the transmission streams [Telatar99].

2.2.2 Parallel Decomposition of the MIMO Channel

As mentioned, when the transmitter or receiver terminals are equipped with multiple antennas, these can be used for achieving diversity gain. When both transmitter and receiver are multi-antenna terminals, spatial multiplexing can be exploited for performance gain. The multiplexing gain of MIMO systems results from the fact that the MIMO channel can be decomposed into z parallel independent channels, being $z = \text{rank}(\mathbf{H})$ the number of independent channel rows or columns. Hence, we can get a z -fold increase in data rate in comparison with a system with just one antenna at the transmitter and receiver, corresponding this to the multiplexing gain.

If the channel is known at the transmitter, the most common way to get multiple parallel subchannels from \mathbf{H} is applying the singular value decomposition (SVD)[Golub96], following

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H,$$

where $\mathbf{U} \in \mathbb{C}^{N \times N}$ and $\mathbf{V} \in \mathbb{C}^{M \times M}$ correspond to the matrices that contain the left and right singular vectors, respectively. The diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{N \times M}$ contains the z non-zero

singular values of \mathbf{H} ordered in descending order. Each singular value, defined as λ_i , can be related with the eigenvalue of the matrix $\mathbf{H}\mathbf{H}^H$ as $\lambda_i = \sqrt{\psi_i}$, being ψ_i the mentioned eigenvalue. Since z cannot exceed the number of columns or rows of \mathbf{H} , $z \leq \min(M, N)$. In addition, if \mathbf{H} is full rank, the number of subchannels becomes $z = \min(M, N)$.

As depicted in Figure 2.2, the parallel decomposition of the channel is obtained by multiplying the channel input vector \mathbf{x} and the output vector \mathbf{y} through transmit precoding and receiver processing matrices \mathbf{V} and \mathbf{U}^H , respectively.

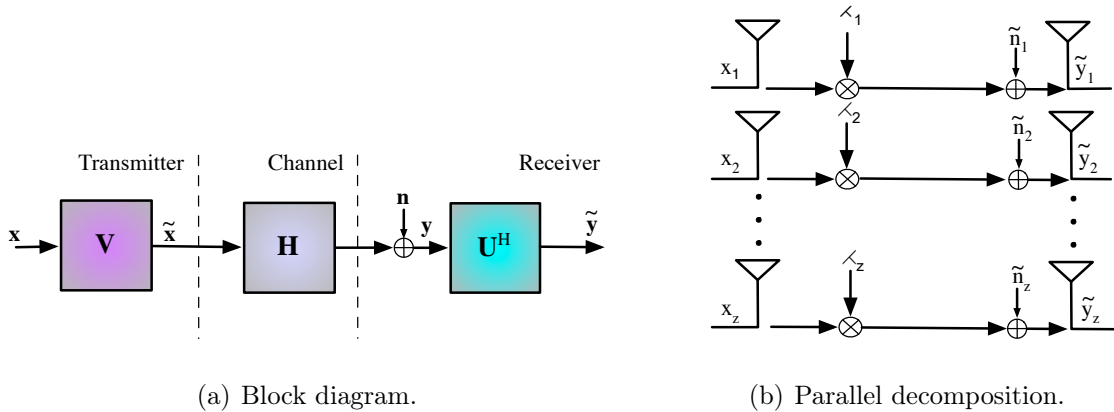


Figure 2.2: Block diagram of the parallel decomposition of the MIMO channel (a) and parallel independent SISO channels (b).

Therefore, the expressions for the precoded ($\tilde{\mathbf{x}}$) and processed symbols ($\tilde{\mathbf{y}}$) are

$$\begin{aligned}\tilde{\mathbf{x}} &= \mathbf{V}\mathbf{x}, & \text{and} \\ \tilde{\mathbf{y}} &= \mathbf{U}^H\mathbf{y}.\end{aligned}\tag{2.2}$$

By replacing the equations in (2.2) into the MIMO model introduced in (2.1), we can transform \mathbf{H} into z subchannels. Since \mathbf{U} and \mathbf{V} are unitary, the covariance matrices of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ are the same as for \mathbf{x} and \mathbf{y} . The new equivalent channel can be expressed as:

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{U}^H\mathbf{y} = \mathbf{U}^H\mathbf{H}\tilde{\mathbf{x}} + \mathbf{U}^H\mathbf{n} = \mathbf{U}^H\mathbf{U}\mathbf{\Lambda}\mathbf{V}^H\mathbf{V}\mathbf{x} + \tilde{\mathbf{n}}, \\ &= \mathbf{\Lambda}\mathbf{x} + \tilde{\mathbf{n}}.\end{aligned}$$

It should be pointed out that the resulting parallel channels do not interfere with each other so they can be considered SISO channels. Since the channel gain diminishes, the last subchannel will present a higher error probability or a lower capacity.

When SVD is applied in transmission, the capacity is defined as

$$C = \sum_{i=1}^z \log_2 \left(1 + \frac{P_i^* \lambda_i^2}{N_0} \right) \quad (\text{bits/s/Hz}),$$

where P_i^* is the water-filling power allocation defined as [Paulraj03]

$$P_i^* = \left(\mu - \frac{N_0}{\lambda_i^2} \right)^+,$$

with μ chosen to satisfy the total power constraint $\sum_{i=1}^z P_i^* = E_{Tr}$ and $(p)^+ = \max(p, 0)$. The optimal power allocation is found iteratively through the water-filling algorithm [Cover91].

2.3 Multiuser MIMO Communications

The efficiency of multiuser communications can be improved with the use of MIMO technique. Multiuser-MIMO, which combines multiuser networks and MIMO, presents a more realistic scenario in comparison with point-to-point MIMO networks because the users have to share resources such as time, bandwidth or transmission power. As it will be introduced below, multiuser-MIMO (MU-MIMO) scenarios are classified by the considered transmission channel, which can be either a multiple access channel (MAC) or broadcast channel (BC).

2.3.1 Multiple Access Channel

Figure 2.3 depicts a multiuser MIMO uplink or MAC channel with K user. Each user k , equipped with M_k antennas, transmits through the wireless channel $\mathbf{H}_k \in \mathbb{C}^{N \times M_k}$, which is the link between the k^{th} user and the BS, equipped with N antennas.

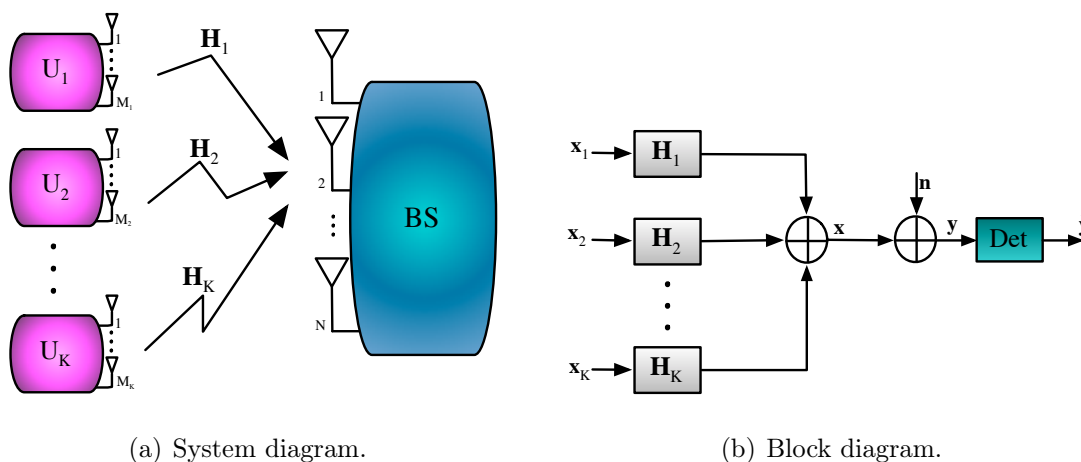


Figure 2.3: MIMO-MAC system diagram (a) and block diagram (b) with K users each one with M_k antennas and an N -antenna BS.

In the uplink channel, the total received signal at the BS is

$$\mathbf{y} = \sum_{k=1}^K \mathbf{H}_k \mathbf{x}_k + \mathbf{n},$$

where $\mathbf{x}_k \in \mathbb{C}^{M_k}$ is the signal sent by the k^{th} user, with covariance matrix $\mathbf{R}_{\mathbf{x},k}$ and $\text{Tr}(\mathbf{R}_{\mathbf{x},k}) \leq P_{Tr,k}$. Vectors $\mathbf{y} \in \mathbb{C}^N$ and $\mathbf{n} \in \mathbb{C}^N$ stand for the received signal and the additive white Gaussian noise (AWGN) with covariance $\mathbf{R}_{\mathbf{n}} = N_0 \mathbf{I}_N$, respectively.

As it can be seen in Figure 2.3, the BS bears the key element for the correct symbol estimation, i.e., the multiuser detector. By virtue of the latter, the BS will separate each users' signal. Depending on the channel and the detector used, the signal will be degraded by the other users' interference and the noise. In spite of the fact that the users do not send synchronized information, the synchronization between them is generally assumed perfect.

Due to the fact that the users share the medium, techniques such as time division multiple access (TDMA), frequency division multiple access (FDMA), space division multiple access (SDMA) or code division multiple access (CDMA) are used in order to access to the channel, being the last one the most employed in the literature [Moshavi96] [Khirallah06]. CDMA, which enables multiple simultaneous transmissions over the same time-bandwidth, presents an alternative in direct spread-CDMA (DS-CDMA). This comes from the fact that the signal occurs over the full bandwidth or spectrum, multiplying each users' signal by a sequence \mathbf{c}_k , being this exclusive for the k^{th} user. By this way, once the signal is received by the BS, it will have the ability to separate each user signal multiplying the total received signal by the code of each user, respectively. The codes applied in DS-CDMA can be either orthogonal or non-orthogonal, being Gold [Prasad96] codes the most referred ones due to their low cross-correlation.

As cited before, the main key for the correct estimation is the design of the detector used at the receiver side, which can be classified into conventional and multiuser detectors. The difference between these is that the former considers the interference as noise, whereas for the latter the interferences are taken into account at the estimation process. The conventional detector introduced in [Moshavi96] is composed by K matched filters, each one matched with the code of the user that has to be detected. The complexity of this kind of detectors in comparison with multiuser detectors is lower, but the performance is degraded because the interference cannot be completely removed.

Additionally, the optimum multiuser detector is the maximum likelihood (ML) detector [Verdú89]. The main drawback of this is that it searches symbols that minimize the Euclidean distance over the set of all possible transmitted symbol vectors. Obviously, this search increases the computational complexity, making the implementation of this kind of detectors harder. In order to reduce the cost, suboptimum detectors can be used. These are classified as linear, non-linear and heuristic. On one hand, among the linear detectors the decor-

relator (DEC) and minimum mean square error (MMSE) [Moshavi96] can be underlined. On the other hand, serial interference cancellation (SIC) and parallel interference cancellation (PIC) detectors [Moshavi96], where matched filters are used to separate the users and the interference is after cancelled serially or in parallel, respectively, are highlighted in the group of non-linear detectors. Finally, heuristic techniques such as genetic algorithms (GA) [Juntti97], ant colony optimization (ACO) [Xu08] or particle swarm optimization (PSO) [Kennedy95] are often used for the detection in multisuser networks.

2.3.2 Broadcast Channel

The MIMO downlink channel or BC [Spencer04a] depicted in Figure 2.4, is composed by a BS sending data to K multi-antenna users.

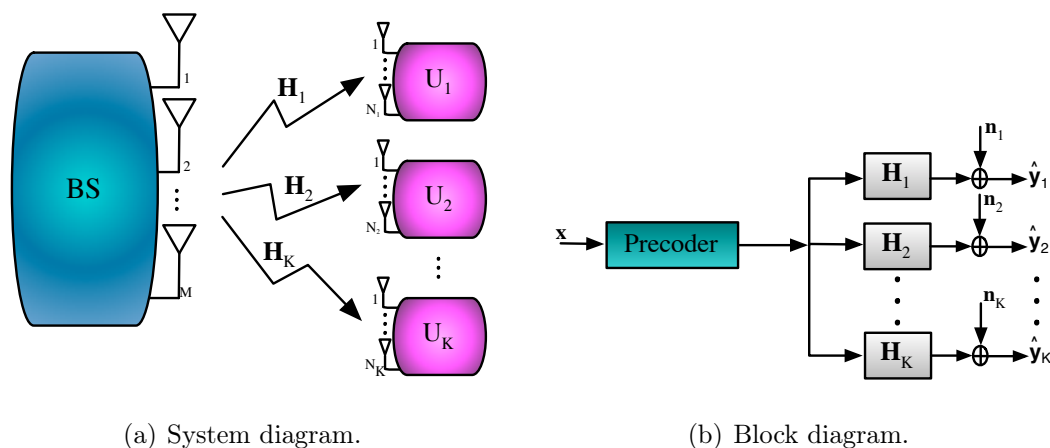


Figure 2.4: MIMO-BC system (a) and simplified block diagram (b) of a system with K N_k -antenna users and an M -antenna BS.

A BS equipped with M transmit antennas sends an independent data sequence for each user through the wireless channel $\mathbf{H} = [\mathbf{H}_1^T, \dots, \mathbf{H}_K^T]^T \in \mathbb{C}^{N \times M}$ being $N = \sum_{k=1}^K N_k$ the total number of receive antennas, where N_k stands for the amount of antennas at the k^{th} terminal and $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ is the channel established between the k^{th} user and the BS. Vector $\mathbf{n}_k \in \mathbb{C}^{N_k}$ denotes the AWGN noise introduced in the link between the BS and the k^{th} user with covariance $N_0 \mathbf{I}_{N_k}$. The main drawback of the downlink channel is that the signal is going to be degraded, not only by the noise introduced by the channel, but also by the interference of the signals targeted by other users since these share the same wireless medium. Due to the lack of cooperation between users, the decoding process plays an important role, making the design of the receiver terminals difficult. With the aim of eliminating the complexity at the receivers, the interference canceller, called *precoder* in BC systems, is moved to the BS. If the BS knows the wireless channel, the interferences can be pre-cancelled at the precoding stage. Several precoding strategies will be showed in more detail in Section 2.5.

2.3.3 MU-MIMO Channel Capacity Region

The capacity of MU-MIMO channel is characterized by the addition of the data rates that can be achieved by the users with arbitrarily low error probability. If the network is composed by K terminals, the capacity is defined by a K -dimensional vector, where each sub-vector is assigned to the data rate of an individual user. As a result of this analysis, the capacity definition introduced by Shannon for SISO systems is extended to the capacity region for multiuser systems, where this includes the set of K -dimensional vectors with the achievable data rates for every user of the network. The capacity analysis of MAC and BC channels will be described in this section.

2.3.3.1 MIMO-MAC Channel Capacity

The capacity region of MIMO-MAC was established in [Goldsmith03]. If joint decoding is used, which implies the cooperative decoding of the signals, the capacity region can be defined as:

$$C_{MAC} = \{(R_1, \dots, R_K) : \sum_{k \in \chi} R_k \leq \log_2 \left(|\mathbf{I}_N + \frac{1}{N_0} \mathbf{H}_\chi \mathbf{R}_{\mathbf{x}, \chi} \mathbf{H}_\chi^H| \right)\} \quad (\text{bits/s/Hz}),$$

where χ is a subset of the set of users $(1, \dots, K)$. Matrix $\mathbf{R}_{\mathbf{x}, \chi}$ corresponds to the input covariance matrix of the terminals included in the subset χ , while R_k is the k^{th} user data rate. Matrix \mathbf{H}_χ contains the channels related to the users grouped into subset χ .

Figure 2.5 shows the capacity region of a $K = 2$ MU-MIMO system, where both single-antenna users transmit data through the wireless channel to the BS. It is known that different points of the capacity region can be achieved applying SIC detection at the receiver side. By means of this, inter-user interference is cancelled reaching the maximum achievable sum-rate described by the line that joins the points A and B, being A the point where the interference of the second user is totally cancelled and B the one where the interference of the first user is completely removed. R_1 and R_2 show the data rate for the first and second users, respectively, when the interference is completely removed.

2.3.3.2 MIMO-BC Channel Capacity

The capacity analysis of MIMO-BC channel is more complex. For single antenna multiuser MIMO networks, the achievable capacity region is introduced in [Caire03]. The extension for multi-antenna systems is analysed in [Yu01] with the application of dirty paper coding (DPC) [Costa83], achieving in this way the maximum sum-rate for the downlink channel. If the transmitter, in this case the BS, knows the non-casual interference to a given user, then the capacity of the channel is the same as if there was no interference [Goldsmith03]. By means

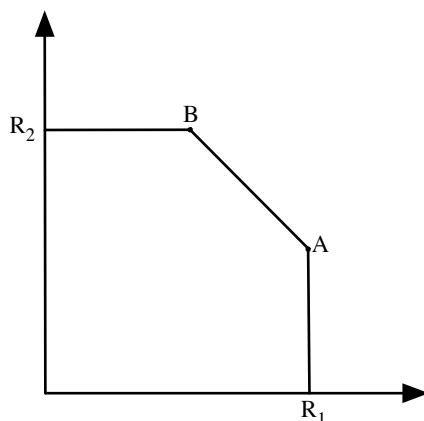


Figure 2.5: Capacity region of a MU-MIMO MAC system with $K = 2$ users.

of DPC, the interference is cancelled before sending data without any penalty in transmission power.

The capacity region is then the convex hull of the union of all such rate vectors over all permutations and all positive semi-definite covariance matrices, satisfying the power constraint. The main drawback about the rate is that the equations defining it can be either concave or non-concave.

In order to simplify the computation of the capacity related to the downlink channel, the existing duality among the BC and MAC channels can be exploited, which establishes that the achievable rates for MIMO-MAC and MIMO-BC are equivalent as long as the sum power constraints are equal. The duality was first proposed in [Vishwanath02] for constant scalar channels while [Jindal04] proposed the duality for multi-antenna systems. Figure 2.6 shows the capacity region for a system with $M = 2$ transmission antennas and two single-antenna users. The capacity region is defined by the outer boundary and the lines inside this boundary correspond to the capacity region of a different dual MIMO-MAC, whose sum-power equals the MIMO-BC's power.

2.4 Multiuser MIMO Relaying Communications

The increased demand in coverage, speed, ubiquity, delay and consumption requires the design and implementation of more reliable wireless networks. In this context, the combination of MIMO and relaying techniques can provide improved coverage and robustness against fading, thanks to the user diversity.

At this point a new concept turns up, MIMO relaying networks. Figure 2.7 shows the typical one-hop relaying network, where a transmission source sends data through a RS to a destination terminal. Furthermore, the source can also establish communication with the

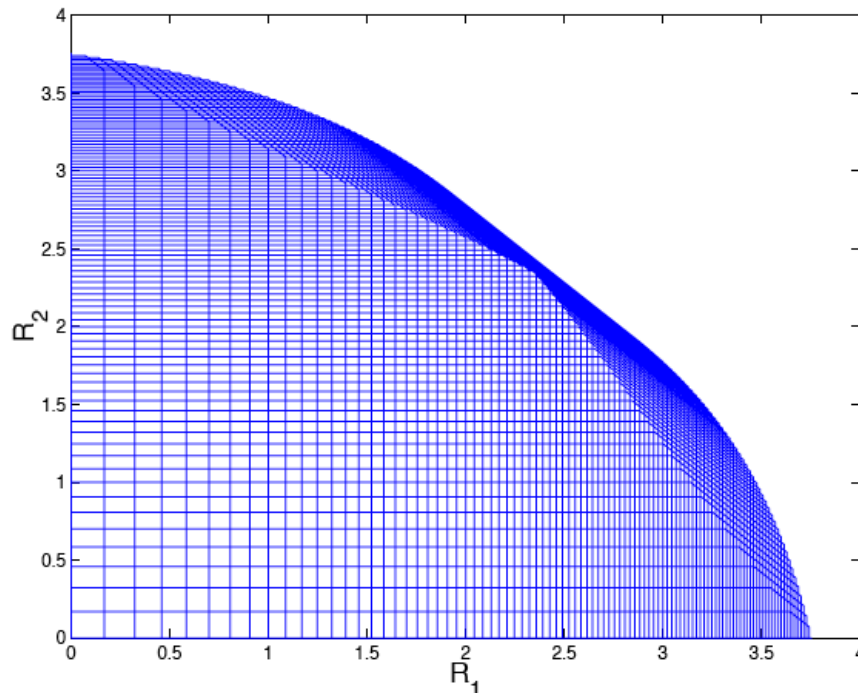


Figure 2.6: Capacity region for a 2-user channel with $M = 2$ and $N_k = 1$ computed by exploiting duality.

destination across a direct link. However, the direct link is omitted in relaying analyses because the lack of communication between the source and the destination is commonly considered.

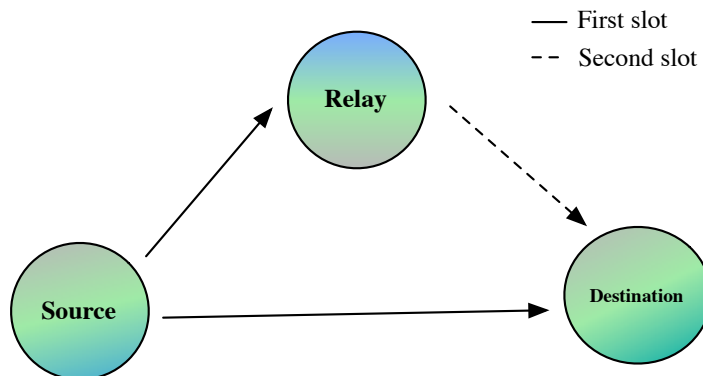


Figure 2.7: Three terminal relaying-based network.

The networks based on fixed relays have been deeply studied. In 1971, Van der Meulen [Van Der Meulen71] introduces three terminal networks while some years later, Cover improved Van der Meulen's results getting the lower and upper capacity bounds for a network composed by three elements: source, relay and destination in [Cover79]. With hindsight, in [Laneman04] a review of the past and recent information-theoretic work of the relay channel has been made.

As it is shown in Figure 2.7, the communication between the transmitter and receiver is established in two time slots. Thus, the RS can receive and transmit on separate time-frequency resources. By means of this communication strategy, also known as half-duplex (HD) relaying, the transmitter will send data to the relay and through the direct link in the first slot, while the relay will retransmit the information in the second slot keeping the source in silence. Usually HD is considered instead of full-duplex (FD) because of its simplicity. While the former establishes the communication in two phases, in the latter the relay has the ability of receiving and transmitting at the same time, supposing a design challenge. Indeed, a large isolation of transmitter and receiver or a interference canceller are needed in order to reduce the error probability. For example, the analysis of the achievable rates for a multi-antenna system is carried out under full duplex transmission in [Ju09], proposing algorithms for interference cancellation.

Hitherto, the research of relaying networks has focused on single user networks, where the transmitter and receiver establish their communication through a RS station, being the main research topic the design of the filters located at the source and the RS. In [Guan08] the design of such filters is done through the minimization of the mean square error (MSE), decomposing the MIMO channels into multiple SISO channels. On the other hand, in [Mo09] the design is also done by means of the minimization of MSE, but the optimum design is proposed through the primal decomposition technique [Boyd04].

2.4.1 Relaying Systems

Once the signal is received by the RS, it is processed before retransmission in order to reduce the error probability at the receiver side. Different strategies can be applied for this purpose [Fan07], being amplify-and-forward (AF) and decode-and-forward (DF), which will be described in the next section, the most extended in the literature.

Other approaches exist such as hybrid relaying (HR), which decodes a training sequence from the source obtaining full channel information and then filters the received signal based on the knowledge of the channel without decoding it. In [Cover79] compress-and-forward (CF), estimate-and-forward (EF) and quantize-and-forward (QF) protocols are proposed, where a three terminal MIMO system with DF strategy is proposed, deriving the capacity when the channel is physically degraded. While CF allows the relay node to compress the received signal before transmitting it to the destination node, being Wyner-Ziv coding known as the optimal compression algorithm, in EF the relay sends an estimate of its channel input to the destination without decoding the source message at all. Finally, in QF the relay quantizes a finite number of bits per sample.

2.4.1.1 Amplify-and-Forward

The aforementioned AF protocol [Tang07] is one of the most extended relaying strategies due to its simplicity, being its main drawback its performance. This technique, also known as non-regenerative relaying, amplifies the received signal at the RS and retransmits it to the destination.

As depicted in Figure 2.8, \mathbf{y}_R is the signal transmitted by the relay, scaled by the factor β^{-1} , which is the amplification factor that ensures that the transmission power at the relay equals E_R . The amplification factor is defined as

$$\beta = \sqrt{\frac{\text{Tr}(\mathbf{y}_1 \mathbf{y}_1^H)}{E_R}},$$

where E_R is the total transmission power at the relay. Once the amplification power is calculated, the signal transmitted by the RS becomes $\mathbf{y}_R = \beta^{-1} \mathbf{y}_1$, being \mathbf{y}_1 the signal received by the relay.

The main problem of this strategy and the main reason why the performance is reduced is the noise amplification. Hence, noisy environments are not suitable for AF relaying strategy.

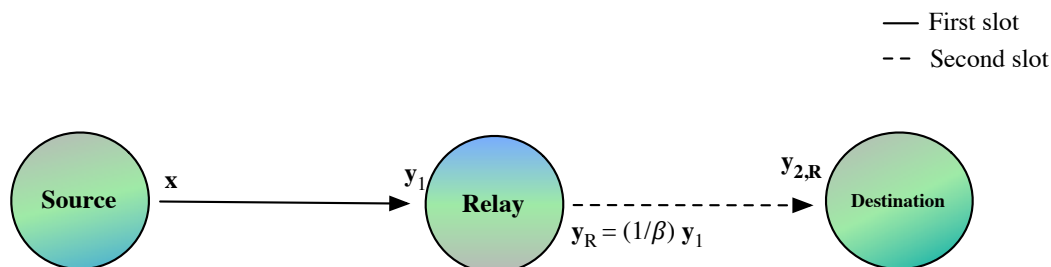


Figure 2.8: Amplify-and-forward relaying strategy.

2.4.1.2 Decode-and-Forward

This technique, known as regenerative relaying due to the correction of the forwarded signal, executes four steps: the received signal \mathbf{y}_1 is detected, decoded and re-encoded. Finally the regenerated signal $\hat{\mathbf{y}}_1$ is forwarded to the destination node. In [Fan07] a SIC detector is used for data detection and decoding. Furthermore, if the direct link is better than the link established between the source and the RS, the latter will not have the ability of improving the transmission. In this PhD dissertation AF relaying algorithm has been chosen for simplicity.

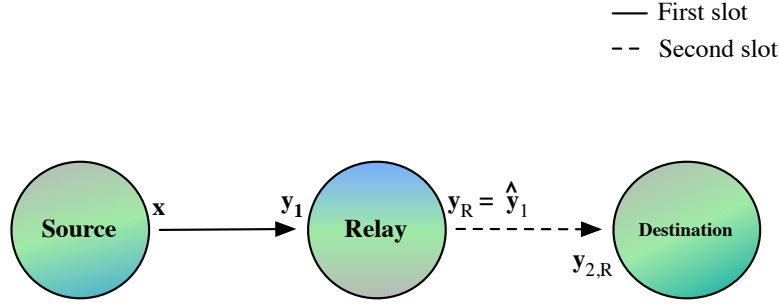
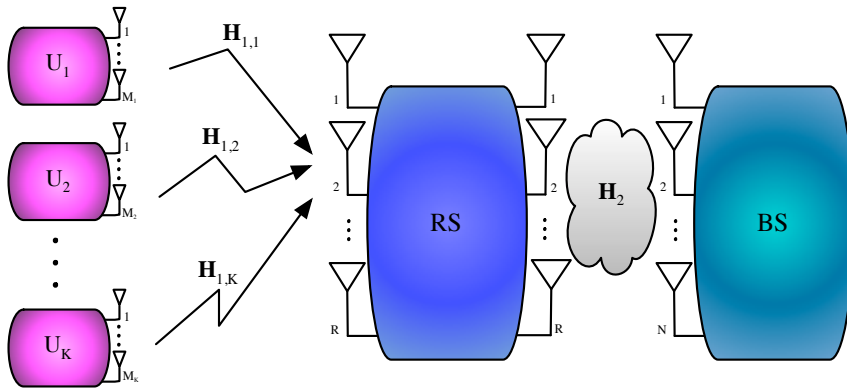


Figure 2.9: Decode-and-forward relaying strategy.

2.4.2 Multiuser MIMO Relaying Model

In the last years the presence of multiple users in relaying schemes has aroused a great interest due to the good results obtained in both performance and capacity. The main drawback of this kind of schemes is that the transceiver design becomes more difficult because they have to share resources such as transmission power or bandwidth.

As happens with MU-MIMO systems, MU-MIMO multihop networks can be classified as either downlink or uplink. Figure 2.10 shows the uplink system where K multi-antenna users send data through a fixed relay equipped with R antennas, which after the application of the appropriate relaying strategy forwards the signal to a base station equipped with N receive antennas.

Figure 2.10: MIMO-MAC relaying system diagram with K users with M_k antennas each, a relay equipped with R antennas and an N -antenna BS.

In the same way, Figure 2.11 depicts the downlink or broadcast channel where an M -antenna BS sends information to K multi-antenna users across a RS station equipped with R antennas. It should be pointed out that this second scheme will be analysed in detail in Section 2.4.2.1 because it has been chosen as the main scenario for this thesis. For both transmission schemes, the communication is assumed half-duplex.

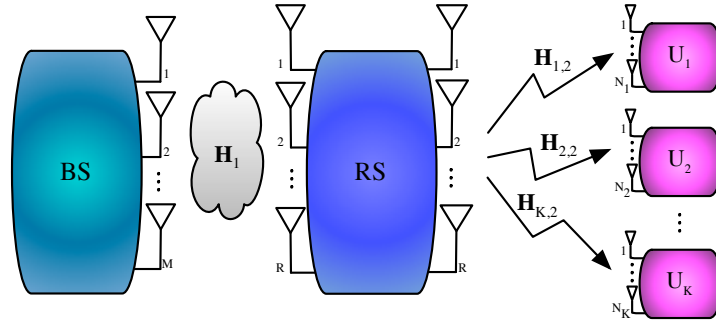


Figure 2.11: MIMO-BC relaying system diagram with K users, each one with N_k antennas, a relay equipped with R antennas and an M -antenna BS.

In [Rong11] the uplink MU-MIMO relaying scheme is analysed for a multihop system where the terminals are multi-antenna. Following the MMSE criterion, the optimal relaying matrix is derived at each relay. As happens with single-user relaying systems, the research focuses on the design of the filters at the BS and RS. An example of this is described in [Li09], where the filters are designed for the MSE minimization from the BS to the end users. Both filters, due to the interdependence of the matrices, are obtained following an iterative procedure. To reduce the computational cost of the filter optimization procedure, a power allocation approach is also proposed through the SVD decomposition of the first channel and assuming an all-pass filter relay.

In [Yu10] and [Chae08b] a theoretical analysis is presented which obtain the optimal filters that maximize the sum-rate. While [Yu10] evaluates both uplink and downlink channels, [Chae08b] pays attention to the BC scenario in order to get upper and lower bounds of the achievable sum-rate assuming zero forcing-dirty paper coding (ZF-DPC). In order to overcome DPC's complexity, THP is proposed at the BS with adaptive modulation.

For MU-MIMO relaying systems, the capacity computation is more complex because MAC and BC concepts are combined. The uplink multiuser relaying capacity is analysed in [Sankaranarayanan04], where outer bounds for a discrete memoryless multiple access relay channels are obtained for CF and AF strategies. The sum-rate for the downlink channel is studied in [Tang06] for single-antenna users and a multi-antenna fixed relay. The proposed algorithms compute the achievable sum-rate based on DPC, for which a lower bound is derived.

2.4.2.1 Multiuser MIMO AF Downlink Relaying

Figure 2.12 depicts a multiuser MIMO-BC AF relaying system. The communication between the BS and the end users across a RS is established in two time slots. In the first slot or phase, a BS equipped with M antennas sends information through the wireless channel \mathbf{H}_1 , whereas in the second time slot an R -antenna relay node forwards the received information

to K multi-antenna mobile stations (MS), equipped with N_k antennas each. It should be pointed out that the total number of receive antennas is $N = \sum_{k=1}^K N_k$. As in the rest of this research work, the direct link is neglected for simplicity.

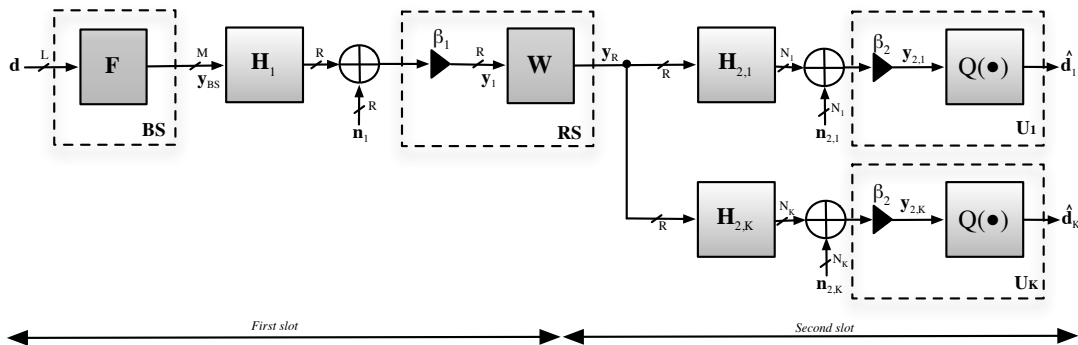


Figure 2.12: Multiuser MIMO AF downlink relaying block diagram.

The strategy employed by the relay to forward the information is AF, where the incoming data is only multiplied by the relaying matrix $\mathbf{W} \in \mathbb{C}^{R \times R}$. In the same way, due to the lack of cooperation between users, in order to separate the data before transmission, a precoding process is applied at the BS, which is carried out throughout the matrix $\mathbf{F} \in \mathbb{C}^{M \times L}$. A scaling factor β_1^{-1} is included in \mathbf{F} in order to limit the power to E_{BS} . In such case, the transmitted signal vector is defined as follows:

$$\mathbf{y}_{BS} = \mathbf{F}\mathbf{d},$$

where $\mathbf{d} = [\mathbf{d}_1^T, \dots, \mathbf{d}_K^T]^T \in \mathbb{C}^L$ is the users' data to be transmitted, being L the total number of streams. In order to guarantee enough degrees of freedom, $L = \min(N, R, M)$ is assumed. Each user sends L_k streams, being $\mathbf{d}_k \in \mathbb{C}^{L_k}$ the symbols transmitted by the k^{th} user. Taking into account the lack of correlation between the input streams, the input signal covariance matrix is defined as $\mathbf{R}_d = \mathbb{E}[\mathbf{d}\mathbf{d}^H] \in \mathbb{R}^{L \times L}$. The symbols are extracted from a quadrature phase shift keying (QPSK) constellation lattice $\nu_{QPSK} = \{c = a + jb : a, b \in \{-1/\sqrt{2}, 1/\sqrt{2}\}\}$.

In the first phase, the data streams travel across the first hop channel $\mathbf{H}_1 \in \mathbb{C}^{R \times M}$ and are received at the relay, where the signal after the re-scaling is

$$\begin{aligned} \mathbf{y}_1 &= \beta_1 \mathbf{H}_1 \mathbf{y}_{BS} + \beta_1 \mathbf{n}_1, \\ &= \beta_1 \mathbf{H}_1 \mathbf{F} \mathbf{d} + \beta_1 \mathbf{n}_1, \end{aligned} \quad (2.3)$$

being $\mathbf{n}_1 \in \mathbb{C}^R$ the AWGN noise introduced at the first hop with covariance matrix equal to $\mathbf{R}_{\mathbf{n}_1} = \mathbb{E}[\mathbf{n}_1 \mathbf{n}_1^H] = N_{01} \mathbf{I}_R$ and N_{01} the power spectral density of the noise. It has to be

underlined that β_1 corresponds to the first hop power scaling factor. After processing the signal by the relaying matrix \mathbf{W} , which includes a β_2^{-1} scaling factor to limit the power to E_R , it is sent to multiple MS as

$$\mathbf{y}_R = \beta_1 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{s} + \beta_1 \mathbf{W} \mathbf{n}_1.$$

Once the signal is received at the final terminal and re-scaled by β_2 , the k^{th} user will obtain the following signal:

$$\begin{aligned} \mathbf{y}_{2,k} &= \beta_2 \mathbf{H}_{2,k} \mathbf{y}_R + \beta_2 \mathbf{n}_{2,k}, \\ &= \beta_2 \beta_1 \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{d} + \beta_2 \beta_1 \mathbf{H}_{2,k} \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{n}_{2,k}. \end{aligned}$$

Furthermore $\mathbf{H}_{2,k} \in \mathbb{C}^{N_k \times R}$ is the second hop channel between the k^{th} user and the relay, and $\mathbf{n}_{2,k} \in \mathbb{C}^{N_k}$ is the AWGN noise introduced in the second hop with covariance matrix $\mathbf{R}_{\mathbf{n}_{2,k}} = \mathbb{E} [\mathbf{n}_{2,k} \mathbf{n}_{2,k}^H] = N_{02} \mathbf{I}_{N_k}$. If $\mathbf{H}_2 = [\mathbf{H}_{2,1}^T, \dots, \mathbf{H}_{2,K}^T]^T \in \mathbb{C}^{N \times R}$, the total received signal can be defined as

$$\mathbf{y}_2 = [\mathbf{y}_{2,1}^T, \dots, \mathbf{y}_{2,K}^T]^T = \beta_2 \beta_1 \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{d} + \beta_2 \beta_1 \mathbf{H}_2 \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{n}_2, \quad (2.4)$$

being $\mathbf{n}_2 \in \mathbb{C}^N$ defined as $\mathbf{n}_2 = [\mathbf{n}_{2,1}^T, \dots, \mathbf{n}_{2,K}^T]^T$.

At each receiver, a detection process is carried out by means of a nearest-neighbour quantizer $Q(x) = \operatorname{argmin}_{c \in \mathcal{V}_{QPSK}} |x - c|^2$, which provides hard decisions on the transmitted data symbols:

$$\hat{\mathbf{d}}_k = Q(\beta_2 \beta_1 \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{d} + \beta_2 \beta_1 \mathbf{H}_{2,k} \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{n}_{2,k}).$$

Obviously β_1 and β_2 have to be known at the RS and MSs, respectively for re-scaling the received signal. By means of feedback channels or training symbols, relay and mobile stations can gain this knowledge. If $\beta_i > 1$ for $i = 1, 2$, the power of the noise vector is amplified, increasing in this way the probability error.

Finally, the SNRs for the first and second hops, respectively, are defined as

$$\rho_1 = \frac{E_{BS}}{\operatorname{Tr}(\mathbf{R}_{\mathbf{n}_1})}, \quad \text{and} \quad \rho_2 = \frac{E_R}{\operatorname{Tr}(\mathbf{R}_{\mathbf{n}_2})}.$$

Simulation results provided in this dissertation assume a unit transmit power constraint per antenna at the BS and RS, i.e, $E_{BS} = M$ and $E_R = R$.

2.5 Precoding techniques

With the aim of reducing the receiver complexity and due to the lack of cooperation between users, the signal processing complexity is transferred to the BS by means of a processing stage

called *precoding*. If the base station knows the channel, the interference can be suppressed before transmission. Combining the precoding phase and the linear filtering at the relay, each user will receive an optimized interference-free signal.

In this section, the multiuser MIMO precoding algorithms are introduced and classified briefly, which will be helpful in the following chapters.

2.5.1 Linear Precoding

Linear precoding multiplies the users signal by a matrix which finds a trade-off between interference nulling and noise reduction. Mainly, linear precoders can be classified as either zero forcing (ZF) or MMSE solution.

ZF criterion cancels the interference between the users completely. By inverting the channel matrix, it presents a good performance at high SNR environments or when the number of users is high enough [Peel05]. The main drawback of ZF is the power increment of the precoded symbols, chiefly at ill-conditioned channels [Peel05], which requires the use of a large power-reduction factor, impacting directly in the detection SNR.

If a limited number of interference or crosstalk between different user streams is admitted, more efficient solutions can be achieved. The optimal regularization factor is derived in [Joham05] following an MMSE or *Wiener* problem formulation. This solution finds a trade-off between noise enhancement and interference by means of a regularized inverse of the channel.

2.5.2 Non-Linear Precoding Techniques

Non-linear precoding techniques improve the performance of linear processing [Hochwald05]. The main drawback of these schemes is that the real implementation is more expensive due to the complexity of the algorithms.

As it can be seen in [Peel05], DPC derives the capacity of the interfering channels when the interference is known at the transmitter. The main problem of DPC is that the increased complexity makes the implementation impossible. In order to reduce the computational cost, THP and VP, both non-linear techniques are generally used, which tend to reach DPC's performance at lower computational cost.

2.5.2.1 Tomlinson-Harashima Precoding

THP was first proposed by Tomlinson [Tomlinson71] and Harashima [Harashima72] for the intersymbol interference (ISI) channel, while Fischer adapted it to MIMO systems in 2002 [Fischer02b]. In order to mitigate DPC's computational complexity, which makes the implementation in real time systems impossible, THP is proposed for multiuser MIMO system precoding, reducing the cost with a reasonable performance. The idea of THP comes from

the decision feedback equalizer (DFE) [Proakis87] filter considered for interference cancellation in point-to-point MIMO systems at the receiver side. Mainly, THP translates DFE filter to the transmitter side for multiuser interference cancellation under the assumption that the channel is perfectly known.

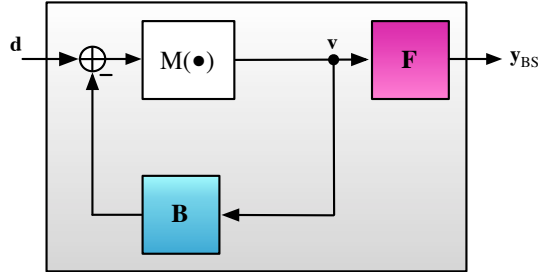


Figure 2.13: Tomlinson-Harashima precoding structure.

As depicted in Figure 2.13, THP is composed by two linear filters: the feedback filter \mathbf{B} and the feedforward filter \mathbf{F} . The former is a lower triangular matrix with zeros in the main diagonal. This structure, which is established to ensure the feedback loop, is also known as *spatial causality* [Joham06], ensuring that only data symbols which have already been precoded are fed back.

The basic idea of THP is the successive interference cancellation, being the symbols of different users precoded one after another. Initially the symbol of the first user is sent unaltered. After that, the symbol of the second user is transmitted taking into account the interference caused by the first one. This procedure is then continuously executed with the rest of users.

Due to the interference cancellation process, the transmission power increases. In order to reduce it, a modulo operator $M(\bullet)$ is applied at both transmitter and receiver, as shown in Figure 2.13. The aim of the modulo operation is to move the symbols to a lower consumption region and it can be defined as

$$M(b) = b - \left\lfloor \frac{\Re(b)}{\tau} + \frac{1}{2} \right\rfloor \tau - j \left\lfloor \frac{\Im(b)}{\tau} + \frac{1}{2} \right\rfloor \tau \in \mathbb{V},$$

where b is the input symbol to the modulo operator while $\Re(b)$ and $\Im(b)$ stand for the real and imaginary parts of b . Moreover, τ is the modulo constant, which depends on the modulation used, being $2\sqrt{2}$ for QPSK, $8/\sqrt{10}$ for 16QAM or $16/\sqrt{10}$ for 64QAM. $\mathbb{V} = \tau\mathbb{CZ}$ denotes the Voronoi fundamental region, where $\mathbb{CZ} = \mathbb{Z} + j\mathbb{Z}$ represents the set of Gaussian integers. The fundamental Voronoi region is a square of side τ centred at the origin, namely $\nu_M = \{x + jy \mid x, y \in [-\tau/2, \tau/2)\}$. Hence, the real and imaginary parts of b are mapped in the interval $[-\tau/2, \tau/2)$, where the power consumption is lower, entailing a better SNR detection. Finally $\lfloor \bullet \rfloor$ is the floor operator, which gets the largest integer smaller than or

equal to its argument.

The design based on THP precoding can be also done following ZF [Fischer02b][Joham04][Joham06] or MMSE [Joham04][Joham06] criteria, where the latter finds a trade-off between interference nulling and noise reduction.

2.5.2.2 Vector Precoding

In THP each symbol traverses the modulo operator located inside the loop. This procedure can be seen as the addition of values of the type $a\tau + jb\tau$, which can be considered as a perturbation vector. Furthermore, the linear filters \mathbf{F} and \mathbf{B} depicted in Figure 2.13 can be combined in a filter \mathbf{F}' , depicted in Figure 2.14, removing the THP transmission loop. Hence, VP is considered a generalization of THP.

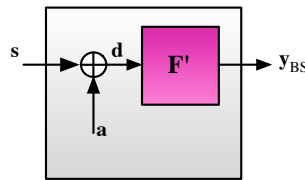


Figure 2.14: Vector precoding structure.

VP was introduced in [Hochwald05] as a simple encoding algorithm that achieves near sum-rate. Basically, VP's main objective is the minimization of the unscaled transmission power, or what is the same, the minimization of the scaling factor that amplifies the noise. Throughout the minimization of the power, the precoded symbols are mapped into a lower consumption region, improving in this way the SNR.

Unlike THP, VP optimizes the perturbation vector directly instead of applying an iterative loop for user cancellation [Schmidt05], considering all the possible perturbation vectors. The optimum vector is searched as the closest point in a lattice, which is a well-documented problem in the literature. Lattice searches have been shown to be NP-hard problems that grow exponentially in complexity with dimensionality [Agrell02].

A solution can be obtained by means of a sphere encoder (SE) [Shao05], which searches only a certain number of vectors of the lattice that are into an hypersphere instead of analysing all the points of the lattice.

2.6 Chapter summary

This chapter has reviewed the main features of MIMO, multiuser MIMO and multiuser MIMO relaying communications, being the latter the main research topic of this dissertation. The chapter has started with a brief analysis of MIMO technology, which is employed for data

transmission rate and robustness incrementation. Furthermore, MIMO channel capacity has been shortly introduced.

Afterwards, the framework has been extended to multiuser MIMO environments, presenting a more realistic scenario where the users share resources such as time, bandwidth or transmission power. A close network classification has been done, labelling multiuser MIMO networks as MAC (uplink) and BD (downlink).

As a consequence of multiuser scenarios the signal is degraded due to the interference created by other users and the noise introduced by the channel. Moreover, the capacity of MIMO-MAC and MIMO-BC systems have been briefly described, introducing the capacity region, which stands for the addition of all the rates achieved by the users.

Once MIMO and multiuser MIMO have been introduced, the main research topic of this dissertation has been given, which is the design of the filters located at the source and the relay in AF multiuser MIMO downlink channels. Then, the multiuser MIMO relaying model is presented, where the main research scenario for this thesis has been introduced.

To conclude precoding techniques have been analysed, being part of the topic of this dissertation. Firstly linear precoding is studied, where a linear transformation is carried out. Non-linear precoding techniques have been next analysed, because of their good performance in multiuser MIMO and their relevance for this research work, being THP and VP processing reviewed.

Linear Precoding in Multiuser MIMO AF Relaying Systems

As it has been introduced before, MIMO systems promise a significant capacity and performance improvement, which can enable parallel transmission of different data streams. When multiple users share the wireless medium, a pre-processing stage called *precoding* has to be exerted due to the absence of cooperation between users in order to adapt the transmitted signal to the channel, cancelling in this way the users' interference. In order to reduce the cost, linear pre-processing can be exploited for complexity reduction. Furthermore, in order to reduce the noise effect, a linear *equalization* process can also be applied at each receiver terminal.

This chapter considers the design of multiuser MIMO AF downlink relaying systems with linear precoding at the BS under the minimization of the mean square error of the complete transmission, subject to two power constraints in order to limit the transmission power at the BS and RS, respectively.

Throughout this chapter the joint linear design of precoding and relaying matrices will be analysed in more detail. In the first section of this chapter, the joint local optimal design is proposed under MSE minimization. Afterwards, a block diagonalization technique is proposed obtaining very good results when multiple antennas are used at each receive terminal. To show the effectiveness of the proposed joint linear designs, simulation results are provided as BER versus SNR, achievable sum-rate and convergence.

3.1 System Model

Henceforward, the set-up used for the analysis of the joint linear transceiver design depicted in the Figure 3.1 will be assumed for the linear optimization problem, where a BS equipped with M antennas sends data through an R -antenna relay to K multi-antenna users, each one equipped with N_k antennas for $k = 1, \dots, K$. The communication is established in two time slots: while in the first one the communication between the BS and the RS is accomplished,

in the second, after the processing at the relay, the signal is forwarded to the multi-antenna users. From now on, the direct link between the base station and the end users is neglected for simplicity. Apart from this, both terminals, base station and relay, have full channel state information of the first and second hops.

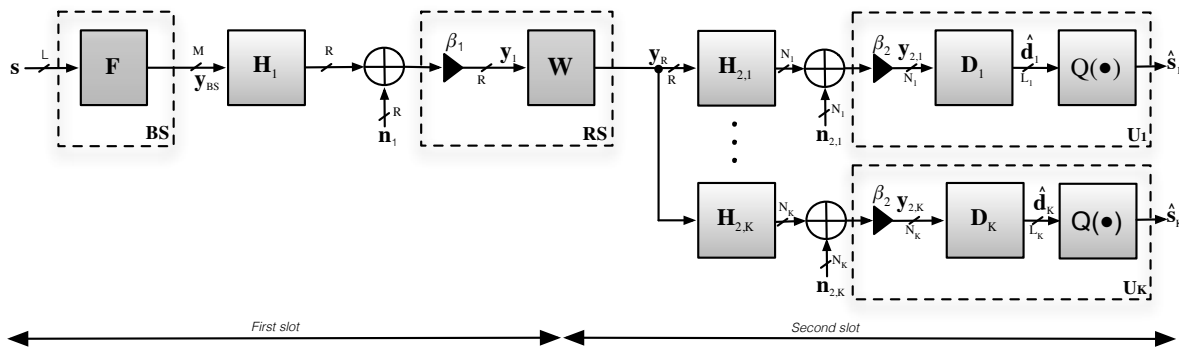


Figure 3.1: Block diagram of multiuser MIMO downlink AF linear relaying system.

The multiuser MIMO AF downlink relaying scheme used for the joint design has been introduced in Section 2.4.2.1. If the possibility of including linear equalizer at the receivers is considered, (2.4) becomes:

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}, \quad (3.1)$$

being β_1 and β_2 the scaling factors that constrain the total transmission power to E_{BS} and E_R at the first and second hops, respectively. Matrix $\mathbf{D}_k \in \mathbb{C}^{L_k \times N_k}$ stands for the equalization matrix, which can be grouped into the matrix $\mathbf{D} = \text{blkdiag}(\mathbf{D}_1, \dots, \mathbf{D}_K) \in \mathbb{C}^{L \times N}$, where L_k denotes the total number of transmitted streams for user k , being $L = \sum_{k=1}^K L_k$ the total number of transmitted streams. In order to provide enough degrees of freedom, $L_k = \min(M, R, N_k)$ is assumed. From now on, if $L_k = N_k$, $\mathbf{D}_k = \mathbf{I}_{L_k \times N_k}$ is considered. The vector $\hat{\mathbf{d}}_k$ is piled up into the vector $\hat{\mathbf{d}} = [\hat{\mathbf{d}}_1^T, \dots, \hat{\mathbf{d}}_K^T]^T \in \mathbb{C}^L$ containing K L_k -vectors belonging to the users data. Therefore, the expression for the total received signal vector is:

$$\hat{\mathbf{d}} = \beta_2 \beta_1 \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{s} + \beta_2 \beta_1 \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{D} \mathbf{n}_2.$$

Assuming linear precoding, the symbols are transmitted without adding any perturbation vector, so that $\mathbf{d} = \mathbf{s}$. Thus, $\mathbf{y}_{BS} = \mathbf{F} \mathbf{s}$ is considered for linear precoding, being $\mathbf{F} \in \mathbb{C}^{M \times L}$ the precoding matrix at the BS and $\mathbf{s} = [\mathbf{s}_1^T, \dots, \mathbf{s}_K^T]^T \in \mathbb{C}^L$ represents the user symbols containing K vectors. The channel matrix $\mathbf{H}_2 = [\mathbf{H}_{2,1}^T, \dots, \mathbf{H}_{2,k}^T, \dots, \mathbf{H}_{2,K}^T]^T \in \mathbb{C}^{N \times R}$ contains the channel matrix gains for the second hop, where $\mathbf{H}_{2,k} \in \mathbb{C}^{N_k \times R}$ stands for the channel created between the k^{th} user terminal and the relay. It should be pointed out

that $N = \sum_{k=1}^K N_k$ represents the total number of receive antennas.

In this chapter, the local optimal transceiver is designed under the sum-MSE minimization criterion, being the optimization metric defined as:

$$\xi = \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{s}_k\|_2^2 \right], \quad (3.2)$$

where $\mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{s}_k\|_2^2 \right]$ denotes the mean square error between the BS and the k^{th} user terminal, while $\|\bullet\|_2^2$ represents the Euclidean norm. Replacing (3.1) in (3.2), we get

$$\begin{aligned} \xi &= \sum_{k=1}^K \text{Tr}(\beta_2^2 \beta_1^2 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{R}_s \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{H}_{2,k}^H \mathbf{D}_k^H + \beta_2^2 \beta_1^2 N_{0_1} \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{W}^H \mathbf{H}_{2,k}^H \mathbf{D}_k^H \\ &+ \beta_2^2 N_{0_2} \mathbf{D}_k \mathbf{D}_k^H - 2\Re(\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{R}_{\mathbf{s}_k}) + \mathbf{I}_{L_k}), \end{aligned}$$

where N_{0_1} and N_{0_2} are the values for noise variances for the first and second hops, respectively and $\mathbf{R}_s = \mathbb{E}[\mathbf{s}\mathbf{s}^H] \in \mathbb{C}^{L \times L}$ is the signal covariance matrix. As introduced before, if there is no correlation between users, we assume that $\mathbf{R}_s = \mathbf{I}_L$. Finally, if we take into consideration the properties of the trace operation, the MSE can be computed as:

$$\begin{aligned} \xi &= \beta_2^2 \beta_1^2 \text{Tr}(\mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{H}_2^H \mathbf{D}^H) + \beta_2^2 \beta_1^2 N_{0_1} \text{Tr}(\mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{W}^H \mathbf{H}_2^H \mathbf{D}^H) \\ &+ \beta_2^2 N_{0_2} \text{Tr}(\mathbf{D} \mathbf{D}^H) - 2\Re(\text{Tr}(\beta_2 \beta_1 \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F})) + L. \end{aligned} \quad (3.3)$$

3.2 Joint Local Optimal Linear Transceiver Design

The joint linear design of precoding and relaying matrices is analysed in this section under ZF and MMSE criteria. It should be underlined that Sections 3.2.1 and 3.2.2 consider $L = N$, being the equalization matrix $\mathbf{D} = \mathbf{I}_{L \times N}$, whereas in Section 3.2.3, for the case $L \neq N$ equalization matrix is derived for each user. Getting a global solution is complicated due to the non-convexity of the optimization problem. A local solution can be obtained through the numerical method introduced in [Fang06], where a local optimal solution is derived in an iterative fashion. Following this numerical method, \mathbf{W} and β_2 are considered known for the derivation of \mathbf{F} and β_1 , and vice versa. For this reason, the approaches analysed in this section are presented as local solution, for which an iterative algorithm has to be applied.

3.2.1 Zero-Forcing Criterion

ZF approach is a very simple filter scheme that projects the channel vectors of each receiver onto the orthogonal subspace spanned by the channel vectors of all other receivers. Put in

other words, the interuser interference is completely cancelled under the ZF design criterion.

The main drawback of the design based on ZF, despite its simplicity, is the noise enhancement, since it disregards the noise term and focuses on perfectly removing the interference term. For ZF design, the constraint $\beta_2\beta_1\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{F} = \mathbf{I}_L$ has to be applied to (3.3), obtaining the following sum-MSE expression:

$$\xi_{ZF} = \beta_2^2\beta_1^2N_{01}\text{Tr}(\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{W}^H\mathbf{H}_2^H\mathbf{D}^H) + \beta_2^2N_{02}\text{Tr}(\mathbf{D}\mathbf{D}^H). \quad (3.4)$$

The optimization problem for the search of the precoding and relaying processing matrices is then defined as follows:

$$\{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\} = \underset{\{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\}}{\text{argmin}} \xi_{ZF} \quad (3.5a)$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{y}_{BS}\mathbf{y}_{BS}^H) = E_{BS}, \quad (3.5a)$$

$$\text{Tr}(\mathbf{y}_R\mathbf{y}_R^H) = E_R, \quad (3.5b)$$

$$\beta_2\beta_1\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{F} = \mathbf{I}_L, \quad (3.5c)$$

where ξ_{ZF} is defined in (3.4). The optimization problem is subject to three constraints. While (3.5a) and (3.5b) are the ones established to limit the total transmission power at the BS and RS, respectively, the constraint (3.5c) is due to the ZF interference cancellation.

The solution to the optimization problem is obtained applying the Lagrangian multipliers [Boyd04] as shown in Appendix B.1. Put in other words, the method of Lagrangian multipliers ignores the dependence between variables and the optimization function is optimized under all the variables, adding some penalty terms to the objective function, which ensures the constraints.

Carrying out the procedure described in Appendix B.1, we get the next expression for the relaying matrix:

$$\mathbf{W} = \frac{1}{\beta_2} \underbrace{(\mathbf{H}_2^H\mathbf{D}^H\mathbf{D}\mathbf{H}_2)^{-1}}_{\mathbf{A}} \mathbf{H}_2^H\mathbf{D}^H \underbrace{\bar{\mathbf{F}}^H\mathbf{H}_1^H (\mathbf{H}_1\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{H}_1^H)^{-1}}_{\mathbf{B}}, \quad (3.6)$$

where the term \mathbf{A} stands for the Moore-Penrose pseudo-inverse of the forward channel $\mathbf{H}_2\mathbf{D}$, whereas \mathbf{B} is the Moore-Penrose pseudo-inverse of the backward channel $\mathbf{H}_1\bar{\mathbf{F}}$, being $\bar{\mathbf{F}} = \beta_1\mathbf{F}$. The scaling factor β_2 is defined as follows:

$$\beta_2 = \sqrt{\frac{\text{Tr}\left(\left(\mathbf{D}\mathbf{H}_2\mathbf{H}_2^H\mathbf{D}^H\right)^{-1}\left(\mathbf{I}_L + \frac{N_{01}\text{Tr}(\bar{\mathbf{F}}\bar{\mathbf{F}}^H)}{E_{BS}}\left(\bar{\mathbf{F}}^H\mathbf{H}_1^H\mathbf{H}_1\bar{\mathbf{F}}\right)^{-1}\right)\right)}{E_R}}.$$

In the same way the precoding matrix is defined as

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H, \quad (3.7)$$

which stands for the pseudo-inverse of matrix $\mathbf{D} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1$ and $\overline{\mathbf{W}} = \beta_2 \mathbf{W}$ is the unscaled version of the relaying matrix. In addition to this, the scaling factor for the first hop is

$$\beta_1 = \sqrt{\frac{\text{Tr} \left(\left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \right)^{-1} \right)}{E_{BS}}},$$

being E_{BS} the total transmit power at the BS.

Due to the dependence between (3.7) and (3.6), an iterative algorithm has to be applied, getting the local optimal \mathbf{F} and \mathbf{W} until convergence is reached, since the value of MSE convergence is monotonously decreasing. Thus, the iterative algorithm is defined as follows, being ϵ_{\min} the minimum error that has to be achieved, ϵ the error updated in each iteration and l the current iteration. Note that this iterative algorithm entails a cost that increases with the number of users or transmit antennas.

Algorithm 1 Iterative algorithm for the computation of the linear precoding matrix \mathbf{F} and the relaying matrix \mathbf{W} under ZF criterion.

- 1: Initialization: $\overline{\mathbf{F}}^0 = \mathbf{I}_{M \times L}$, $\overline{\mathbf{W}}^0 = \mathbf{I}_R$ and $\epsilon = \infty$.
 - 2: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 3: Calculate $\overline{\mathbf{F}}^{l+1}$ as defined in (3.7) for a fixed $\overline{\mathbf{W}}^l$.
 - 4: Calculate $\overline{\mathbf{W}}^{l+1}$ following (3.6) for the updated $\overline{\mathbf{F}}^{l+1}$.
 - 5: Compute the convergence error: $\epsilon = \|\overline{\mathbf{F}}^{l+1} - \overline{\mathbf{F}}^l\|^2 + \|\overline{\mathbf{W}}^{l+1} - \overline{\mathbf{W}}^l\|^2$.
 - 6: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 7: $\overline{\mathbf{F}}^l = \overline{\mathbf{F}}^{l+1}$.
 - 8: $\overline{\mathbf{W}}^l = \overline{\mathbf{W}}^{l+1}$.
 - 9: **end if**
 - 10: **end while**
-

3.2.2 Minimum Mean Square Error Criterion

The MMSE approach, also known as *Wiener* solution, offers better performance than ZF designs due to the inclusion of noise in the optimization. The main aim of the design based on minimum mean square error criterion is the joint design of the precoding and relaying matrices that minimize the overall MSE defined in (3.3), providing the best balance between interference mitigation and the noise enhancement. Therefore, the optimization problem is

defined by:

$$\begin{aligned} \{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\} &= \underset{\{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\}}{\operatorname{argmin}} \xi \\ \text{s.t.} \quad \operatorname{Tr}(\mathbf{y}_{BS}\mathbf{y}_{BS}^H) &= E_{BS}, \end{aligned} \quad (3.8a)$$

$$\operatorname{Tr}(\mathbf{y}_R\mathbf{y}_R^H) = E_R. \quad (3.8b)$$

The solution can be again obtained by means of Lagrangian multipliers, which provide a local maxima and minima of a function subject to equality constraints (3.8a) and (3.8b). This derivation has been included in Section B.2 of the Appendix B, where the following expression is obtained for the relaying matrix

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \frac{N_{02} \operatorname{Tr}(\mathbf{D}\mathbf{D}^H)}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H \mathbf{D}^H \bar{\mathbf{F}}^H \mathbf{H}_1^H \mathbf{A}^{-1}, \quad (3.9)$$

where $\mathbf{A} = \left(\mathbf{H}_1 \bar{\mathbf{F}} \bar{\mathbf{F}}^H \mathbf{H}_1^H + \frac{N_{01} \operatorname{Tr}(\bar{\mathbf{F}}\bar{\mathbf{F}}^H)}{E_{BS}} \mathbf{I}_R \right)$.

Comparing (3.9) with (3.6), which is the ZF solution for the processing at the relay, it can be seen that MMSE expression involves a regularized pseudo-inverse. This regularization introduces a bias that gives more reliable results than ZF when the matrix is ill-conditioned and/or the channel is noisy. Furthermore, the power scaling factor at the relay becomes:

$$\beta_2 = \sqrt{\frac{\operatorname{Tr} \left(\mathbf{A}^{-1} \mathbf{H}_1 \bar{\mathbf{F}} \left(\mathbf{D} \mathbf{H}_2 \mathbf{H}_2^H \mathbf{D}^H + \frac{N_{02} \operatorname{Tr}(\mathbf{D}\mathbf{D}^H)}{E_R} \mathbf{I}_R \right)^{-2} \mathbf{D} \mathbf{H}_2 \mathbf{H}_2^H \mathbf{D}^H \bar{\mathbf{F}}^H \mathbf{H}_1^H \right)}{E_R}}.$$

In the same way, taking \mathbf{W} and β_2 fixed, we set the derivatives of the Lagrangian function respect to \mathbf{F} and β_1 . Executing the steps described in Section B.2 of the Appendix B, we get:

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \mathbf{B} \mathbf{H}_1 + \frac{N_{01} \operatorname{Tr}(\mathbf{B})}{E_{BS}} \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H, \quad (3.10)$$

and

$$\beta_1 = \sqrt{\frac{\operatorname{Tr} \left(\mathbf{D} \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \left(\mathbf{H}_1^H \mathbf{B} \mathbf{H}_1 + \frac{N_{01} \operatorname{Tr}(\mathbf{B})}{E_{BS}} \mathbf{I}_M \right)^{-2} \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \right)}{E_{BS}}},$$

where $\mathbf{B} = \bar{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \frac{N_{02} \operatorname{Tr}(\mathbf{D}\mathbf{D}^H)}{E_R} \mathbf{I}_R \right) \bar{\mathbf{W}}$ and matrix $\bar{\mathbf{W}} = \beta_2 \mathbf{W}$ is the unscaled relaying matrix.

As it happens with the ZF criterion, due to the interdependence of (3.10) and (3.9) an iterative approach has to be applied. At each iteration l , the precoding and relaying matrices are updated until a convergence error is reached. Thus, the iterative algorithm can be summed up as described in Algorithm 1, where \mathbf{F} and \mathbf{W} are computed as (3.10) and (3.9).

3.2.3 Receive Equalization Based Design

The classical way to deal with the distortions generated by the channel and the perturbation caused by the noise is the receiver processing, also known as *equalization*. In multiuser schemes, the equalization process can be specially beneficial when the number of transmitted streams is different to the number of receive antennas.

As it happens for the design of BS and RS processing algorithms, the equalizer can be designed under either the ZF or MMSE criterion, which are described in the following sections.

3.2.3.1 Zero-Forcing Equalizer

The design based on ZF considers an interference-free signal by means of the constraint $\beta_2\beta_1\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{F} = \mathbf{I}_L$. The optimization problem is defined in (3.5), being the error between the BS and the end users after the application of ZF constraint defined in (3.4).

The ZF equalization matrix at each user is obtained by means of the derivative of the Lagrangian respect to \mathbf{D}_k , getting:

$$\mathbf{D}_k = \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_{2,k}^H \left(\mathbf{H}_{2,k} \bar{\mathbf{W}} \mathbf{H}_1 \bar{\mathbf{F}} \bar{\mathbf{F}}^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_{2,k}^H \right)^{-1}, \quad (3.11)$$

where $\bar{\mathbf{F}}_k \in \mathbb{C}^{M \times L_k}$ is the precoding matrix belonging to the user k , grouped into the matrix $\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_k, \dots, \mathbf{F}_K]$.

Due to the dependence of (3.7), (3.6) and (3.11), an iterative approach is established as described in the Algorithm 1 being \mathbf{D} updated after $\bar{\mathbf{W}}$.

3.2.3.2 Minimum Mean Square Error Equalizer

In the same manner, the design of the equalizer can be included in the joint linear design under the MMSE criterion. The MMSE equalizer at each user is defined by the next expression:

$$\mathbf{D}_k = \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_{2,k}^H \left(\mathbf{H}_{2,k} \mathbf{A} \mathbf{H}_{2,k}^H + \frac{N_{02}}{E_R} \text{Tr}(\mathbf{A}) \right)^{-1}, \quad (3.12)$$

where $\mathbf{A} = \overline{\mathbf{W}} \left(\mathbf{H}_1 \overline{\mathbf{F}} \overline{\mathbf{F}}^H \mathbf{H}_1^H + N_{01} \frac{\text{Tr}(\overline{\mathbf{F}} \overline{\mathbf{F}}^H)}{E_{BS}} \mathbf{I}_R \right) \overline{\mathbf{W}}^H$. Analysing deeply (3.12), we can assert that the Wiener equalizer tends to the expression (3.11) when ρ_1 and ρ_2 tend to infinity.

3.2.4 Simulation Results

In this section, simulation results are presented for the multiuser MIMO AF system with linear processing. The BER performance is evaluated for the linear ZF and Wiener solutions with and without equalizer, fixing the SNR of one hop each time. Moreover, the achievable sum-rate and the convergence results of the system will be given.

The simulations assume M antennas at the BS, R antennas at the RS and K users, each one with N_k antennas, being N the total number of receive antennas. From now on, the set-ups will be known as $M \times R \times \{K \times N_k\}$. Three different scenarios have been chosen to validate the performance of the evaluated set-ups:

- A. $M = 4, R = 4, K = 4$ and $N_k = 1$, i.e, $4 \times 4 \times \{4 \times 1\}$.
- B. $M = 6, R = 6, K = 4$ and $N_k = 6$, i.e, $6 \times 6 \times \{6 \times 1\}$.
- C. $M = 6, R = 6, K = 2$ and $N_k = 3$, i.e, $6 \times 6 \times \{2 \times 3\}$.

It should be underlined that $4 \times 4 \times \{4 \times 1\}$ set-up will be a constant in the remaining chapters whilst other set-ups will be tested depending on the characteristics of the presented algorithms.

An ensemble of 10^4 channel matrices \mathbf{H}_i , for $i = 1, 2$, have been simulated with Rayleigh distribution, being \mathbf{H}_1 and \mathbf{H}_2 channel matrices for hops 1 and 2, respectively. Simulations with QPSK modulation have been conducted, where $N_B = 100$ symbols are sent. Finally, the SNRs at the first (ρ_1) and second hops (ρ_2) have been considered independent.

Four different schemes have been evaluated: linear zero-forcing approach (*Lin-ZF*), linear MMSE solution (*Lin-MMSE*), *Lin-ZF-EQ* and *Lin-MMSE-EQ*, from which the last two are the linear approaches with equalizers at the receivers.

3.2.4.1 BER Performance

Figure 3.2 shows the bit error rate performance for a fixed ρ_2 and ρ_1 , depicted in Figures 3.2(a) and 3.2(b). Two values have been considered for the fixed signal-to-noise ratio: 15 and 20 dB.

Some conclusions hold for both figures. The worst BER performance is presented for the designs based on ZF criterion, named *Lin-ZF* and *Lin-ZF-EQ*, being outperformed by Wiener solutions. Moreover, provided BER simulations show that zero-forcing design with equalizer matches up with the framework without it. This means that ZF approaches

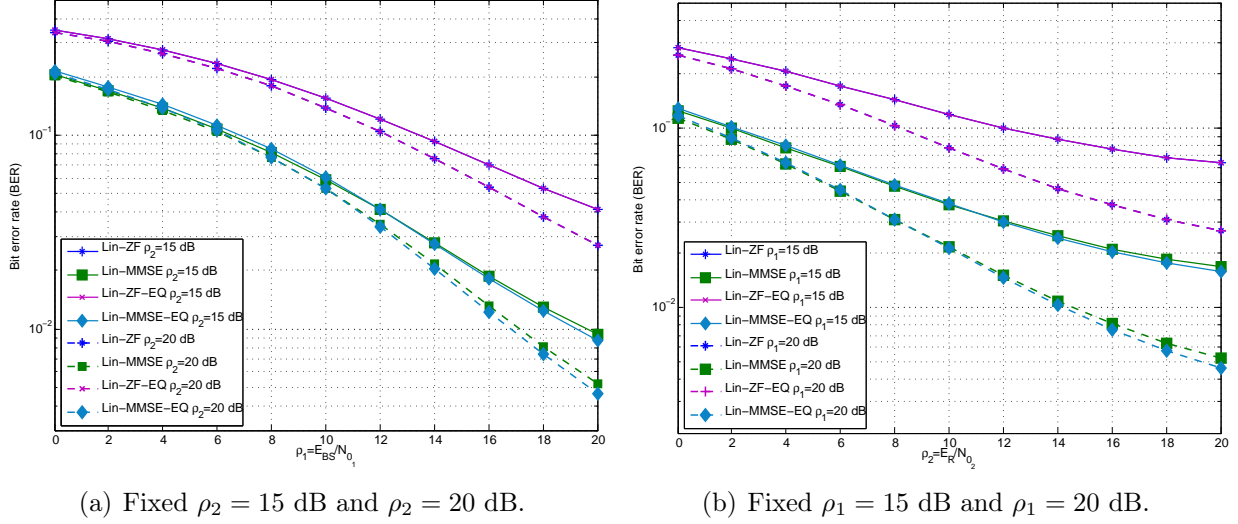


Figure 3.2: BER performance of the ZF and MMSE linear precoding techniques with and without equalizer in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation.

converge in the first iteration, obtaining an identity matrix as equalizer. Furthermore, as expected *Lin-MMSE-EQ* approach outperforms the rest of the systems.

When ρ_2 is considered fixed, as depicted in Figure 3.2(a), we observe that the gain that *Lin-MMSE-EQ* achieves respect to *Lin-MMSE* is negligible when $\rho_2 = 15$ dB. Nevertheless, this difference gap increases when ρ_2 is set to 20 dB. Furthermore, the maximum gain achieved by MMSE set-ups respect to ZF approaches is about 8 dB when $\rho_2=15$ dB and 7 dB when $\rho_2 = 20$ dB.

Assuming ρ_1 fixed (Figure 3.2(b)), we observe that the difference between *Lin-MMSE* and *Lin-MMSE-EQ* continues negligible when $\rho_1 = 15$ dB, whilst for $\rho_1 = 20$ dB the difference is greater, being the maximum difference of 1 dB. Furthermore, the gain achieved by MMSE approaches respect to ZF reaches 14 dB and 11 dB when $\rho_1 = 15$ dB and $\rho_1 = 20$ dB, respectively. Therefore, increasing the fixed SNR the difference between ZF and MMSE approaches is lower. Apart from this, comparing both figures we show that the first hop SNR (ρ_1) has more influence on the performance of the system.

Figure 3.3 shows the performance for the linear set-up $6 \times 6 \times \{6 \times 1\}$, being ρ_1 set to 15 and 20 dB in Figure 3.3(b), whilst ρ_2 has been fixed to the same values in Figure 3.3(a). The same conclusions are derived for this set-up. It should be underlined that the best performance is achieved by *Lin-MMSE-EQ* approach, being the gain respect to *Lin-MMSE* little when the SNR is fixed to 15 dB and more evident when this changes to 20 dB. Furthermore, the performance of the linear ZF approaches continues being the same.

As depicted in Figure 3.3(a), when $\rho_2 = 15$ dB, the maximum gain achieved by MMSE approaches over ZF frameworks is about 10 dB, while it decreases to 8 dB when $\rho_2=20$ dB. In the same fashion, when ρ_1 is set to 15 dB the gain is greater, reaching 16 dB when $\rho_1 = 20$

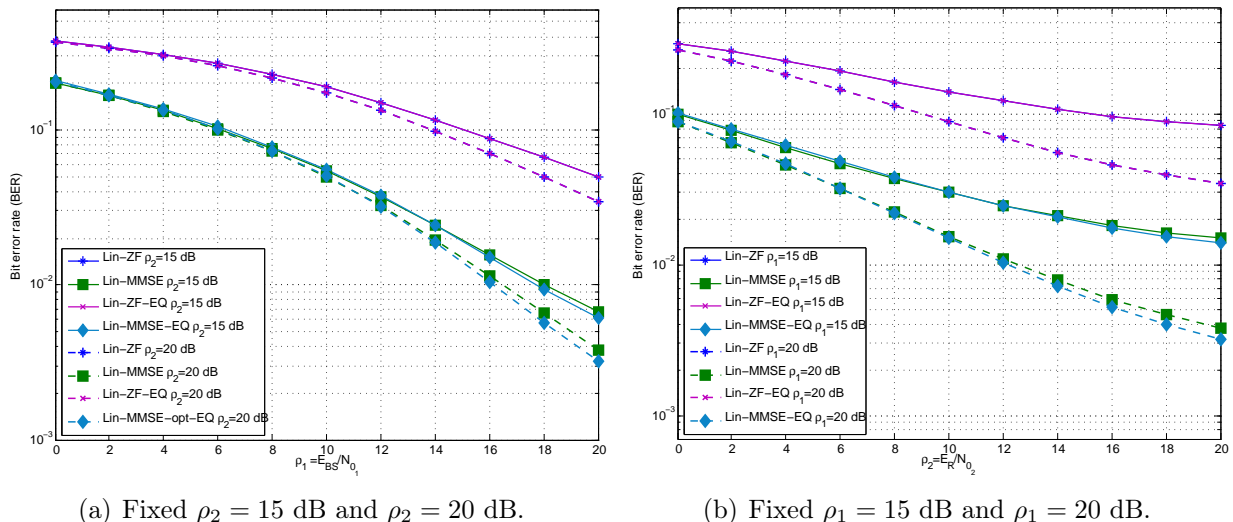


Figure 3.3: BER performance of the ZF and MMSE linear precoding techniques with and without equalizer in the $6 \times 6 \times \{6 \times 1\}$ set-up (B) with QPSK modulation.

dB.

Finally Figure 3.4 depicts the performance for $6 \times 6 \times \{2 \times 3\}$ set-up when ρ_1 and ρ_2 are fixed (Figures 3.4(b) and 3.4(a)). Here, the same conclusions are obtained. Furthermore, we observe the effectiveness of the equalization process, specially interesting when $N_k > 1$.

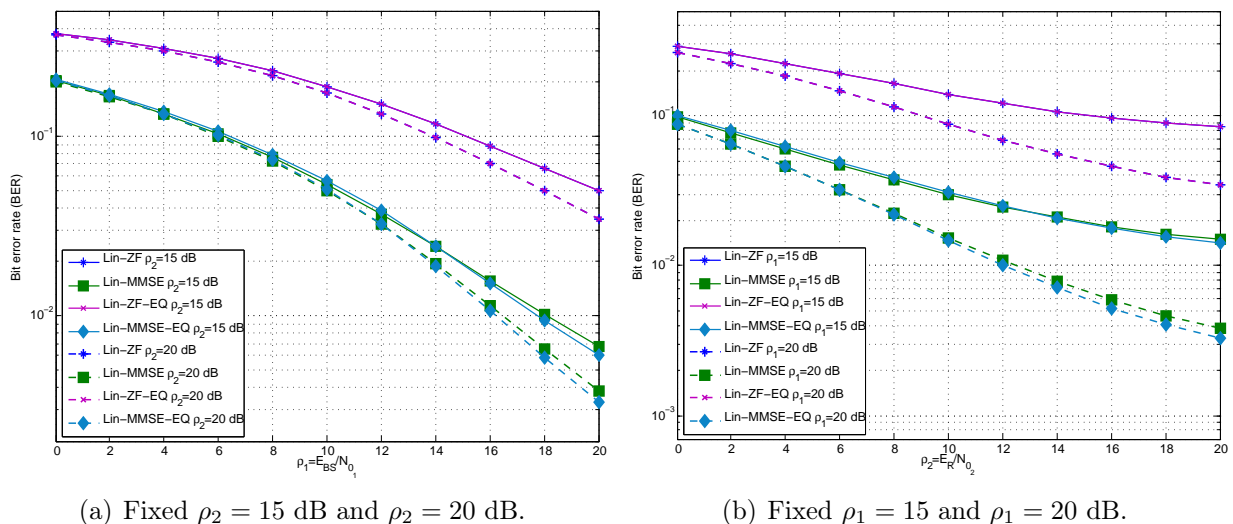


Figure 3.4: BER performance of the ZF and MMSE linear precoding techniques with and without equalizer in the $6 \times 6 \times \{2 \times 3\}$ set-up (C) with QPSK modulation.

To sum-up, we observe that the performance of the systems will be limited by ρ_1 . Moreover, increasing the fixed SNR the performance is better. Hereinafter, due to the results obtained, only the set-up $4 \times 4 \times \{4 \times 1\}$ will be analysed for fixed $\rho_1 = 15$ dB and $\rho_2 = 15$ dB. Depending on the considered algorithms, other set-ups will be also considered.

3.2.4.2 Achievable Sum-Rate

The achievable sum-rate stands for the data rate that can be achieved individually by each user with low probability error. The achievable sum-rate, analysed deeply for single-user MIMO relaying networks in [Wang05], can be computed in the scenario here considered as:

$$\begin{aligned} C &= \frac{1}{2} \sum_{k=1}^K \log_2 (|\mathbf{I}_{L_k} + \mathbf{R}_{\mathbf{n}_k}^{-1} \mathbf{H}_k \mathbf{R}_d \mathbf{H}_k^H|), \\ &= \frac{1}{2} \sum_{k=1}^K \log_2 (|\mathbf{I}_{L_k} + \mathbf{R}_{\mathbf{n}_k}^{-1} \mathbf{D}_k \mathbf{H}_{2,k} \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}} \mathbf{F}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_{2,k}^H \mathbf{D}_k^H|), \end{aligned} \quad (3.13)$$

where $\mathbf{R}_{\mathbf{n}_k} = \beta_1^2 N_{01} \mathbf{D}_k \mathbf{H}_{2,k} \overline{\mathbf{W}} \mathbf{W}^H \mathbf{H}_{2,k}^H \mathbf{D}_k^H + \beta_2^2 N_{02} \mathbf{D}_k \mathbf{D}_k^H$ and $\overline{\mathbf{W}}$ and $\overline{\mathbf{F}}$ are the unscaled relaying and precoding matrices. The matrix channel created from the BS to the user k , while \mathbf{R}_d denotes the input signal covariance matrix, assumed $\mathbf{R}_d = \mathbf{I}_L$ for linear approaches.

In this section the achievable sum-rate is analysed for the optimal linear approaches introduced before. Figure 3.5 depicts the capacity for set-up A for a fixed $\rho_1=15$ dB and $\rho_2=15$ dB, illustrated in Figures 3.5(b) and 3.5(a), respectively.

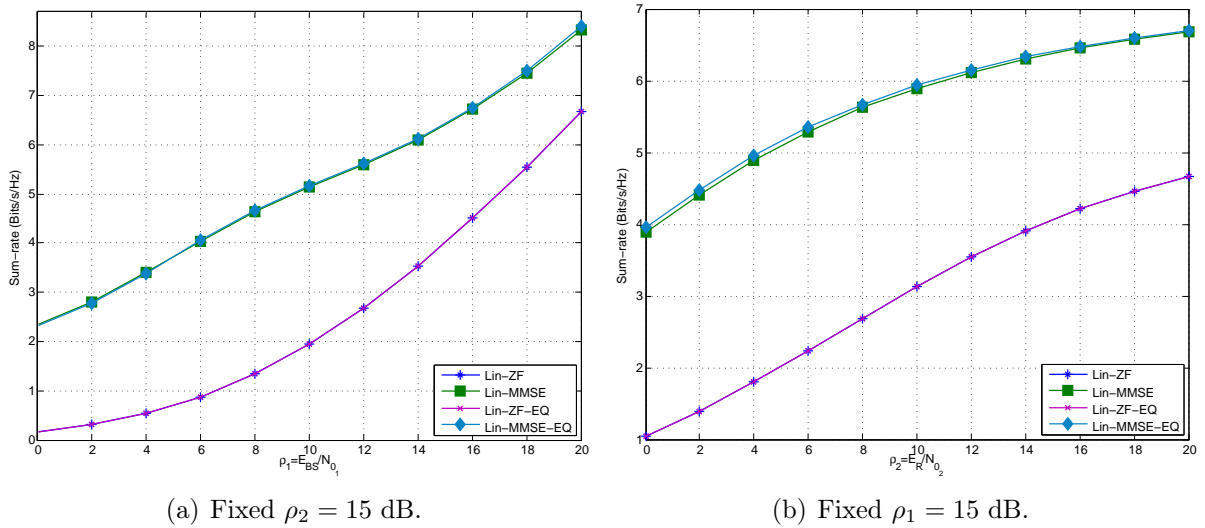


Figure 3.5: Achievable sum-rate of the ZF and MMSE linear precoding techniques with and without equalizer in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

As expected, MMSE solutions, *Lin-MMSE* and *Lin-MMSE-EQ* outperform the approaches designed under ZF criterion, which achieve the same sum-rate. Oppositely to BER performance, ρ_1 will have more influence in the performance of the system positively. Furthermore, the improvement achieved by *Lin-MMSE-EQ* is not as evident as expected. Nevertheless, we expect that increasing the number of receive antennas or increasing the fixed SNR this will be more evident.

When $\rho_2 = 15$ dB, the difference between MMSE and ZF system decreases. Hence, when

$\rho_1 = 0$ dB the difference gap is about 13 dB, whilst at 11 dB is reduced to 9 dB. In addition, when $\rho_1 = 15$ dB the difference increases, oppositely to the previously presented set-up.

3.2.4.3 Convergence

In order to measure the convergence of the proposed iterative algorithms, two different simulations have been carried out. In the first one, we compute the BER versus SNR curves for a set of iterations. In the second one, the mean square error is computed for different numbers of iterations.

Before analysing the results, it should be underlined that the convergence of ZF approaches named *Lin-ZF* and *Lin-ZF-EQ* has been omitted due to the results obtained in the previous section. The mean square error and the BER performance are constant over the different number of iterations which means that the convergence error is reached in the first iteration, breaking the iterative loop.

Figure 3.6 shows the BER versus ρ_1 curves for set-up A for both *Lin-MMSE* and *Lin-MMSE-EQ* approaches when the number of iterations are constant and the signal-to-noise ratio of the second hop is 15 dB.

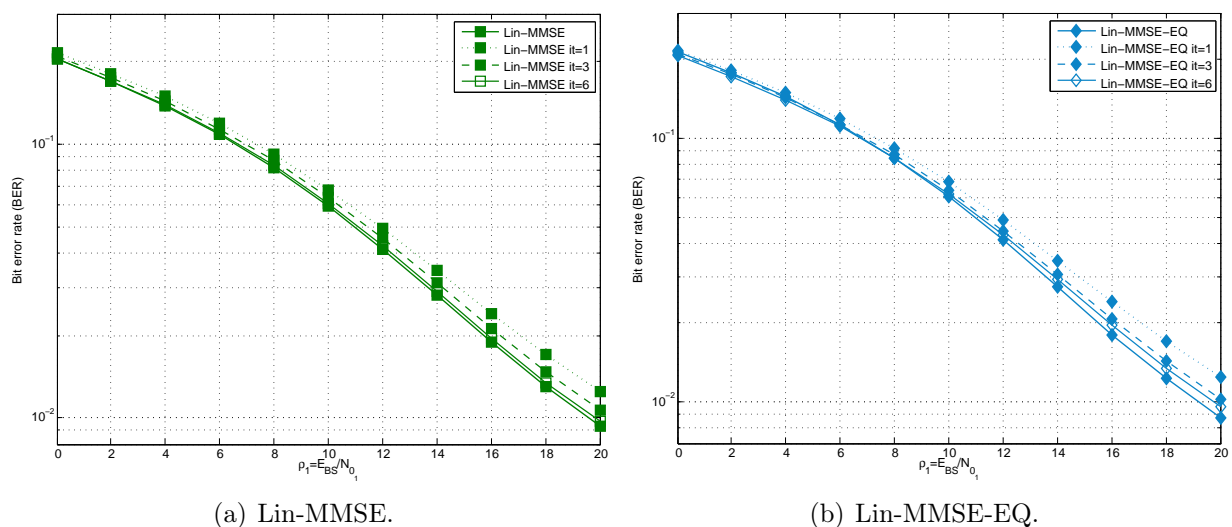


Figure 3.6: BER versus ρ_1 results for different iteration numbers of *Lin-MMSE* and *Lin-MMSE-EQ* linear approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation and $\rho_2 = 15$ dB.

Figure 3.6(a) depicts the BER convergence for *Lin-MMSE*, which stands for the joint local solution's performance. Comparing the best curve obtained with the one we get when the number of iterations is equal to one, the gain obtained by the optimal is close to 2 dB. Increasing the number of iterations, we can state that the gain is reduced to 1 dB when 3 iterations are carried out. Finally, when 6 iterations are accomplished the curve falls near the optimal. This indicates that the computational cost can be reduced limiting the number

of iterations.

The same happens for *Lin-MMSE-EQ* approach as depicted in Figure 3.6(b). However, it should be pointed out that in this case executing 6 iterations the system does not perform near the optimal being advisable the realization of a greater number of iterations.

Finally Figure 3.7 shows the MSE convergence of the linear approaches for the set-up A for several values of fixed ρ_1 and ρ_2 . When $\rho_1 < \rho_2$, as shown in Figure 3.7(a), the mean square error is over one. The effectiveness of *Lin-MMSE-EQ* over *Lin-MMSE* can be seen in both figures, being the error lower. Watching carefully the figure, we can state that the convergence is achieved in a few number of iterations. For $\rho_1 > \rho_2$, depicted in Figure 3.7(b), the MSE is lower than the unit and the error difference is greater comparing it with the one obtained when $\rho_1 < \rho_2$. In this case, the convergence is also reached in ten iterations.

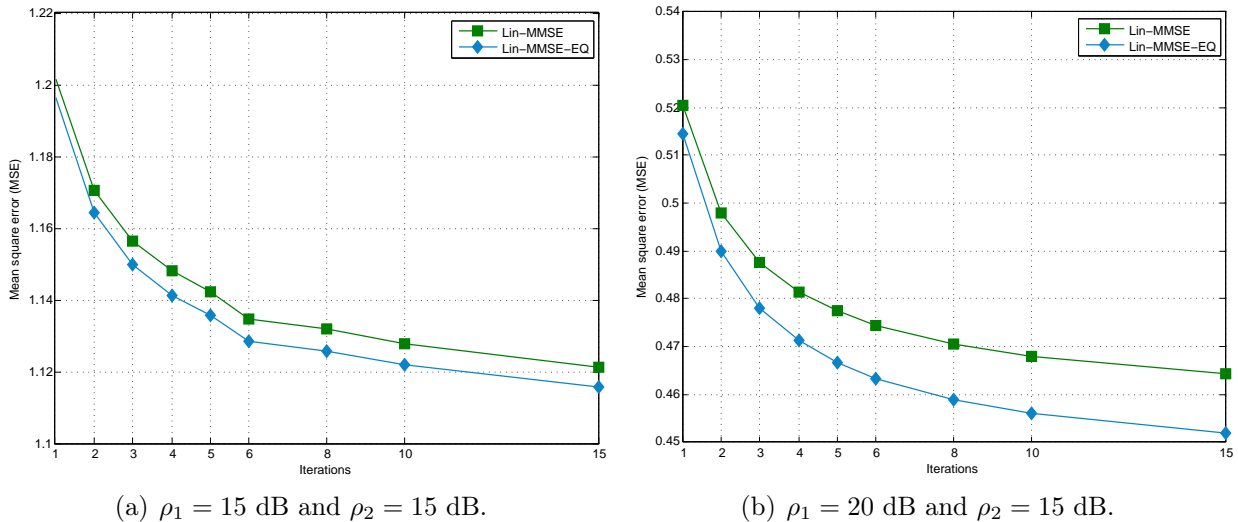


Figure 3.7: MSE convergence of *Lin-MMSE* and *Lin-MMSE-EQ* linear approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

3.3 Block Diagonalization Based Designs

In multiuser MIMO schemes, where multiple users share the channel, the performance of the system decreases due to the interference created by other users and the noise. Furthermore, as shown at the second chapter, for multiuser MIMO Gaussian channels, the sum-rate, defined as the addition of the achievable rates of the users, can only be achieved using DPC [Costa83]. The main drawback of dirty paper coding is the complexity that presents, which makes the implementation in real-time systems impossible. In order to reduce the cost, block diagonalization [Spencer04b][Shen07][Shim08] have been proposed as a precoding method which can be exploited for multiuser interference cancellation. This approach processes the transmitted symbol vector by a precoding matrix that assures no interference, dividing the

multiuser channel into multiple parallel point-to-point subchannels.

Basically, BD can be viewed as a generalization of channel inversion for the maximization of the throughput and for power allocation problems. In [Spencer04b] block diagonalization is analysed for multiuser MIMO systems, where $M > \max\{\text{rank}(\tilde{\mathbf{H}}_1), \dots, \text{rank}(\tilde{\mathbf{H}}_K)\}$ is derived as the basic condition for BD application, being M the total number of transmit antennas and $\tilde{\mathbf{H}}_k$ the equivalent channel matrix containing the channel matrices of the interfering users. This research concludes that the performance of the aforementioned system comes close to DPC's when user selection algorithms are implemented. Equally important is [Shen07], where two conclusions are obtained for MU-MIMO systems after BD application. On one hand, if the channels are orthogonal, BD reaches DPC's capacity, whereas, on the other hand, if the user channels lie in a common row vector space, the gain of DPC over BD can be bounded by the minimum of the number of transmit and receive antennas as well as the number of users. Alternatively, BD can be used for other-cell interference (OCI) cancellation in multiuser MIMO frameworks as described in [Shim08] which applies a whitening filter at each receiver for the noise reduction.

In spite of its computational complexity, BD has been also applied to multiuser relaying systems. For example, in [Hwang09] BD is employed for multiuser MIMO AF relaying, being the relaying matrix designed for sum-rate maximization. Furthermore, the optimum power allocation is carried out through the SVD decomposition under the MMSE criterion. The system is based on source-destination pairs and no precoding process is performed at each BS. The peculiarity of this research work resides on the gradient descent algorithm application [Boyd04] for the design of the relaying matrix.

3.3.1 Block Diagonalization System Model

Taking into account the good results for multiuser schemes, BD is applied in this section to our multiuser MIMO relaying system, which has been depicted in Figure 3.1. It is known that after re-scaling the signal the terminal k receives:

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_k \mathbf{s}_k + \underbrace{\sum_{\substack{j=1 \\ j \neq k}}^K \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_j \mathbf{s}_j}_{\text{Interference}} + \underbrace{\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}}_{\text{Noise}}.$$

For BD application, the constraint $\mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_j = \mathbf{0}_{N_k \times L_k}$ has to be applied for interference cancellation, mapping the precoding matrix of each user in the null space of the equivalent channel created between the BS and the end users. We assume that the precoding matrix can be decomposed as: $\mathbf{F} = \mathbf{F}_2 \mathbf{F}_1 \in \mathbb{C}^{M \times L}$, being $\mathbf{F}_2 = [\mathbf{F}_{2,1} \dots \mathbf{F}_{2,K}] \in \mathbb{C}^{M \times L}$ and $\mathbf{F}_1 = \text{blkdiag}(\mathbf{F}_{1,1}, \dots, \mathbf{F}_{1,K}) \in \mathbb{C}^{L \times L}$ the interference cancellation matrix and the transmit processing matrix, respectively. By means of the former, the equivalent channel created

between the BS and the end users is diagonalized, creating multiple parallel subchannels, whereas the latter corresponds to the precoding matrix for signal processing and MUI cancellation. After the application of the defined precoding matrix, the received signal at user terminal k is reduced to

$$\begin{aligned} \hat{\mathbf{d}}_k &= \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_{2,k} \mathbf{F}_{1,k} \mathbf{s}_k + \underbrace{\sum_{\substack{j=1 \\ j \neq k}}^K \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_{2,j} \mathbf{F}_{1,j} \mathbf{s}_j}_{\text{Interference}} \\ &+ \underbrace{\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}}_{\text{Noise}}, \end{aligned} \quad (3.14)$$

being $\mathbf{F}_k = \mathbf{F}_{2,k} \mathbf{F}_{1,k} \in \mathbb{C}^{M \times L_k}$ the precoding matrix related to the k^{th} user, which can be stacked into the matrix \mathbf{F} previously defined. In order to cancel the interference the constraint $\mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_{2,j} = \mathbf{0}_{N_k \times L_k}$ is set for $i = 1, \dots, j-1, j+1, \dots, K$. Unsurprisingly, the matrix $\mathbf{F}_{2,j} \in \mathbb{C}^{M \times L_k}$ has to lie in the null space of $\tilde{\mathbf{H}}_j = [\tilde{\mathbf{H}}_1^T, \dots, \tilde{\mathbf{H}}_{j-1}^T, \tilde{\mathbf{H}}_{j+1}^T, \tilde{\mathbf{H}}_K^T]^T \in \mathbb{C}^{N-N_k \times M}$, being $\tilde{\mathbf{H}}_{j-1} = \mathbf{H}_{2,j-1} \mathbf{W} \mathbf{H}_1$. This equivalent channel excludes the user j from the equivalent channel matrix, thus it can be renamed as interference channel matrix. If SVD decomposition is applied to $\tilde{\mathbf{H}}_j$, we get

$$\tilde{\mathbf{H}}_j = \tilde{\mathbf{U}}_j \tilde{\mathbf{\Lambda}}_j \begin{bmatrix} \tilde{\mathbf{V}}_j^{(1)} & \tilde{\mathbf{V}}_j^{(0)} \end{bmatrix}^H,$$

where $\tilde{\mathbf{U}}_j \in \mathbb{C}^{N-N_k \times N-N_k}$ and $\tilde{\mathbf{\Lambda}}_j \in \mathbb{R}^{N-N_k \times M}$ denote the right singular vectors and the singular values ordered in descending order, respectively. In addition, $\tilde{\mathbf{V}}_j^{(1)} \in \mathbb{C}^{M \times M-N_j}$ stands for the right singular vectors corresponding to the non-zero singular values, while $\tilde{\mathbf{V}}_j^{(0)} \in \mathbb{C}^{M \times N_j}$ contains the right singular vectors corresponding to the null singular values. The latter forms an orthogonal basis from the null space of $\tilde{\mathbf{H}}_j$, thus, $\mathbf{F}_{2,j} = \tilde{\mathbf{V}}_j^{(0)}$ fulfils the constraint $\mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_{2,j} = \mathbf{0}_{N_k \times L_k}$, considering $L_k = N_k$. Once the block diagonalization constraint is applied to (3.14), the signal received at the k^{th} MS is reduced to

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{F}_{1,k} \mathbf{s}_k + \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{n}_1 + \beta_2 \beta_1 \mathbf{D}_k \mathbf{n}_{2,k},$$

where $\mathbf{H}_{e,k} = \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_{2,k} \in \mathbb{C}^{N_k \times L_k}$ is the equivalent channel created from the BS to the end users. After the application of BD, multiple parallel subchannels are created between the transmitter and the receivers, each one belonging to one user.

3.3.2 Design of Precoding and Relaying Matrices

In the same way, as it has been done for linear approaches, the relaying and precoding matrices are jointly optimized in order to minimize the mean square error, defined as

$$\begin{aligned} \xi &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{s}_k\|_2^2 \right], \\ &= \sum_{k=1}^K \text{Tr}(\beta_2^2 \beta_1^2 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_{2,k} \mathbf{F}_{1,k} \mathbf{F}_{1,k}^H \mathbf{F}_{2,k}^H \mathbf{H}_1^H \mathbf{W} \mathbf{H}_{2,k}^H \mathbf{D}_k^H + \beta_2^2 \beta_1^2 N_{0_1} \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{W}^H \\ &\quad \mathbf{H}_{2,k}^H \mathbf{D}_k^H + \beta_2^2 N_{0_2} \mathbf{D}_k \mathbf{D}_k^H - 2\beta_2 \beta_1 \Re(\mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{F}_{2,k} \mathbf{F}_{1,k}) + \mathbf{I}_{L_k}), \end{aligned} \quad (3.15)$$

where, $\mathbf{R}_s = \mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_L$ and N_{0_1} and N_{0_2} denote the AWGN noise spectral densities of the first and second hops, respectively. When no equalization is considered, $\mathbf{D}_k = \mathbf{I}_{L_k \times N_k}$ is assumed. Furthermore, two power constraints are established in order to limit the power at the base station and the relay:

$$\sum_{k=1}^K \text{Tr}(\mathbf{F}_{2,k} \mathbf{F}_{1,k} \mathbf{F}_{1,k}^H \mathbf{F}_{2,k}^H) = \sum_{k=1}^K \text{Tr}(\mathbf{F}_{1,k} \mathbf{F}_{1,k}^H) = E_{BS}, \quad \text{and} \quad (3.16a)$$

$$\beta_1^2 \sum_{k=1}^K \text{Tr}(\mathbf{W}(\mathbf{H}_1 \mathbf{F}_{2,k} \mathbf{F}_{1,k} \mathbf{F}_{1,k}^H \mathbf{F}_{2,k}^H \mathbf{H}_1^H + N_{0_1} \mathbf{I}_R) \mathbf{W}^H) = E_R. \quad (3.16b)$$

The expression (3.16a) can be reduced to the second term of the equation due to the orthogonality of the matrix $\mathbf{F}_{2,k}$, i.e., $\mathbf{F}_{2,k}^H \mathbf{F}_{2,k} = \mathbf{I}_{L_k}$.

If ZF is applied for the joint design of the matrices, the constraint $\beta_2 \beta_1 \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F}_2 \mathbf{F}_1 = \mathbf{I}_{N \times L}$ has to be included. After applying this to the MSE defined in (3.15), we get

$$\xi_{ZF} = \sum_{k=1}^K \text{Tr}(\beta_2^2 \beta_1^2 N_{0_1} \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{W}^H \mathbf{H}_{2,k}^H \mathbf{D}_k^H + \beta_2^2 N_{0_2} \mathbf{D}_k \mathbf{D}_k^H). \quad (3.17)$$

The optimization problem for the design under the ZF criterion is outlined as

$$\begin{aligned} \{\mathbf{F}_1, \mathbf{W}, \beta_1, \beta_2\} &= \underset{\{\mathbf{F}_1, \mathbf{W}, \beta_1, \beta_2\}}{\text{argmin}} \quad \xi_{ZF} \\ \text{s.t.} \quad &\sum_{k=1}^K \text{Tr}(\mathbf{F}_{1,k} \mathbf{F}_{1,k}^H) = E_{BS}, \\ &\beta_1^2 \sum_{k=1}^K \text{Tr}(\mathbf{W}(\mathbf{H}_1 \mathbf{F}_{2,k} \mathbf{F}_{1,k} \mathbf{F}_{1,k}^H \mathbf{F}_{2,k}^H \mathbf{H}_1^H + N_{0_1} \mathbf{I}_R) \mathbf{W}^H) = E_R, \\ &\beta_2 \beta_1 \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} = \mathbf{0}_L, \end{aligned}$$

where ξ_{ZF} is defined in (3.17). Once the optimization problem is solved, as shown in Section

B.1 of Appendix B, the following expressions are obtained

$$\mathbf{W} = \frac{1}{\beta_2} (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2)^{-1} \mathbf{H}_2^H \mathbf{D}^H \mathbf{F}_2^H \bar{\mathbf{F}}_1^H \mathbf{H}_1^H \left(\mathbf{H}_1 \mathbf{F}_2 \bar{\mathbf{F}}_1 \bar{\mathbf{F}}_1^H \mathbf{F}_2^H \mathbf{H}_1^H \right)^{-1}, \quad \text{and}$$

$$\mathbf{F}_{1,k} = \frac{1}{\beta_1} (\mathbf{H}_{e,k}^H \mathbf{D}_k^H \mathbf{D}_k \mathbf{H}_{e,k})^{-1} \mathbf{H}_{e,k}^H \mathbf{D}_k^H,$$

being $\mathbf{H}_{e,k} = \mathbf{H}_{2,k} \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{F}_{2,k}$ the equivalent channel created between the BS and the k^{th} user and $\bar{\mathbf{W}} = \beta_2 \mathbf{W}$ and $\bar{\mathbf{F}}_1 = \beta_1 \mathbf{F}_1$ are the equivalent unscaled versions of matrices \mathbf{W} and \mathbf{F}_1 , respectively.

Otherwise, if MMSE is chosen as design criterion, the optimization problem is set as

$$\{\mathbf{F}_1, \mathbf{W}, \beta_1, \beta_2\} = \underset{\{\mathbf{F}_1, \mathbf{W}, \beta_1, \beta_2\}}{\text{argmin}} \quad \xi$$

s.t. $\text{Tr}(\mathbf{y}_{BS} \mathbf{y}_{BS}^H) = E_{BS},$
 $\text{Tr}(\mathbf{y}_R \mathbf{y}_R^H) = E_R,$

being ξ the mean square error defined in (3.15). The expressions obtained for the precoding and relaying matrices become:

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \frac{N_{02} \text{Tr}(\mathbf{D} \mathbf{D}^H)}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H \mathbf{D}^H \mathbf{F}_2^H \bar{\mathbf{F}}_1^H \mathbf{H}_1^H \mathbf{A}^{-1}, \quad \text{and}$$

$$\mathbf{F}_{1,k} = \frac{1}{\beta_1} \left(\bar{\mathbf{H}}_{e,k}^H \mathbf{D}_k^H \mathbf{D}_k \bar{\mathbf{H}}_{e,k} + \frac{N_{02} \text{Tr}(\mathbf{D} \mathbf{D}^H)}{E_R} \mathbf{F}_{2,k}^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{F}_{2,k} + \mathbf{B} \right)^{-1} \bar{\mathbf{H}}_{e,k}^H \mathbf{D}_k^H,$$

where $\mathbf{A} = \left(\mathbf{H}_1 \mathbf{F}_2 \bar{\mathbf{F}}_1 \bar{\mathbf{F}}_1^H \mathbf{F}_2^H \mathbf{H}_1^H + N_{01} \frac{\text{Tr}(\bar{\mathbf{F}}_1 \bar{\mathbf{F}}_1^H)}{E_{BS}} \mathbf{I}_R \right)$ and $\mathbf{F}_{1,k}$ denotes the precoding matrix related to the k^{th} user for MUI cancellation. In the same way, $\mathbf{B} = \frac{N_{01}}{E_{BS}} \text{Tr} \left(\bar{\mathbf{W}}^H (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \frac{N_{02} \text{Tr}(\mathbf{D} \mathbf{D}^H)}{E_R} \mathbf{I}_R) \bar{\mathbf{W}} \right) \mathbf{I}_M$. As ρ_1 and ρ_2 increase, the Wiener expression comes close to the expression obtained through the zero forcing criterion.

As a result of the interdependence of the matrices, an iterative approach has to be applied. However, in opposition to the iterative algorithms proposed in previous section, this iterative approach does not converge. Put in other words, once $\bar{\mathbf{F}}_1$ is optimized, the convergence error increases due to the maximization of $\bar{\mathbf{W}}$ and vice versa. Since the convergence is not achieved, the application of this algorithm is unfeasible.

Therefore, in order to give a local optimal solution based on BD, an alternative approach is proposed in the following section, where a linear precoding stage is applied at the BS and a linear block diagonalization process is employed at the relaying terminal, which helps with the interference cancellation.

3.3.3 Partial Block Diagonalization Linear Approach

The partial block diagonalization approach introduced in this section is depicted in Figure 3.8, where it can be observed that the transmission is divided into a point-to-point and a point-to-multipoint communication.

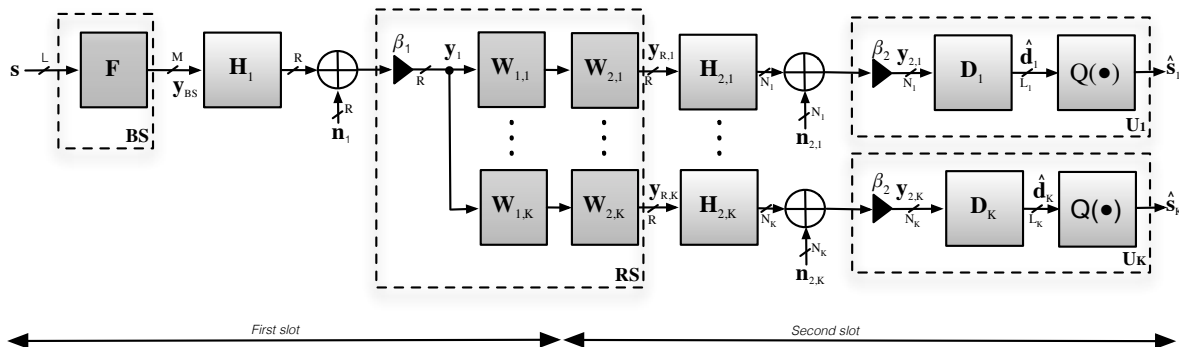


Figure 3.8: Partial block diagonalized system composed by a BS transmitting through a relay to K multi-antenna users.

The partial block diagonalization approach proposed in this section combines linear precoding and linear block diagonalization precoding. While the former is applied at the BS for users' signal separation, the latter, applied at the relay, cancels other users' interference. Since it has been underlined throughout this section, the linear precoding arouses a high interest for signal processing in multiuser environments because of its simplicity. Apart from that, [Spencer04b] proposes BD for multiuser MIMO downlink systems as an alternative to linear precoders. If the channel is known at the relay, the other users' interference can be cancelled before the transmission giving diagonalized structure to the signal.

From Figure 3.8, the precoding matrix at the relay can be divided into two submatrices, one for interference cancellation and the other for signal processing and MUI cancellation, such as

$$\mathbf{W} = \mathbf{W}_2 \mathbf{W}_1 = [\mathbf{W}_{2,1}, \dots, \mathbf{W}_{2,k}, \dots, \mathbf{W}_{2,K}] \begin{bmatrix} \mathbf{W}_{1,1} \\ \vdots \\ \mathbf{W}_{1,k} \\ \vdots \\ \mathbf{W}_{1,K} \end{bmatrix}, \quad (3.18)$$

where the matrix $\mathbf{W}_2 \in \mathbb{C}^{R \times N}$ denotes the interference suppression matrix that diagonalizes the second hop channel, being $\mathbf{W}_{2,k} \in \mathbb{C}^{R \times N_k}$ the matrix related to the user k . Moreover, $\mathbf{W}_1 \in \mathbb{C}^{N \times R}$ stands for the signal processing matrix at the relay containing K matrices $\mathbf{W}_{1,k} \in \mathbb{C}^{N_k \times R}$, each one related to one user. After the application of (3.18) in (3.14) the

signal received at user k can be expressed as:

$$\begin{aligned} \hat{\mathbf{d}}_k &= \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F} \mathbf{s} + \underbrace{\sum_{\substack{j=1 \\ j \neq k}}^K \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,j} \mathbf{W}_{2,j} \mathbf{W}_{1,j} (\mathbf{H}_1 \mathbf{F} \mathbf{s} + \mathbf{n}_1)}_{\text{Interference}} \\ &+ \underbrace{\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}}_{\text{Noise}}, \end{aligned}$$

being β_1 and β_2 the scaling factors employed for the power constraints at the first and second hops, respectively. For interference suppression at the second hop, it is necessary the application of the constraint $\mathbf{H}_{2,k} \mathbf{W}_{2,j} = \mathbf{0}_{N_k}$ for all $j \neq k$ and $1 \leq j, k \leq K$, in order to make the channel block-diagonal.

As well as the BD approach introduced in the Section 3.3, two steps are carried out for the diagonalization. Firstly, the interfering channel matrix is defined as $\tilde{\mathbf{H}}_j = [\mathbf{H}_{2,1}^T, \dots, \mathbf{H}_{2,j-1}^T, \mathbf{H}_{2,j+1}^T, \dots, \mathbf{H}_{2,K}^T]^T \in \mathbb{C}^{N-N_k \times R}$ for each user. Additionally, the singular value decomposition of $\tilde{\mathbf{H}}_j$ is accomplished. Before continuing, it should be pointed out that the minimum condition for block diagonalization at the second hop is $R \geq \{\text{rank}(\tilde{\mathbf{H}}_1), \dots, \text{rank}(\tilde{\mathbf{H}}_K)\}$. After SVD decomposition, $\tilde{\mathbf{V}}_j^{(0)} \in \mathbb{C}^{R \times N_j}$ denotes the right zero singular vectors that forms an orthogonal basis from the null space of $\tilde{\mathbf{H}}_j$, becoming $\mathbf{W}_{2,j} = \tilde{\mathbf{V}}_j^{(0)}$. Once BD constraint is applied, the received signal at each user is represented as:

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k} \mathbf{F} \mathbf{s} + \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}. \quad (3.19)$$

Analysing (3.19), the interference from the BS to the end users is not suppressed completely due to the application of block diagonalization at the second hop. However, with the help of the precoding stage at the base station the impact of the interferences created by other users is reduced. After the aforementioned block diagonalization at the second hop, the channel is divided into parallel blocks, each one related to one user.

$$\mathbf{H}_e = \begin{bmatrix} \mathbf{H}_{2,1} \mathbf{W}_{2,1} & \cdots & \mathbf{H}_{2,1} \mathbf{W}_{2,k} & \cdots & \mathbf{H}_{2,1} \mathbf{W}_{2,K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{2,k} \mathbf{W}_{2,1} & \cdots & \mathbf{H}_{2,k} \mathbf{W}_{2,k} & \cdots & \mathbf{H}_{2,k} \mathbf{W}_{2,K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{H}_{2,K} \mathbf{W}_{2,1} & \cdots & \mathbf{H}_{2,K} \mathbf{W}_{2,k} & \cdots & \mathbf{H}_{2,K} \mathbf{W}_{2,K} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{e,1} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{H}_{e,k} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{H}_{e,K} \end{bmatrix},$$

being $\mathbf{H}_e = \mathbf{H}_2 \mathbf{W}_2 \in \mathbb{C}^{N \times N}$ the equivalent block-diagonalized channel.

The main design strategy of this approach is the joint optimization of the precoding matrix \mathbf{F} and the relaying matrix \mathbf{W}_1 for the minimization of the mean square error. For

the optimization under both ZF and MMSE criterion, the MSE is defined as:

$$\begin{aligned}
 \xi &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{s}_k\|_2^2 \right], \\
 &= \sum_{k=1}^K \text{Tr} \left(\beta_2^2 \beta_1^2 \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{D}_k^H + \beta_2^2 \beta_1^2 N_{01} \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{D}_k^H \right. \\
 &\quad \left. + \beta_2^2 N_{02} \mathbf{D}_k \mathbf{D}_k^H - 2\Re(\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F}_k) + \mathbf{I}_{L_k} \right). \tag{3.20}
 \end{aligned}$$

The design is going to be subject to two power constraints, one to restrict the power at the BS to $E_{BS} = \text{Tr}(\mathbf{F}\mathbf{F}^H)$ and the other at the relay to make the power equal to $E_R = \sum_{k=1}^K \beta_1^2 \text{Tr}(\mathbf{W}_{1,k}(\mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H + N_{01} \mathbf{I}_R) \mathbf{W}_{1,k}^H)$, where $\mathbf{W}_{2,k}^H \mathbf{W}_{2,k} = \mathbf{I}_{N_k}$ has been applied, considering in this way the orthogonality of the matrix.

In the following points, the partial block diagonalization approach will be analysed for multiuser MIMO AF downlink relaying systems under zero forcing and MMSE design criteria.

3.3.3.1 Zero-Forcing Partial BD

For ZF approach the $\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_e \mathbf{W}_1 \mathbf{H}_1 \mathbf{F}_k = \mathbf{I}_{L_k}$ constraint is set. Accomplishing the procedure depicted in Section B.1 of Appendix B, we get

$$\mathbf{W}_{1,k} = \frac{1}{\beta_2} (\mathbf{H}_{e,k}^H \mathbf{D}_k^H \mathbf{D}_k \mathbf{H}_{e,k})^{-1} \mathbf{H}_{e,k}^H \mathbf{D}_k^H \bar{\mathbf{F}}_k^H \mathbf{H}_1^H (\mathbf{H}_1 \bar{\mathbf{F}} \bar{\mathbf{F}}^H \mathbf{H}_1^H)^{-1}, \tag{3.21}$$

where $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$ stands for the unscaled precoding matrix and

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \bar{\mathbf{W}}_1^H \mathbf{H}_e^H \mathbf{D}^H \mathbf{D} \mathbf{H}_e \bar{\mathbf{W}}_1 \mathbf{H}_1 \right)^{-1} \mathbf{H}_1^H \bar{\mathbf{W}}_1^H \mathbf{H}_e^H \mathbf{D}^H, \tag{3.22}$$

being $\bar{\mathbf{W}}_1 = \beta_2 \mathbf{W}_1$ the unscaled \mathbf{W}_1 . As a consequence of the dependence of (3.22) and (3.21) an iterative process has to be applied until a convergence criterion is met.

3.3.3.2 Minimum Mean Square Error Partial BD

The MMSE optimization problem, subject to the already mentioned two power constraints is defined as

$$\begin{aligned}
 \{\mathbf{F}, \mathbf{W}_1, \beta_1, \beta_2\} &= \underset{\{\mathbf{F}, \mathbf{W}_1, \beta_1, \beta_2\}}{\text{argmin}} \quad \xi \\
 \text{s.t.} \quad &\text{Tr}(\mathbf{y}_{BS} \mathbf{y}_{BS}^H) = E_{BS}, \\
 &\text{Tr}(\mathbf{y}_R \mathbf{y}_R^H) = E_R,
 \end{aligned}$$

where \mathbf{y}_{BS} and \mathbf{y}_R denote the transmitted signals at the BS and RS, respectively, while ξ stands for the MSE defined in (3.20). The same conclusions as for the ZF case hold for the design based on MMSE, where the following expressions are obtained for the precoding and relaying matrices:

$$\mathbf{W}_{1,k} = \frac{1}{\beta_2} \left(\mathbf{H}_{e,k}^H \mathbf{D}_k^H \mathbf{D}_k \mathbf{H}_{e,k} + \frac{N_{02} \text{Tr}(\mathbf{D}\mathbf{D}^H)}{E_R} \mathbf{I}_{N_k} \right)^{-1} \mathbf{H}_{e,k}^H \mathbf{D}_k^H \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \mathbf{A}^{-1}, \quad (3.23)$$

where $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$ represents the unscaled \mathbf{F} and $\mathbf{A} = \left(\mathbf{H}_1 \bar{\mathbf{F}} \bar{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R \right)$. In the same way, considering the unscaled relaying matrix $\bar{\mathbf{W}}_1 = \beta_2 \mathbf{W}_1$, the precoding matrix for the case of MMSE becomes:

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \mathbf{B} \mathbf{H}_1 + \frac{N_{01} \text{Tr}(\mathbf{B})}{E_{BS}} \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_e^H \mathbf{D}^H \quad (3.24)$$

being $\mathbf{B} = \left(\bar{\mathbf{W}}_1^H \left(\mathbf{H}_e^H \mathbf{D}^H \mathbf{D} \mathbf{H}_e + \frac{N_{02} \text{Tr}(\mathbf{D}\mathbf{D}^H)}{E_R} \mathbf{I}_N \right) \bar{\mathbf{W}}_1 \right)$ obtained from the Lagrangian multipliers.

Analysing (3.23) and (3.24), an iterative process is again necessary due to the interdependence of the computed matrices. As it will be shown in the next section, the simulation results confirm the convergence and effectiveness of the proposed partial diagonalization approach, specially for multi-antenna set-ups, improving the local optimum solution analyzed in Section 3.2. The iterative algorithm that computed the precoding and relaying matrices is described in Algorithm 2.

Algorithm 2 Search of \mathbf{F} and \mathbf{W}_1 under the MMSE criterion for the partial BD approach.

- 1: For channel block diagonalization, definition of the equivalent interferent channel $\tilde{\mathbf{H}}_j$ for each user
 - 2: Singular value decomposition of $\tilde{\mathbf{H}}_j \implies \mathbf{W}_2$ definition.
 - 3: Initialize the variables: $\bar{\mathbf{F}}^0 = \mathbf{I}_{M \times L}$, $\bar{\mathbf{W}}_1^0 = \mathbf{I}_{N \times R}$ and $\epsilon = \infty$.
 - 4: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 5: Calculate $\bar{\mathbf{F}}^{l+1}$ (3.24) for a fixed $\bar{\mathbf{W}}_1^l$.
 - 6: Computation of $\bar{\mathbf{W}}_1^{l+1}$ (3.23) for the updated $\bar{\mathbf{F}}^{l+1}$.
 - 7: Error calculation: $\epsilon = \|\bar{\mathbf{F}}^{l+1} - \bar{\mathbf{F}}^l\|^2 + \|\bar{\mathbf{W}}_1^{l+1} - \bar{\mathbf{W}}_1^l\|^2$.
 - 8: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 9: $\bar{\mathbf{F}}^l = \bar{\mathbf{F}}^{l+1}$.
 - 10: $\bar{\mathbf{W}}_1^l = \bar{\mathbf{W}}_1^{l+1}$.
 - 11: **end if**
 - 12: **end while**
-

3.3.3.3 Receiver Equalization

When $N_k \geq 2$ or simply to mitigate the noise effect over the received signal, each user can be equipped with an equalizer. As introduced in Section 3.2.3, the equalizer can be designed under ZF and MMSE criteria. Accomplishing the same procedure depicted in the aforementioned section, the equalization matrices designed under ZF and MMSE are, respectively:

$$\mathbf{D}_k = \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \bar{\mathbf{W}}_{1,k}^H \mathbf{H}_{e,k}^H \left(\mathbf{H}_{e,k} \bar{\mathbf{W}}_{1,k} \mathbf{H}_1 \bar{\mathbf{F}} \bar{\mathbf{F}}^H \mathbf{H}_1^H \bar{\mathbf{W}}_{1,k}^H \mathbf{H}_{e,k}^H \right)^{-1}, \quad \text{and}$$

$$\mathbf{D}_k = \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \bar{\mathbf{W}}_{1,k}^H \mathbf{H}_{e,k}^H \left(\mathbf{H}_{e,k} \bar{\mathbf{W}}_{1,k} \left(\mathbf{H}_1 \bar{\mathbf{F}} \bar{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 \mathbf{N}_{01} \mathbf{I}_R \right) \bar{\mathbf{W}}_{1,k}^H \mathbf{H}_{e,k}^H + \beta_2^2 \mathbf{N}_{02} \mathbf{I}_{N_k} \right)^{-1},$$

being $\bar{\mathbf{W}}_1 = \beta_2 \mathbf{W}_1$ and $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$.

3.3.4 Simulation Results

In this section, we show the performance of the systems based on the previously defined block diagonalization approaches, named *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD*, which correspond to the approaches based on BD without and with equalizer, respectively. In order to show the effectiveness of multiuser block diagonalization approaches, these are compared with *Lin-MMSE* and *Lin-MMSE-EQ*, which have been analysed in the previous section.

Before proceeding, it should be pointed out that the set-up parameters are the ones described in Section 3.2.4. It should be underlined that as a consequence of the results obtained for $6 \times 6 \times \{6 \times 1\}$ whose conclusions match up with the ones obtained for $4 \times 4 \times \{4 \times 1\}$ framework, the latter has only been considered.

Furthermore, to show the effectiveness of BD approaches, two multi-antenna set-ups are evaluated, named $8 \times 8 \times \{4 \times 2\}$ (D) and $8 \times 8 \times \{2 \times 4\}$ (E). Finally, due to the poor results obtained, the approaches designed under zero-forcing criterion are here omitted.

3.3.4.1 BER Performance

Figure 3.9 shows the BER performance for the set-up $4 \times 4 \times \{4 \times 1\}$, for fixed $\rho_2 = 15$ dB (Figure 3.9(a)) and $\rho_1 = 15$ dB (Figure 3.9(b)). Generally speaking, in both figures we observe that at low SNR, from 0 to 10 dB, linear approaches (*Lin-MMSE* and *Lin-MMSE-EQ*) outperform BD systems *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD*. Oppositely, at high SNR, from 10 dB on, BD systems improve linear's performance. Furthermore, after some tests, we can state that increasing the fixed signal-to-noise ratio, the difference between linear processing and block diagonalization approaches is greater. As expected, equalization based designs outperform the ones with equalizer, although the gain is negligible for $N_k = 1$ set-ups.

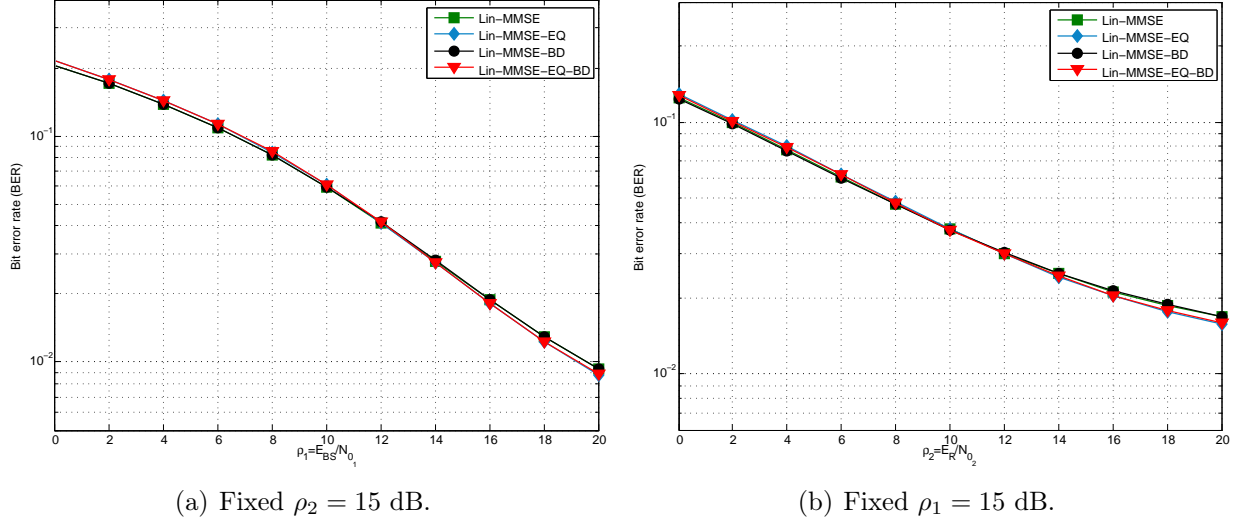


Figure 3.9: BER performance obtained by MMSE-BD approaches for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation.

Figure 3.9(a) depicts the BER performance when ρ_2 is fixed to 15 dB. The difference between *Lin-MMSE-EQ-BD* and *Lin-MMSE-BD* is negligible. In addition, the gain obtained by BD approaches over linear systems reach 0.25 dB. On the other hand, when ρ_1 is set to 15 dB, this difference increases to a dB.

In the same fashion, Figure 3.10 shows the BER performance of BD systems for set-up D when $\rho_2 = 15$ dB and $\rho_1 = 15$ dB are set, shown in Figures 3.10(a) and 3.10(b), respectively. The same conclusions are obtained, being the effectiveness of the equalizer visible at high SNR when $N_k > 1$. We can state that *Lin-MMSE-EQ-BD* outperforms the rest of the systems presented in Figure 3.10, being the maximum gain with respect to *Lin-MMSE-BD* of about 0.5 dB and 2 dB when $\rho_2 = 15$ dB and $\rho_1 = 15$, respectively.

Finally, the BER is given for set-up E when $\rho_2 = 15$ dB and $\rho_1 = 15$ dB, as shown in Figures 3.11(a) and 3.11(b), respectively. The same conclusions hold for this configuration, where we observe that the difference achieved by BD approaches over *Lin-MMSE* and *Lin-MMSE-EQ* is greater, being this about 2.5 dB and 5 dB when $\rho_2 = 15$ and $\rho_1 = 15$ dB, respectively.

Summing up, we can state that increasing N_k the difference of *Lin-MMSE-EQ-BD* respect to *Lin-MMSE* and *Lin-MMSE-BD* is greater. In fact, BD apart from being used for MUI cancellation, can be also employed for ISI cancellation. To conclude, it should be underlined that N_{01} will limit the performance of the system.

3.3.4.2 Achievable Sum-Rate

The achievable sum-rate of the system measures the maximum capacity obtained by the users defined as (3.13). Figure 3.12 depicts the performance obtained by simulations for

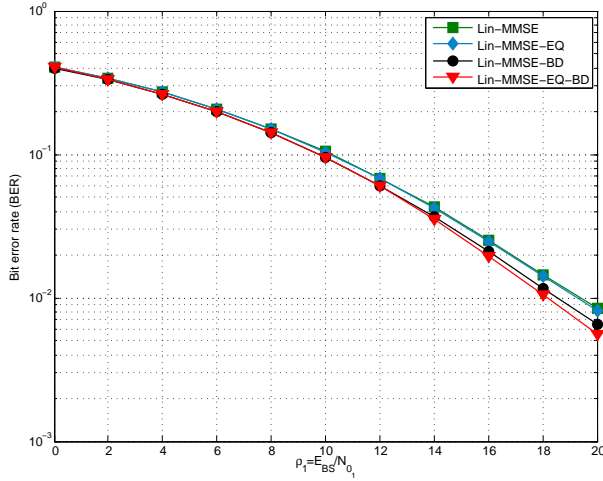
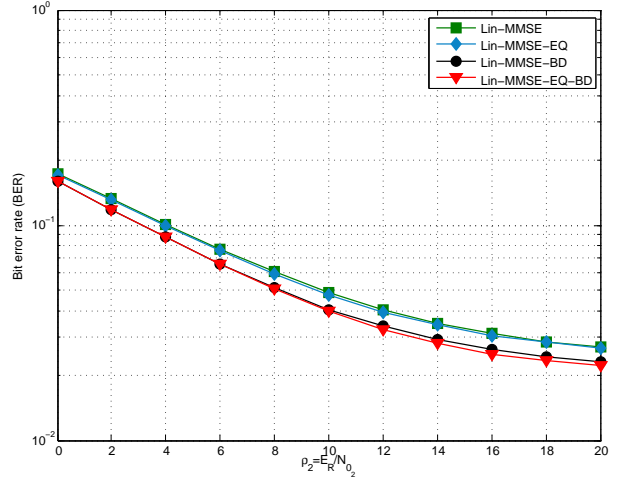

 (a) Fixed $\rho_2 = 15$ dB.

 (b) Fixed $\rho_1 = 15$ dB.

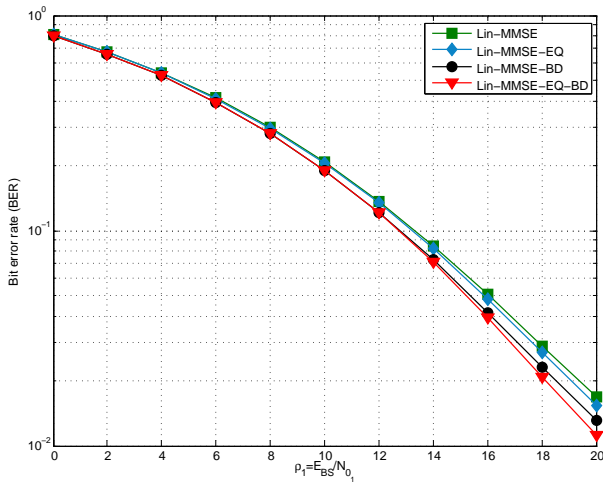
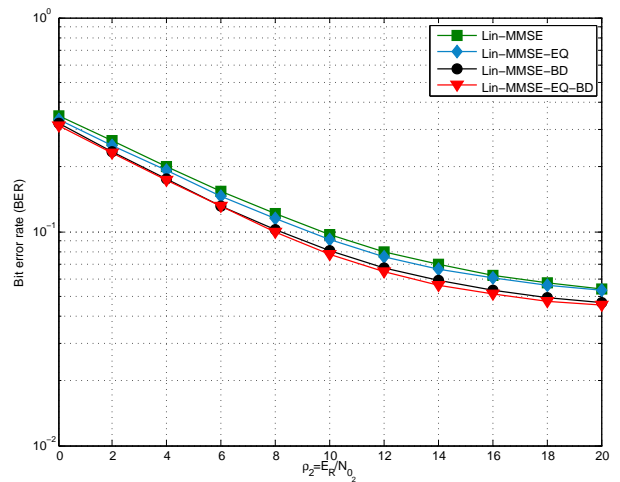
 Figure 3.10: BER performance obtained by MMSE-BD approaches for $8 \times 8 \times \{4 \times 2\}$ set-up (D) with QPSK modulation.

 (a) Fixed $\rho_2 = 15$ dB.

 (b) Fixed $\rho_1 = 15$ dB.

 Figure 3.11: BER performance obtained by MMSE-BD approaches for $8 \times 8 \times \{2 \times 4\}$ set-up (E) with QPSK modulation.

$4 \times 4 \times \{4 \times 1\}$ set-up when $\rho_2=15$ dB (Figure 3.12(a)) and $\rho_1=15$ dB (Figure 3.12(b)).

As it can be seen in Figures 3.12(a) and 3.12(b), *Lin-MMSE-EQ-BD* performs better than *Lin-MMSE-BD*, specially at high SNR. Furthermore, BD improves slightly the sum-rate of MMSE systems. Despite this improvement, the gain is not as evident as expected. Furthermore, we observe that the sum-rate achieved by the approaches with and without equalizer is not perceptible when $N_k = 1$.

When ρ_2 is set to 15 dB, which is the case considered in Figure 3.12(a), the maximum gain of BD approaches respect to MMSE linear systems is achieved at high ρ_1 , from $\rho_1 = 16$ dB to $\rho_1 = 20$ dB. Nevertheless, this difference is negligible.

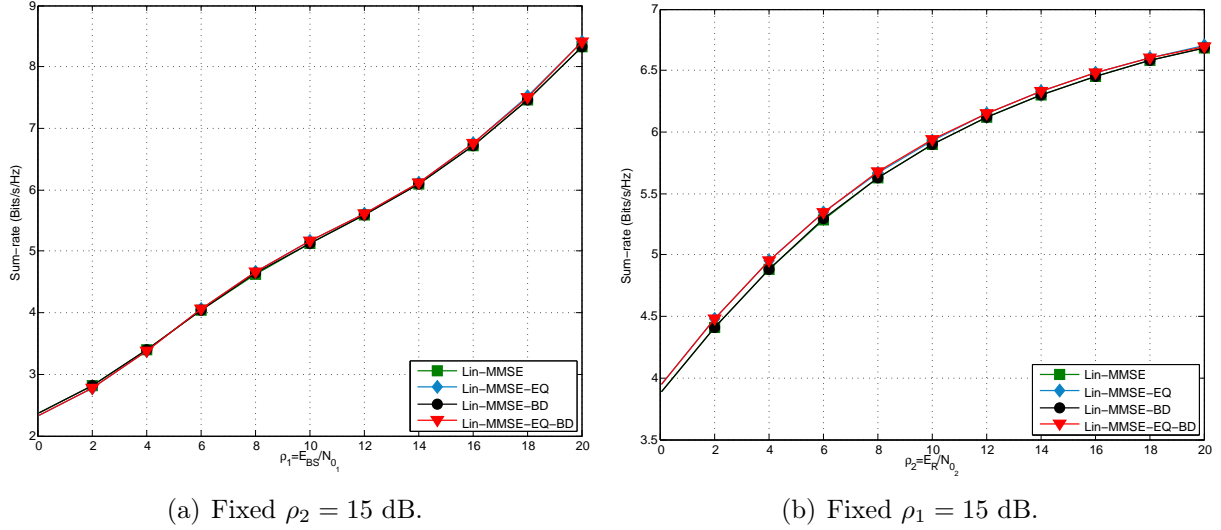


Figure 3.12: Sum-rate achieved by MMSE-BD approaches in $4 \times 4 \times \{4 \times 1\}$ set-up (A).

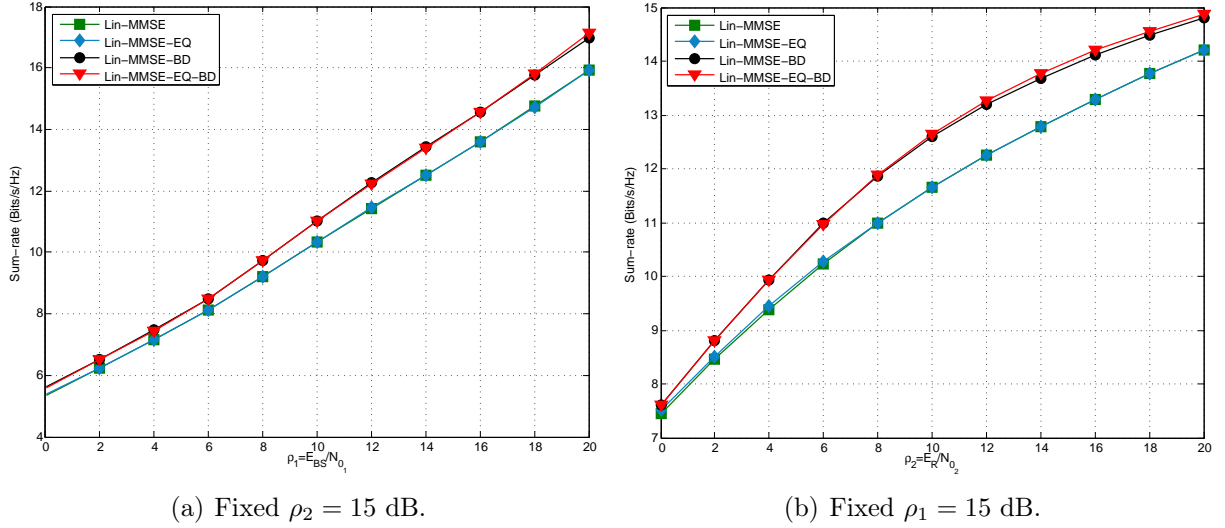
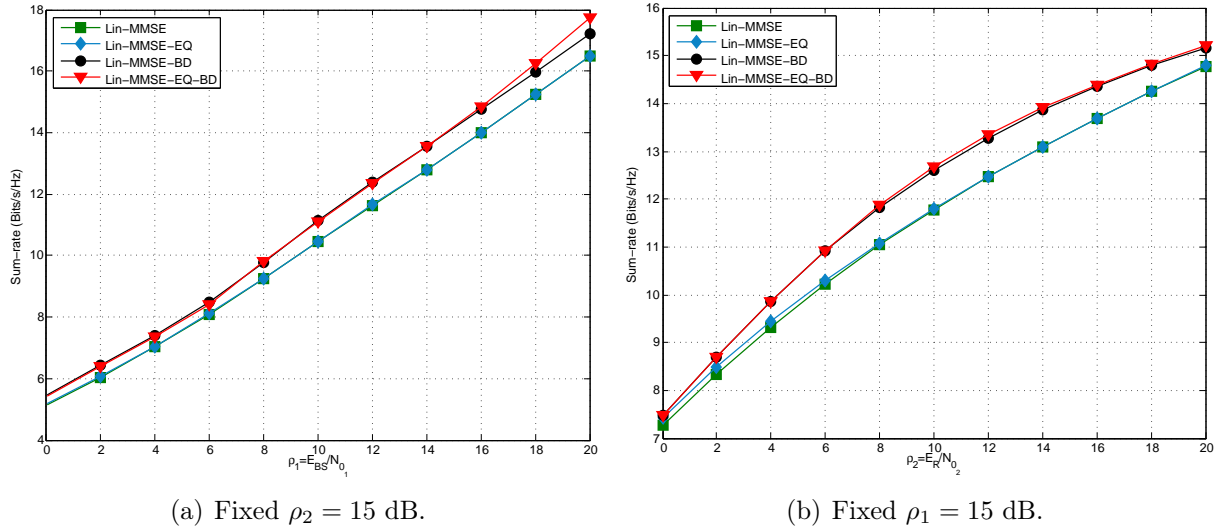
In addition, Figure 3.12(b) shows the sum-rate achieved when ρ_1 is set to 15 dB. Oppositely to the previous case, the gain of BD approaches is obtained at low SNR, being this less than 0.25 dB.

The sum-rate achieved by BD approaches for set-up D is shown in Figure 3.13 when $\rho_2 = 15$ dB (Figure 3.13(a)) and $\rho_1 = 15$ dB (Figure 3.13(b)). Globally, it can be said that *Lin-MMSE-EQ-BD* outperforms the rest of the approaches. When ρ_2 is set to 15 dB, a gain of 2 dB is achieved by BD approaches over linear systems. Furthermore, *Lin-MMSE-EQ-BD* performs slightly better than *Lin-MMSE-BD*, being the maximum gain of 1 dB from $\rho_1 = 14$ dB on. When ρ_1 is fixed to 15 dB, the difference between the approaches with and without equalizer is not perceptible. At middle ρ_2 the maximum difference between BD and linear approaches is given, being the maximum gain of 2.5 dB.

Finally, Figure 3.14 depicts the sum-rate achieved for set-up E when ρ_2 is fixed to 15 dB (Figure 3.14(a)) and $\rho_1 = 15$ dB (Figure 3.14(b)). For this set-up, the effect of the equalizers and block diagonalization is more evident, being the maximum difference between linear and BD approaches greater. For both configuration set-ups, the maximum sum-rate is achieved by *Lin-MMSE-EQ-BD* approach, specially at high SNR. When $\rho_2 = 15$, the best sum-rate is achieved by BD approaches, being the maximum difference respect to linear MMSE approaches of about 2 dB. Oppositely, when $\rho_1 = 15$ dB the gain obtained by BD approaches increases with the SNR, being the maximum gain achieved of 4 dB.

3.3.4.3 Convergence

As mentioned before, the convergence of the system will measure the effectiveness of the iterative algorithm, tested on one hand, measuring the BER when the system is forced to a certain number of iterations, and on the other hand, getting the MSE error after stopping


 Figure 3.13: Sum-rate achieved by MMSE-BD approaches in $8 \times 8 \times \{4 \times 2\}$ set-up (D).

 Figure 3.14: Achievable sum-rate of the MMSE block diagonalization technique with and without equalizer in the $8 \times 8 \times \{2 \times 4\}$ set-up (E).

the algorithm after the execution of a concrete number of iterations. Figure 3.15 shows the bit error rate performance of the systems *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD* when the systems are forced to stop after 1, 3 and 6 iterations.

In both figures it is observed that increasing the number of the iterations the bit error rate decreases, being *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD* the local optimal BD approaches. For example, executing three iterations we gain a dB over *Lin-MMSE-BD* $it=1$ and *Lin-MMSE-EQ-BD* $it=1$, while executing six iterations the systems perform near the ones that carry out three. Therefore, to reduce the computational cost, we can execute a certain number of iterations, avoiding the calculation of the convergence error.

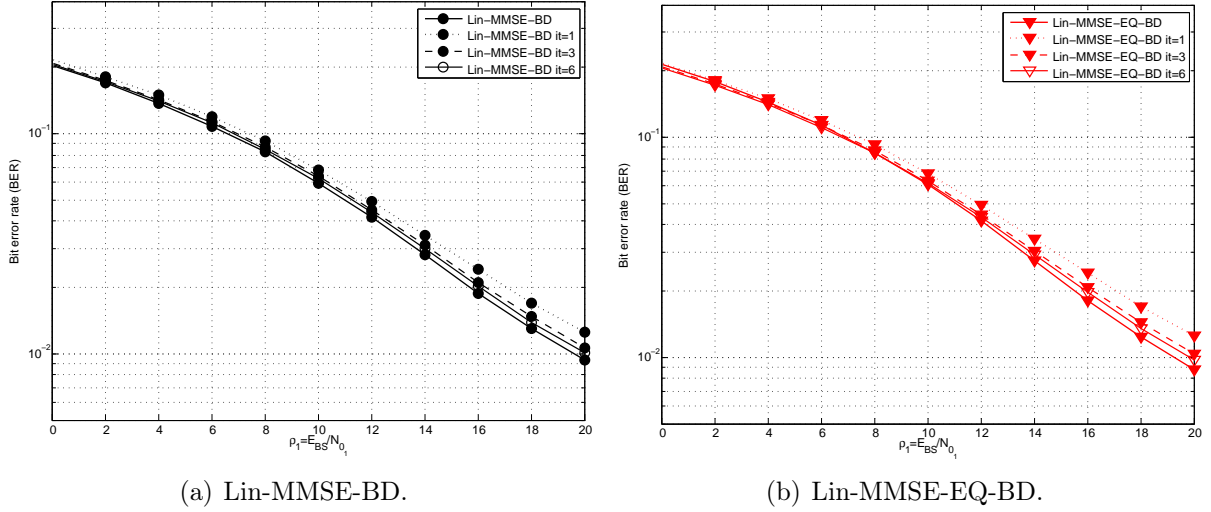


Figure 3.15: BER performance obtained by MMSE-BD approaches for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_2=15$ dB for a fixed number of iterations.

To show the MSE convergence, the mean square error has been plotted for a certain number of iterations in Figure 3.16. This figure depicts the MMSE error of *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD* for a certain number of iterations compared with *Lin-MMSE* and *Lin-MMSE-EQ* approaches.

A and E set-ups have been tested, defining the SNR in pairs, such as:

- $\rho_1=10$ dB and $\rho_2=15$ dB ($\rho_1 < \rho_2$).
- $\rho_1=20$ dB and $\rho_2=15$ dB ($\rho_1 > \rho_2$).

Watching carefully the figures, it can be seen that the error of block diagonalized systems match up with the MMSE results presented before when $4 \times 4 \times \{4 \times 1\}$ set-up is tested. Some conclusions are derived from the results depicted in the Figure 3.16. Firstly when $\rho_1 < \rho_2$ the error is greater than one and is worst increasing the number of antennas and users. On the contrary, if $\rho_1 > \rho_2$ the error is reduced and the effect of the equalizer is more evident.

3.4 Computational Complexity

Throughout this section, the word *complexity* has been used more than once to talk about the cost of the algorithms. This is a crucial fact since the implementation of an algorithm will be possible or not depending on its complexity.

Before continuing, two concepts have to be introduced. The first one, called *complexity order* provides a description of the scaling behaviour of the algorithm's complexity in one or multiple design parameters. For example, a complexity order of $O(n^2)$ specifies that the

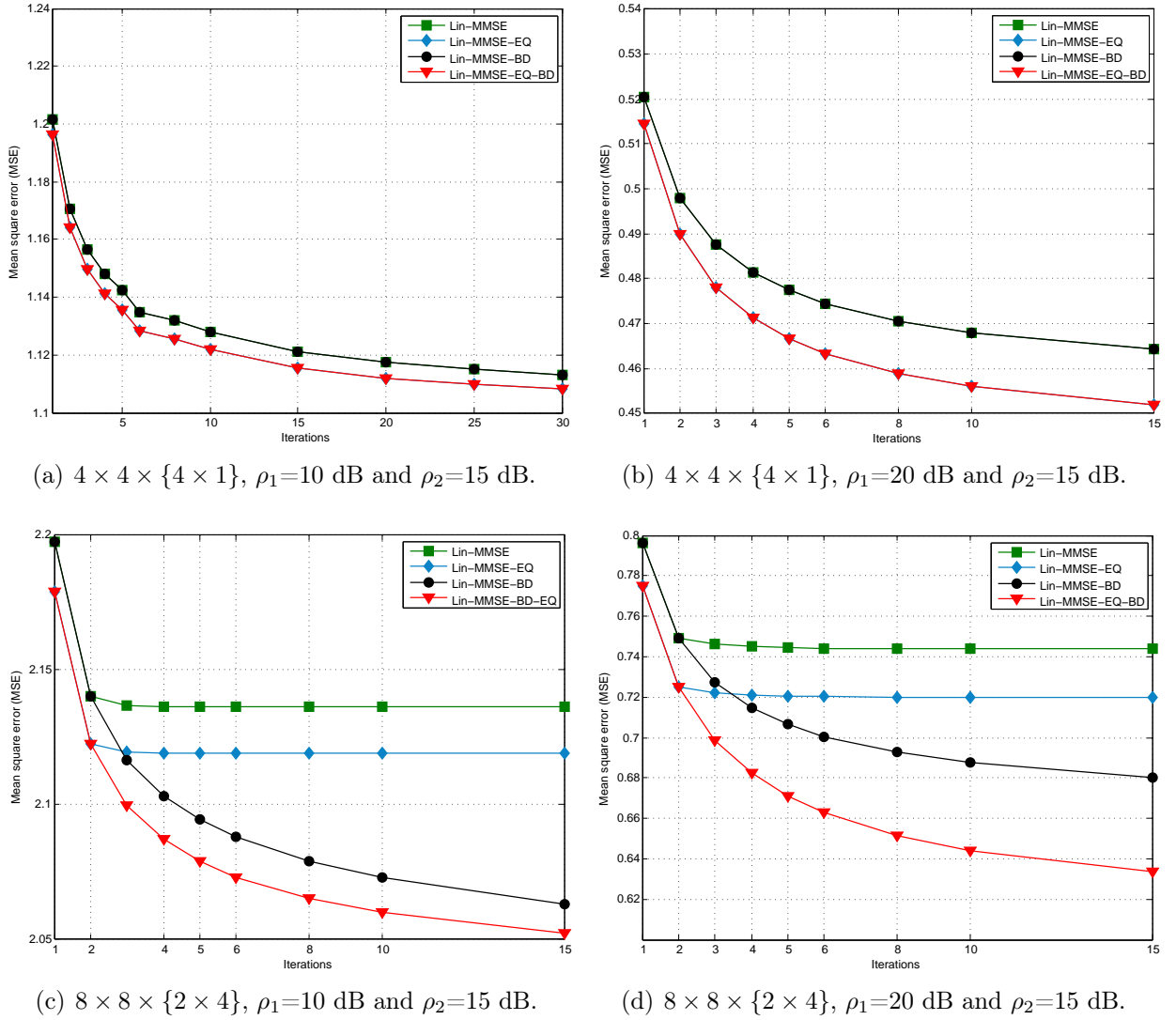


Figure 3.16: Convergence of the MMSE block diagonalization technique with and without equalizer.

complexity is quadratic. The second and the last, named *computational complexity* describes the complexity of an algorithm in terms of costly operations.

After the analysis, we can firmly say that the computational complexity of our algorithm will depend on the number of inverse operations and the loops. For example, the computational complexity of two matrices $\mathbf{A} \in \mathbb{R}^{m \times p}$ and $\mathbf{B} \in \mathbb{R}^{p \times n}$ is defined as $(2p + 1)nm$ floating-point operations (FLOPs). If $n = m = p$, the complexity order is cubic $O(n^3)$ [Golub96]. The number of FLOPs for the computation of the inverse of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is $4n^3 + 6n^2 + 3n$, being the complexity order $O(n^3)$.

In this section, the complexity will be analysed by means of the following measurements. On one hand, the computational complexity of the proposed schemes will be measured in terms of FLOPs. On the other hand, the algorithm execution time will be obtained, computing the average time that the algorithm needs to obtain the precoding and relaying

matrices. This simulation time is computed as the addition of the running-time needed at the BS, the RS and K multi-antenna terminals.

For simplicity, a concrete number of iterations, denoted by it , is assumed. Provided simulations, which show the run-time of the proposed algorithms, are carried out for a fixed $\rho_2 = 15$ dB.

3.4.1 Complexity analysis for the joint linear approaches

This section analyses the computational complexity in FLOPs and the computational order of the linear approaches named *Lin-ZF*, *Lin-MMSE*, *Lin-ZF-EQ* and *Lin-MMSE-EQ*.

For example, Table 3.1 shows the number of operations that are carried out for the computation of the precoding (**F**), relaying (**W**) and equalization (**D**) matrices for both ZF and MMSE criteria.

| Algorithm | Operation | FLOPs |
|-------------|-----------|--|
| Lin-ZF | F | $4M^3 + (4R+7)M^2 + 6MR^2 + (4+6N)MR + (2N+3)M$ |
| | W | $8R^3 + (15+4N+4M)R^2 + (6NM+2N+2M+6)R$ |
| Lin-MMSE | F | $4M^3 + (6R+8)M^2 + (2N+6M+3)R^2 + (4R+3)M + (2R+1)MN + 1$ |
| | W | $10R^3 + (19+4N+4M)R^2 + (6N+2)MR + 2NR + 1$ |
| Lin-ZF-EQ | F | $4M^3 + (4R+7)M^2 + (2N+1+3M)R^2 + (5R+3+2RN+N)R$ |
| | W | $10R^3 + (15+4N+2M)R^2 + (6+2NL+2M+3L+4ML+2N)R$ |
| | D | $4N^3 + (7+2R)N^2 + (2L+6N)R^2 + (2R+4L+1)NM + (2M+2N)RL + (2L+3R)N + 2RM$ |
| Lin-MMSE-EQ | F | $4M^3 + (4R+9)M^2 + (2N+3+6M)R^2 + (5M+N+4NL+2MN+L)R + (3N+2NL+L)M$ |
| | W | $10R^3 + (19+4N+2M)R^2 + (6+M+2N+6NL+L)R + 1$ |
| | D | $4N^3 + (7+2R)N^2 + (2L+6N)R^2 + (2R+4L+1)NM + (2M+2N)RL + (2L+3R)N + 24M$ |

Table 3.1: Number of FLOPs for linear ZF and MMSE approaches.

As depicted in Table 3.1, the number of FLOPs is a function of M , R , N and L , which

stand for the number of antennas at the BS and RS, the total number of receive antennas and the total number of transmitted streams, respectively. Assuming $M = R = N = L = n$ for simplicity, Table 3.2 depicts the total number of FLOPs that are carried out for the computation of the iterative algorithms.

| Algorithm | Comp. | FLOPs | | |
|--------------------|--------------|--------------------------------|---------------------|----------------------|
| | | M=R=N=n | M=R=N=4 | M=R=N=6 |
| Lin-ZF | $O(it\ n^3)$ | $it(42n^3 + 38n^2 + 8n + 6)$ | $3 \times 10^3\ it$ | $10 \times 10^3\ it$ |
| Lin-MMSE | $O(it\ n^3)$ | $it(52n^3 + 54n^2 + 5n + 16)$ | $4 \times 10^3\ it$ | $13 \times 10^3\ it$ |
| Lin-ZF-EQ | $O(it\ n^3)$ | $it(68n^3 + 60n^2 + 12n + 7)$ | $5 \times 10^3\ it$ | $16 \times 10^3\ it$ |
| Lin-MMSE-EQ | $O(it\ n^3)$ | $it(104n^3 + 49n^2 + 13n + 1)$ | $7 \times 10^3\ it$ | $24 \times 10^3\ it$ |

Table 3.2: Complexity order and number of FLOPs for linear ZF and MMSE approaches.

The complexity of joint linear approaches depends on the number of iterations and n . We can state that, for a fixed number of iterations, the complexity order is cubic for all the algorithms. However, the number of FLOPs is higher when an equalization process is included.

Figure 3.17 shows the average time required for the algorithm execution when $\rho_2 = 15$ dB and ρ_1 is kept variable for the set-ups $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$.

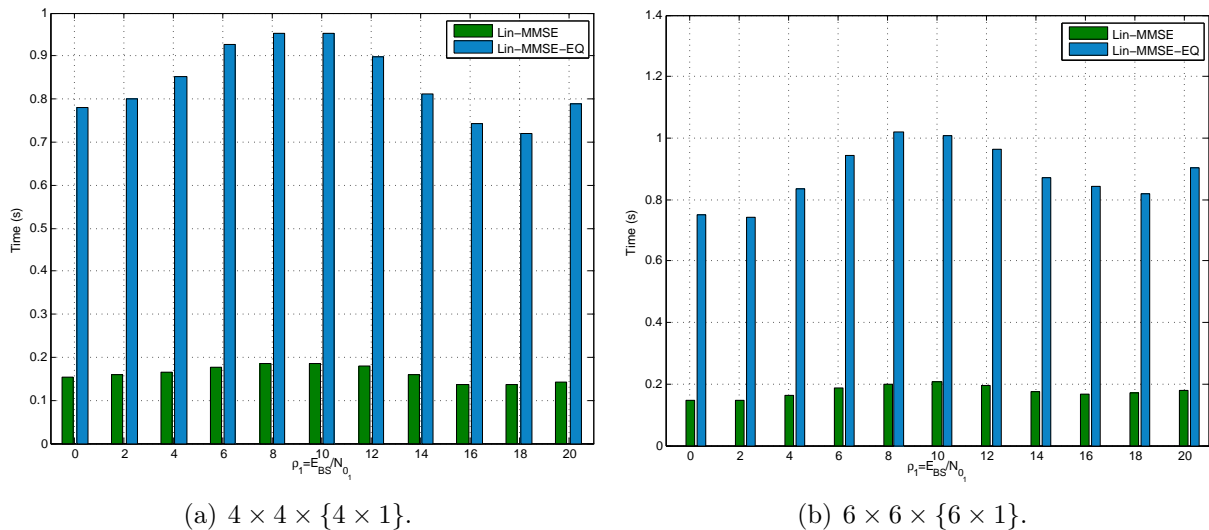


Figure 3.17: Running-time for *Lin-MMSE* and *Lin-MMSE-EQ* for $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ set-ups (A and B) when $\rho_2 = 15$ dB.

As it can be observed in both figures, the execution time required for *Lin-ZF* and *Lin-ZF-EQ* have been omitted due to the low value obtained. For example *Lin-ZF* needs 4.74×10^{-4} seconds to converge while *Lin-ZF-EQ* converges in 9.5×10^{-4} seconds.

Lin-MMSE needs 0.1-0.2 seconds to converge in the set-ups $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ as it can be seen in Figures 3.17(a) and 3.17(b), respectively. As expected, *Lin-MMSE-EQ* algorithm increases the run-time reaching the second when ρ_1 is between 8 and 10 dB, whereas the maximum run-time is achieved for $\rho_1 = 20$ dB when the set-up changes to $6 \times 6 \times \{6 \times 1\}$.

3.4.2 Complexity analysis for block diagonalization based approaches

This section analyses the computational cost related to the proposals called *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD* introduced in Section 3.3.

Table 3.3 shows the amount of FLOPs required for the computation of \mathbf{F} , \mathbf{W} and \mathbf{D} . This operations include matrix products, additions or inverses, being the operations related to the latter expressed as $4n^3 + 6n^2 + 3n$, where n stands for the dimension of the matrix.

| Algorithm | Operation | FLOPs |
|-----------------------|-----------|--|
| Lin-MMSE-BD | F | $2N^3 + 4M^3 + (2 + M)N^2 + (2R + 4N + 7)M^2 + (4R + 4)NM + (2R + 3)M$ |
| | W | $4R^3 + (2N + 2M + 8)R^2 + 10N^3 + (10 + 2M)N^2 + (3 + 6M)NR + (3 + M)N + (3 + M)R + 1$ |
| Lin-MMSE-EQ-BD | F | $6N^3 + 4M^3 + (4 + M)N^2 + (2R + 4N + 7)M^2 + (4R + 4)NM + (2R + 3)M$ |
| | W | $4R^3 + (2N + 2M + 8)R^2 + 14N^3 + (12 + 2M)N^2 + (3 + 6M)NR + (3 + M)N + (3 + M)R + 1$ |
| | D | $4N^3 + 4R^3 + (6 + 2M + 4M + 2L)R^2 + (6 + 2L)N^2 + (6M + 2N + 3)LR + (2L + M + 24 + 3)N + 3$ |

Table 3.3: Number of FLOPs for *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD*.

Assuming $M = R = N = n$ and taking into consideration that $L = \min(M, R, N)$ the expressions are simplified, as it can be seen in Table 3.4. Furthermore, $n = 4$ and $n = 8$ have been assumed as a consequence of the simulations accomplished for BD approaches in Section 3.3.4. As expected, the one with more complexity or the one that executes more operations is *Lin-MMSE-EQ-BD*, which apart from BD executes an equalization process.

Figure 3.18 shows the time spent for the computation of each algorithm versus ρ_1 , assuming $\rho_2 = 15$ dB and computed at three terminals: BS, RS and the end multi-antenna users. The simulations have been performed for $4 \times 4 \times \{4 \times 1\}$, $8 \times 8 \times \{4 \times 2\}$ and $8 \times 8 \times \{2 \times 4\}$ set-ups, represented in Figures 3.18(a), 3.18(b) and 3.18(c), respectively.

Watching carefully the simulations comparing *Lin-MMSE-BD* with *Lin-MMSE*, we can assert that the former adds a time penalty due to the block diagonalization process. The

| Algorithm | Comp. order | FLOPs | | |
|----------------|-------------|---------------------------------|--------------------|---------------------|
| | | M=R=N=n | M=R=N=4 | M=R=N=8 |
| Lin-MMSE-BD | $O(it n^3)$ | $it (71n^3 + 26n^2 + 13n + 10)$ | $5 \times 10^3 it$ | $38 \times 10^3 it$ |
| Lin-MMSE-EQ-BD | $O(it n^3)$ | $it (94n^3 + 76n^2 + 14n + 12)$ | $7 \times 10^3 it$ | $54 \times 10^3 it$ |

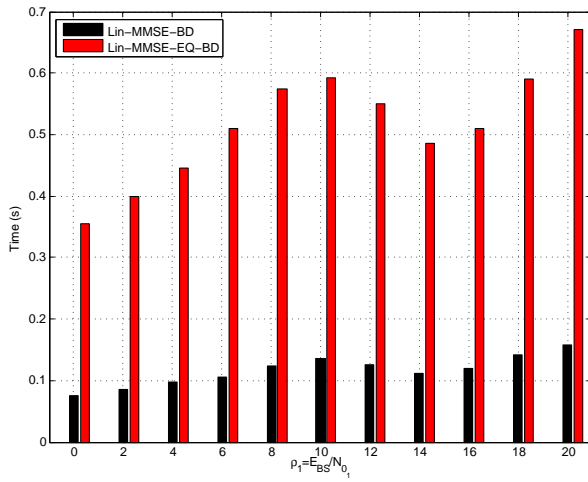
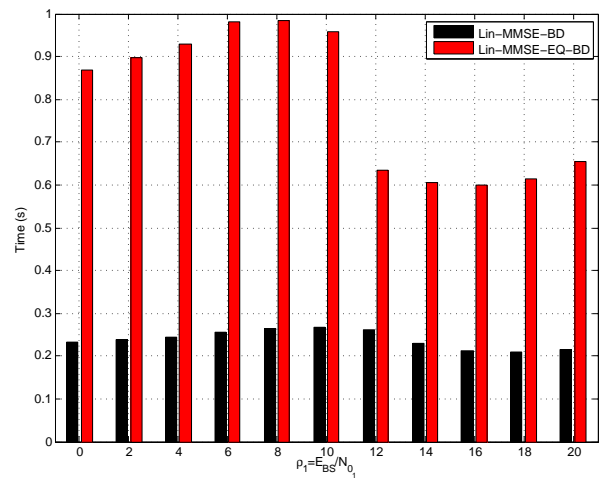
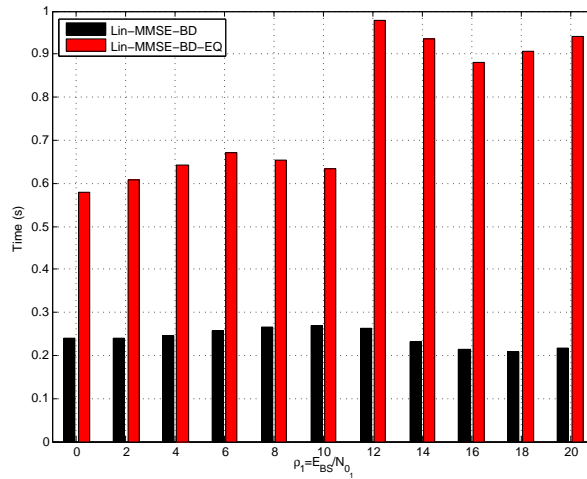
 Table 3.4: Complexity order and number of FLOPs for *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD*.

 (a) $4 \times 4 \times \{4 \times 1\}$.

 (b) $8 \times 8 \times \{4 \times 2\}$.

 (c) $8 \times 8 \times \{2 \times 4\}$.

 Figure 3.18: Running-time for *Lin-MMSE-BD* and *Lin-MMSE-EQ-BD* for $4 \times 4 \times \{4 \times 1\}$, $8 \times 8 \times \{4 \times 2\}$ and $8 \times 8 \times \{2 \times 4\}$ set-ups (A, D and E) when $\rho_2 = 15$ dB.

same conclusion is obtained for the approaches with equalizers. For $4 \times 4 \times \{4 \times 1\}$ set-

up, *Lin-MMSE-BD* presents a continuous run-time over the different values of ρ_1 while the maximum (0,55 seconds) and minimum (0,45 seconds) run-times are achieved when $\rho_1 = 12$ dB and $\rho_1 = 20$ dB respectively for *Lin-MMSE-EQ-BD* proposal.

Obviously and as expected, increasing the number of transmit, receive or relaying antennas, the computational complexity also increases as shown in Figure 3.18. For both multi-antenna configurations (D and E set-ups), the maximum running time is required by *Lin-MMSE-EQ-BD*, largely because the equalization process and BD stage. Figure 3.18(b) shows that up to 10 dB the time required by this approach is higher than from 10 dB on. Furthermore, the maximum time needed is obtained at 8 dB, where *Lin-MMSE-EQ-BD* needs 2 hours for the execution of 10^4 Monte Carlo simulations, whilst for *Lin-MMSE-BD*, *Lin-MMSE-EQ* and *Lin-MMSE* only require 50 minutes, 48 minutes and 13 minutes, respectively.

Finally Figure 3.18(c) depicts set-up E's run-time, where the maximum time needed is obtained from 10 dB on. Up to 10 dB the running-time is constant from 0.6 and 0.7 seconds for *Lin-MMSE-EQ-BD*. The maximum time is employed for 12 dB computation, being required 0.98 seconds for *Lin-MMSE-EQ-BD*. This triplicates the time needed for *Lin-MMSE-BD* and *Lin-MMSE-EQ*, whilst is 10 times higher than the time needed for *Lin-MMSE* computation.

3.5 Chapter summary

This chapter has considered a multiuser MIMO AF downlink relaying systems with linear precoding at the BS under minimization of the MSE of the system subject to two power constraints. Basically, linear approaches have been proposed for the joint design of the precoding and relaying matrices. The chapter has started with the presentation of the model, that will be the one employed throughout this thesis. Section 3.1 has presented a framework composed by $K + 2$ terminals: a base station, a relay and K multi-antenna users.

Afterwards, the chapter has continued with the analysis of the algorithms carried out in Section 3.2 under ZF and MMSE criteria. Furthermore, due to the non-convexity of the optimization problem a numerical method has been proposed as a solution, achieving in this way a local optimal solution.

Moreover, block diagonalization based designs have been proposed in Section 3.3, presenting BD as an alternative to DPC. By means of BD the interference created by other users have been cancelled, creating multiple point-to-point parallel subchannels. As a consequence of the lack of convergence of BD approach that diagonalizes the system from the BS to the end users, a novel local optimal solution has been proposed, combining linear precoding at the BS and BD at the relay.

Provided simulation results, given in terms of BER or achievable sum-rate, have showed

the effectiveness of BD based approach over local optimal linear frameworks in Section 3.3.4, specially for multi-antenna receivers. Finally the computational cost of the aforementioned proposals has been given in Section 3.4, taking into consideration the complexity order, the computational complexity and the required running-time, measured at the BS, RS and the end multi-antenna users.

Chiefly, the contributions of this chapter revolve around the joint design of the precoding and relaying matrices based on block diagonalization approach, being the combination of BD at the relaying station and linear precoding at the base station proposed firstly. Furthermore, a complete complexity analysis is given in terms of complexity order and algorithm execution time.

Suboptimal Linear Precoding in Multiuser MIMO AF Relaying Systems

As it has been analysed in the previous chapter, the research of multiuser MIMO relaying networks has focused on the joint design of precoding and relaying matrices under linear processing [Li09][Xu11]. Apart from the non-convexity of the optimization problem, which makes unfeasible to obtain a global solution, the main drawback of the aforementioned joint design is the high computational cost that it entails due to the iterative processing algorithm required, which makes the implementation of the algorithms difficult.

This computational cost, directly related to the number of transmit, relaying and receive antennas, can be heavily reduced by the suboptimal approaches proposed in this chapter. As it will be shown, the considered suboptimal approaches can be divided into three groups. While the first one consists of the design of the precoding matrix considering an all-pass filter at the relay and block diagonalization of the overall system, the second one is based on the minimization of the mean square error of each hop independently. Finally, the third and last group assumes the total diagonalization of the system from the BS to the end users, reducing the cost of the design considerably.

In the first section of this chapter the linear BD design based on all-pass filtering at the relay is proposed and optimized based on MSE minimization. The all-pass filter will be only in charge of limiting the transmission power at the relay and will not apply any processing to the received signal. In the following section, a simple but efficient suboptimal independent hop-by-hop strategy is proposed for the design of the matrices, ending with the system diagonalization based approaches, which can target either MSE minimization or sum-rate maximization. For all the design approaches considered in this chapter, the system model depicted in Figure 3.1 is considered, which has been analysed in Section 3.1.

4.1 Suboptimal Linear Block Diagonalization with All-Pass Relaying

Block diagonalization has been considered as precoding strategy in multiuser MIMO systems in [Spencer04b] [Shim08] for inter-users' interference cancellation. Due to the good results obtained in multiuser MIMO schemes, this precoding technique has been considered for the local joint optimization of the precoding and relaying matrices in Section 3.3.3. Nevertheless, its computational complexity has been shown to be prohibitive, mostly due to the required iterative process.

In order to reduce the computational cost, a suboptimal approach based on block diagonalization is considered in this section employing an all-pass filter at the relay. Hence, the signal received at the relay does not suffer any change, being its mainly aim the amplification of the output power up to E_R .

In few words, the proposed technique provides a simple solution for BD-based AF multiuser MIMO relaying systems. This solution, also known as *naive* solution, exerts the relaying matrix $\mathbf{W} = \beta_2^{-1} \mathbf{I}_R$, which just amplifies the signal and forwards it without further processing. The power is limited at the relay station to E_R by $1/\beta_2$, being the scaling factor β_2 defined as

$$\beta_2 = \sqrt{\frac{\text{Tr} \left(\mathbf{H}_1 \overline{\mathbf{F}} \overline{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R \right)}{E_R}}.$$

The straightforwardness of this naive solution, in opposite to the local optimal solution introduced in Section 3.3.3, is that the application of an iterative algorithm to obtain the solution is not necessary. Being \mathbf{W} independent of \mathbf{F} , the precoding matrix can be easily computed. For simplicity, we assume $\mathbf{D}_k = \mathbf{I}_{L_k \times N_k}$, considering no equalization and $L_k = N_k$. Hence, the total received signal at the k^{th} user is reduced to

$$\hat{\mathbf{d}}_k = \beta_1 \mathbf{H}_{2,k} \mathbf{H}_1 \mathbf{F}_k \mathbf{s}_k + \underbrace{\beta_1 \sum_{\substack{j=1 \\ j \neq k}}^K \mathbf{H}_{2,k} \mathbf{H}_1 \mathbf{F}_j \mathbf{s}_j}_{\text{Interference}} + \underbrace{\beta_1 \mathbf{H}_{2,k} \mathbf{n}_1 + \beta_2 \mathbf{n}_{2,k}}_{\text{Noise}},$$

where $\mathbf{F}_k \in \mathbb{C}^{M \times L_k}$ is the precoded matrix related to the k^{th} user grouped into the matrix $\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_k, \dots, \mathbf{F}_K] \in \mathbb{C}^{M \times L}$. If the equivalent channel is defined as $\overline{\mathbf{H}}_k = \mathbf{H}_{2,k} \mathbf{H}_1$, the constraint for BD is the following: $\overline{\mathbf{H}}_k \mathbf{F}_j = \mathbf{0}_{N_k \times L_k}$ for all $j \neq k$ $1 \leq j, k \leq K$. With the help of block diagonalization, \mathbf{F}_j is forced to lie in the null space of $\tilde{\mathbf{H}}_j = [\overline{\mathbf{H}}_1^T, \dots, \overline{\mathbf{H}}_{j-1}^T, \overline{\mathbf{H}}_{j+1}^T, \dots, \overline{\mathbf{H}}_K^T]^T \in \mathbb{C}^{N-N_k \times M}$, which contains the channel matrices of interfering users. Assuming $\mathbf{F}_k = \mathbf{F}_{2,k} \mathbf{F}_{1,k}$, the precoding matrix at the BS is defined as

$\mathbf{F} = [\mathbf{F}_1 \dots \mathbf{F}_k \dots \mathbf{F}_K] = \mathbf{F}_2 \mathbf{F}_1$, where the matrix \mathbf{F}_2 is used for the diagonalization of the channel and \mathbf{F}_1 is employed for channel adaptation. After the singular value decomposition of $\tilde{\mathbf{H}}_j$ and carrying out the same procedure described in Section 3.3, the following expression is obtained: $\mathbf{F}_{2,j} = \tilde{\mathbf{V}}_j^{(0)} \in \mathbb{C}^{M \times L_k}$, being $\tilde{\mathbf{V}}_j^{(0)}$ the first L_k zero right singular vectors. After the application of the block diagonalization constraint, the signal received at the k^{th} user is given by:

$$\hat{\mathbf{d}}_k = \beta_1 \mathbf{H}_{e,k} \mathbf{F}_{1,k} \mathbf{s}_k + \beta_1 \mathbf{H}_{2,k} \mathbf{n}_1 + \beta_2 \mathbf{n}_{2,k}, \quad (4.1)$$

where $\mathbf{H}_{e,k} = \mathbf{H}_{2,k} \mathbf{H}_1 \mathbf{F}_{2,k} \in \mathbb{C}^{N_k \times L_k}$ denotes the equivalent interference-free channel created from the BS to the end user.

For the optimization of \mathbf{F}_1 , the metric employed is the minimization of the mean square error, thus, the local optimal precoding matrix will minimize the sum-MSE, which can be designed as

$$\begin{aligned} \xi &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{s}_k\|_2^2 \right], \\ &= \sum_{k=1}^K \text{Tr}(\beta_1^2 \mathbf{H}_{e,k} \mathbf{F}_{1,k} \mathbf{F}_{1,k}^H \mathbf{H}_{e,k}^H + \beta_1^2 N_{01} \mathbf{H}_{2,k} \mathbf{H}_{2,k}^H + \beta_2^2 N_{02} \mathbf{I}_{N_k} - 2\Re(\beta_1 \mathbf{H}_{e,k} \mathbf{F}_{1,k}) + \mathbf{I}_{N_k}), \end{aligned} \quad (4.2)$$

being $\hat{\mathbf{d}}_k$ defined in (4.1). In the following sections, the precoding matrix will be derived under both ZF and MMSE criteria.

4.1.1 Zero Forcing

This section obtains the precoding matrix \mathbf{F} and the scaling factor β_1 under zero-forcing criterion, i.e., setting the constraint $\beta_1 \mathbf{H}_e \mathbf{F}_1 = \mathbf{I}_{N \times L}$. Applying this to the error definition in (4.2), we get

$$\xi_{ZF} = \sum_{k=1}^K \beta_1^2 N_{01} \text{Tr}(\mathbf{H}_{2,k} \mathbf{H}_{2,k}^H) + \beta_2^2 N_{02} N.$$

The optimization problem is now defined as

$$\begin{aligned} \{\mathbf{F}_1, \beta_1\} &= \underset{\{\mathbf{F}_1, \beta_1\}}{\text{argmin}} \quad \xi_{ZF} \\ \text{s.t.} \quad &\sum_{k=1}^K \text{Tr}(\mathbf{F}_{1,k} \mathbf{F}_{1,k}^H) = E_{BS}, \\ &\beta_1 \mathbf{H}_e \mathbf{F}_1 = \mathbf{I}_{N \times L}, \end{aligned}$$

where $\mathbf{H}_e = \text{blkdiag}(\mathbf{H}_{e,1}, \dots, \mathbf{H}_{e,k}, \dots, \mathbf{H}_{e,K}) \in \mathbb{C}^{N \times L}$ is the total equivalent channel matrix. The second constraint is defined for the zero forcing solution. Due to the convexity of the optimization problem, the problem can be solved using Lagrange multipliers as described in Appendix B, obtaining:

$$\mathbf{F}_{1,k} = \frac{1}{\beta_1} (\mathbf{H}_{e,k}^H \mathbf{H}_{e,k})^{-1} \mathbf{H}_{e,k}^H \quad \text{and} \quad \beta_1 = \sqrt{\frac{\text{Tr} \left((\mathbf{H}_e \mathbf{H}_e^H)^{-1} \right)}{E_{BS}}},$$

where the expression obtained for \mathbf{F}_1 equals the Moore-Penrose pseudo inverse of the matrix $\mathbf{H}_{e,k}$, piled up into the matrix $\mathbf{F}_1 = \text{blkdiag}(\mathbf{F}_{1,1}, \dots, \mathbf{F}_{1,k}, \dots, \mathbf{F}_{1,K})$.

4.1.2 Minimum Mean Square Error

Alternatively, the design based on MMSE criterion is represented by the next optimization problem, where the local optimal precoding matrix and scaling factor are obtained by means of the minimization of the mean square error:

$$\begin{aligned} \{\mathbf{F}_1, \beta_1\} &= \underset{\{\mathbf{F}_1, \beta_1\}}{\text{argmin}} \quad \xi \\ \text{s.t.} \quad &\text{Tr}(\mathbf{F}_1 \mathbf{F}_1^H) = E_{BS}, \end{aligned}$$

where ξ is the mean square error defined in (4.2). After setting the derivatives with respect to the variables to be optimized and continuing the procedure described in Appendix B.2, the following expressions are obtained

$$\begin{aligned} \mathbf{F}_{1,k} &= \frac{1}{\beta_1} \left(\mathbf{H}_{e,k}^H \mathbf{H}_{e,k} + \frac{N_{0_1}}{E_{BS}} \text{Tr}(\mathbf{H}_{2,k}^H \mathbf{H}_{2,k}) \mathbf{I}_{N_k} \right)^{-1} \mathbf{H}_{e,k}^H \quad \text{and} \\ \beta_1 &= \sqrt{\frac{\sum_{k=1}^K \text{Tr} \left(\mathbf{H}_{e,k} \left(\mathbf{H}_{e,k}^H \mathbf{H}_{e,k} + \frac{N_{0_1}}{E_{BS}} \text{Tr}(\mathbf{H}_{2,k}^H \mathbf{H}_{2,k}) \mathbf{I}_{N_k} \right)^{-2} \mathbf{H}_{e,k}^H \right)}{E_{BS}}}, \end{aligned}$$

where $\frac{N_{0_1}}{E_{BS}} \text{Tr}(\mathbf{H}_{2,k}^H \mathbf{H}_{2,k})$ is the regularization factor.

To sum up, it should be pointed out that the naive solution presented throughout this section under ZF and MMSE criteria does not require any iterative process for the optimization, reducing the computational cost considerably.

4.1.3 Simulation Results

The performance of the BD-based all-pass linear filter approach introduced in this section is here analysed. Three different measurements have been conducted: BER performance,

sum-rate analysis and MSE behaviour of the system.

We assume the same notation employed in Section 3.2.4 for set-up definition: $M \times R \times \{K \times N_k\}$, considering only the set-up A ($M = 4$, $R = 4$, $K = 4$ and $N_k = 1$) to validate the performance of the evaluated framework.

An ensemble of 10^4 channel matrices \mathbf{H}_i for $i = 1, 2$ have been simulated with Rayleigh distribution, being \mathbf{H}_1 and \mathbf{H}_2 the channel matrices for hops 1 and 2, respectively. QPSK symbols have been transmitted for simulations and SNR values are considered independent for first (ρ_1) and second (ρ_2) hops.

As it will be seen in the following figures, the performance of BD all-pass filter approach, called *Lin-MMSE-BD-All-Pass-filter*, has been compared with the joint linear ZF design (*Lin-ZF*) and the joint linear BD partial approach (*Lin-MMSE-BD*) proposed in Chapter 3. It should be pointed out that the results for ZF-BD all pass solution has been omitted due to the results obtained.

4.1.3.1 BER Performance

This section analyses the BER performance of *Lin-MMSE-BD-All-Pass-Filter* approach compared with joint linear designs *Lin-ZF* and *Lin-MMSE-BD*. Figure 4.1 shows BER performance when $\rho_2 = 15$ dB (Figure 4.1(a)) and $\rho_1 = 15$ dB (Figure 4.1(b)).

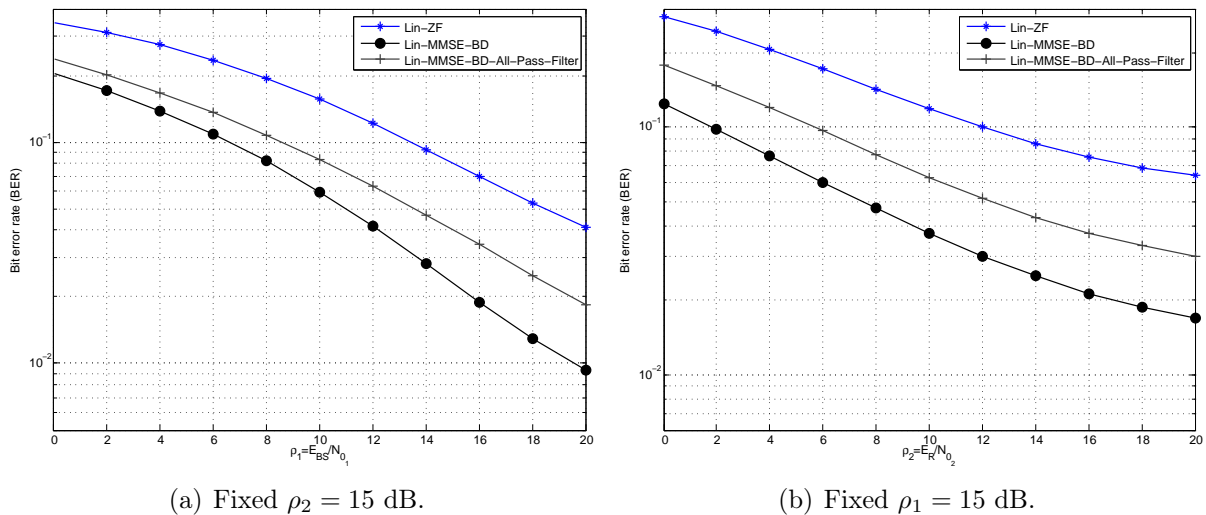


Figure 4.1: BER curves for *Lin-MMSE-BD-All-Pass-Filter* approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation.

Obviously and as expected, *Lin-MMSE-BD-All-Pass-Filter* performs worst than *Lin-MMSE-BD* but better than *Lin-ZF*. Furthermore, comparing *Lin-MMSE-BD-All-Pass-Filter* with the former, we can observe the effectiveness of the processor at the relay. When $\rho_2 = 15$ dB, the maximum gain that *Lin-MMSE-BD* achieves over *Lin-MMSE-BD-All-Pass-Filter*

solution is 4 dB, whereas the maximum gain obtained by the BD naive approach with respect to *Lin-ZF* is about 5 dB, being this constant from $\rho_1 = 0$ to $\rho_1 = 20$ dB.

The same conclusions hold when $\rho_1 = 15$ dB as depicted in Figure 4.1(b). The difference gap with respect to *Lin-MMSE-BD* reaches 8 dB, whilst it increases to 10 dB in comparison with *Lin-ZF*. For a fixed ρ_2 the performance get is worst than when ρ_1 is set to a certain value. We derive that effect of ρ_1 is greater, deteriorating the performance of the system.

4.1.3.2 Achievable Sum-Rate

The achievable sum-rate of the system, obtained by simulation, measures the maximum capacity obtained by the users, defined as in (3.13). Figure 4.2 depicts the sum-rate of $4 \times 4 \times \{4 \times 1\}$ set-up when $\rho_2=15$ dB (Figure 4.2(a)) and $\rho_1=15$ dB (Figure 4.2(b)).

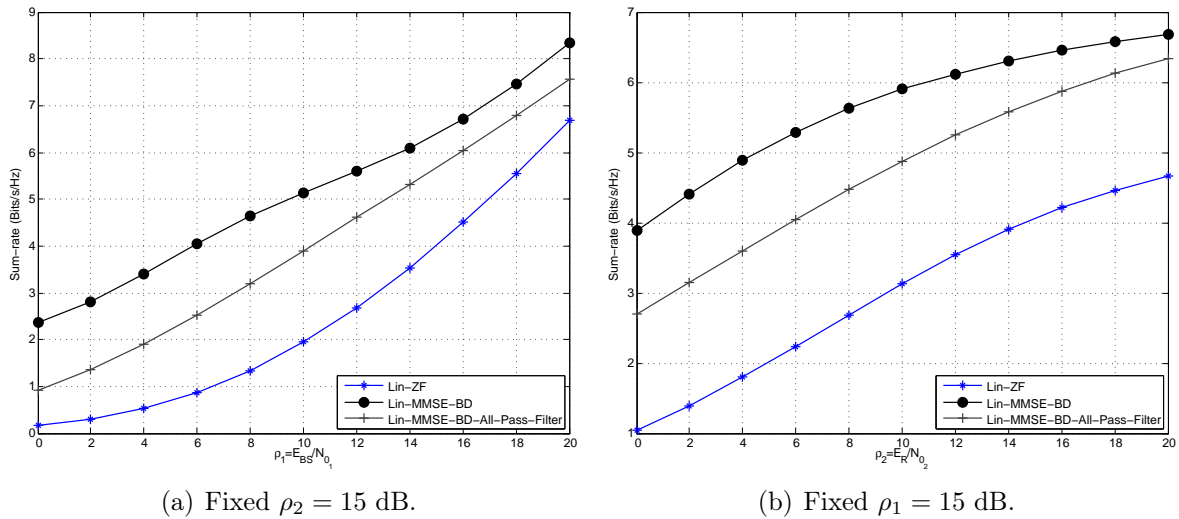


Figure 4.2: Achievable sum-rate of *Lin-MMSE-BD-All-Pass-Filter* approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

As expected, the achievable sum-rate of the joint design *Lin-MMSE-BD* is better compared to *Lin-MMSE-BD-All-Pass-filter* approach when both $\rho_2 = 15$ dB and $\rho_1 = 15$ dB. However, for both configurations we observe that the difference gain achieved decreases as the SNR increases, obtaining a gain of 5 dB when $\rho_2 = 15$ dB and 6 dB when $\rho_1 = 15$ dB. Furthermore, as happens with the BER performance, *Lin-MMSE-BD-All-Pass-Filter* outperforms *Lin-ZF* in both configurations, being the maximum difference about 6 dB and 8 dB for $\rho_2 = 15$ dB and $\rho_1 = 15$ dB, respectively. Provided sum-rate simulations depict that when ρ_1 is fixed the performance obtained is worst, having more impact on the performance the first hop noise.

4.1.3.3 Error Measurement

As mentioned in the previous chapter, the error of the proposal will measure the effectiveness of the systems. Before proceeding, it should be underlined that the error of *Lin-MMSE-All-Pass-Filter* approach is constant over the different number of iterations owing to the non-iterative process nature of the approach.

Figure 4.3 shows the convergence for set-up A when the SNRs are fixed in pairs. Figures 4.3(a) and 4.3(b) depict the error measurement for $\{\rho_1 = 10 \text{ dB}, \rho_2 = 15 \text{ dB}\}$ and $\{\rho_1 = 20 \text{ dB}, \rho_2 = 15 \text{ dB}\}$, respectively.

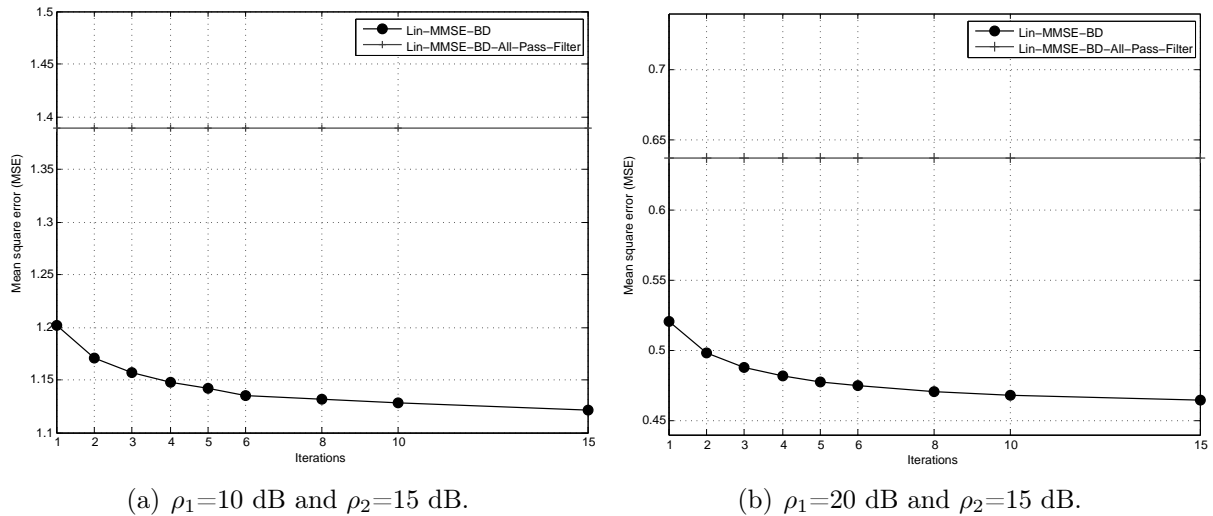


Figure 4.3: MSE error of *Lin-MMSE-BD-All-Pass-Filter* approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

It can be seen that when $\rho_1 < \rho_2$ the error obtained is over the unit whereas when $\rho_1 > \rho_2$ the error reached is under the unit, being the error higher with a higher number of users and antennas.

4.2 Independent Hop-by-Hop MMSE Approach

The independent hop-by-hop MMSE solution (*Lin-MMSE-HH*), is here proposed for the design of the precoding and relaying matrices, avoiding the application of an iterative algorithm and the complexity of BD approaches.

Through *Lin-MMSE-HH* approach, the mean square error of each hop is minimized independently. Thus, with the reduction of the MSE at the first hop, the precoding matrix \mathbf{F} and the power scaling factor β_1 are optimized, whereas with the minimization of the MSE of the second hop, the relaying matrix \mathbf{W} and the scaling factor β_2 are obtained, where $\mathbf{D}_k = \mathbf{I}_{L_k \times N_k}$ has been assumed for simplicity.

The framework employed for the analysis is again the one described in the Figure 3.1. Briefly, it should be underlined that *Lin-MMSE-HH* solution obtains a non-dependent solution at the cost of introducing some performance degradation. MMSE or Wiener solution is considered due to its effectiveness over the zero forcing solution.

As mentioned before, the precoding matrix \mathbf{F} and the scaling factor β_1 are computed, minimizing the MSE from the BS to the relay, defined as

$$\xi_1 = \mathbb{E} [\|\mathbf{y}_1 - \mathbf{s}\|_2^2] = \text{Tr} (\beta_1^2 \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R - 2\Re (\beta_1 \mathbf{H}_1 \mathbf{F}) + \mathbf{I}_M),$$

where \mathbf{y}_1 is the signal received by the relay after re-scaling process, defined in (2.3). For *Lin-MMSE-HH* analysis, $M = R = N$ is assumed. In order to get a solution, the optimization problem is defined as

$$\begin{aligned} \{\mathbf{F}, \beta_1\} &= \underset{\{\mathbf{F}, \beta_1\}}{\text{argmin}} \quad \xi_1 \\ \text{s.t.} \quad &\text{Tr} (\mathbf{F} \mathbf{F}^H) = E_{BS}. \end{aligned} \quad (4.3a)$$

The constraint (4.3a) represents the limit for the transmission power at the BS. After some manipulations, the precoding matrix is computed as

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \mathbf{H}_1 + \frac{N_{01} R}{E_{BS}} \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H. \quad (4.4)$$

The filter MMSE for the BS defined in (4.4) is a regularized inverse of the first hop channel, where $\frac{N_{01}}{E_{BS}}$ corresponds to the signal-to-noise ratio of the first hop. The mentioned regularization term introduces a bias that gives more reliable results when the matrix is ill-conditioned or the channel is noisy. The scaling factor β_1 is equally computed as:

$$\beta_1 = \sqrt{\frac{\text{Tr} \left(\mathbf{H}_1^H \left(\mathbf{H}_1 \mathbf{H}_1^H + \frac{N_{01} R}{E_{BS}} \mathbf{I}_R \right)^{-2} \mathbf{H}_1 \right)}{E_{BS}}}.$$

Similarly to the first hop, the relaying matrix \mathbf{W} and the scaling factor β_2 that minimize the MSE of the second hop should be found, being the optimization problem defined as

$$\begin{aligned} \{\mathbf{W}, \beta_2\} &= \underset{\{\mathbf{W}, \beta_2\}}{\text{argmin}} \quad \xi_2 \\ \text{s.t.} \quad &\text{Tr} (\mathbf{y}_R \mathbf{y}_R^H) = E_R, \end{aligned}$$

where $\mathbf{y}_R \in \mathbb{C}^R$ is the signal transmitted by the relay after the processing and ξ_2 the mean

square error of the second hop, defined as:

$$\xi_2 = \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{y}_1\|_2^2 \right] = \text{Tr} \left(\beta_2^2 \mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H \mathbf{H}_2^H + \beta_2^2 N_{0_2} \mathbf{I}_N - 2\Re \left(\beta_2 \mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1} \right) + \mathbf{R}_{\mathbf{y}_1} \right),$$

being $\mathbf{R}_{\mathbf{y}_1} \in \mathbb{C}^{R \times R}$ the covariance matrix of the input signal at the relay and $\hat{\mathbf{d}}$ the total received signal after re-scaling, defined as $\hat{\mathbf{d}} = \beta_2 \mathbf{H}_2 \mathbf{W} \mathbf{y}_1 + \beta_2 \mathbf{n}_2$. Following the same procedure, the relaying matrix is computed identically, leading to:

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} N}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H. \quad (4.5)$$

In addition, the scaling factor β_2 is determined by

$$\beta_2 = \sqrt{\frac{\text{Tr} \left(\left(\mathbf{H}_1 \overline{\mathbf{F}} \mathbf{F}^H \mathbf{H}_1^H + \beta_1^2 N_{0_1} \mathbf{I}_R \right) \mathbf{H}_2 \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} N}{E_R} \mathbf{I}_R \right)^{-2} \mathbf{H}_2^H \right)}{E_R}},$$

where $\overline{\mathbf{F}} = \beta_1 \mathbf{F}$ denotes the unscaled version of the precoding matrix. Evidently, looking at the expressions (4.4) and (4.5), it can be seen that there is no dependence between them, reducing the computational cost of the joint design approaches considerably.

4.2.1 Simulation Results

This section analyses the performance of the proposed suboptimal *Lin-MMSE-HH* approach. In order to show the effectiveness of this system, we compare this approach with the joint design approaches presented in Chapter 3 named *Lin-ZF* and *Lin-MMSE*, as well as with *Lin-MMSE-BD-All-Pass-Filter* introduced in the previous section for the parameters specified in Section 4.1.3.

4.2.1.1 BER Performance

Figure 4.4 depicts the BER performance of *Lin-MMSE-HH* compared with the joint local optimal approaches *Lin-ZF* and *Lin-MMSE*, and the all-pass relaying BD approach named *Lin-MMSE-BD-All-Pass-Filter* for the set-up A ($4 \times 4 \times \{4 \times 1\}$) when $\rho_2=15$ dB (Figure 4.4(a)) and $\rho_1=15$ dB (Figure 4.4(b)).

When $\rho_2 = 15$ dB, as depicted in Figure 4.4(a), until $\rho_1 = 10$ dB *Lin-ZF* performs better than *Lin-MMSE-HH*. From $\rho_1 = 16$ dB on, *Lin-MMSE-BD-All-Pass-Filter* is outperformed by *Lin-MMSE-HH*. Obviously and as expected, *Lin-MMSE* improves *Lin-MMSE-HH*'s performance. Nevertheless, *Lin-MMSE-HH* might perform better than *Lin-MMSE* at high SNR if more than $\rho_1 = 20$ dB is simulated. Furthermore, note that the gain obtained by *Lin-MMSE* over *Lin-MMSE-HH* decreases while ρ_1 increases.

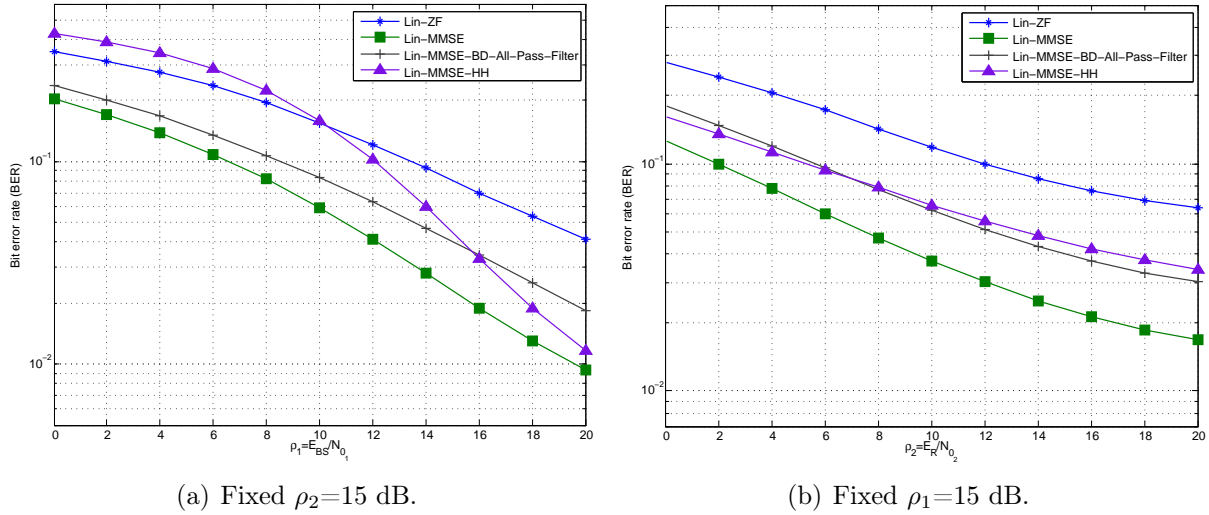


Figure 4.4: BER performance of the proposed *Lin-MMSE-HH* approach for the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation.

In the same way, when $\rho_1 = 15$ dB is established as shown in Figure 4.4(b), where *Lin-MMSE-HH* outperforms *Lin-ZF*, being a gain of 10 dB obtained with respect to this approach. In the same fashion, from $\rho_2 = 6$ dB on, *Lin-MMSE-All-Pass-Filter* improves *Lin-MMSE-BD-HH*'s performance slightly. Oppositely to the previous case, *Lin-MMSE* is not outperformed by *Lin-MMSE-HH*, being the difference greater as ρ_2 increases.

Summing up, it should be pointed out that ρ_1 affects more than ρ_2 , limiting the performance of the system. Despite its performance, *Lin-MMSE-HH* can be considered a good candidate for the design thanks to the complexity obtained.

4.2.1.2 Achievable Sum-Rate

In this section we show the achievable sum-rate for the aforementioned suboptimal approach. When $\rho_2 = 15$ dB is set, as depicted in Figure 4.5(a) for set-up A, the system *Lin-MMSE-HH* performs better than *Lin-ZF* until 16 dB are reached. From 16 dB on, the situation changes achieving a greater capacity for *Lin-ZF* approach. The greatest capacity is obtained by *Lin-MMSE* and *Lin-MMSE-BD-All-Pass-Filter* approaches, being their performance similar at high SNRs.

For fixed $\rho_1 = 15$ dB, the results obtained are the opposite as shown in Figure 4.5(b), where *Lin-MMSE-HH* achieves a better capacity compared to *Lin-MMSE-BD-All-Pass-Filter* and *Lin-ZF*, being the difference gap between *Lin-MMSE-HH* and *Lin-ZF* remarkable.

4.2.1.3 Error Measurement

As it happens with the naive solution, *Lin-MMSE-HH* does not update the precoding and relaying matrices iteratively, since these are not interdependent. Therefore, the error will be

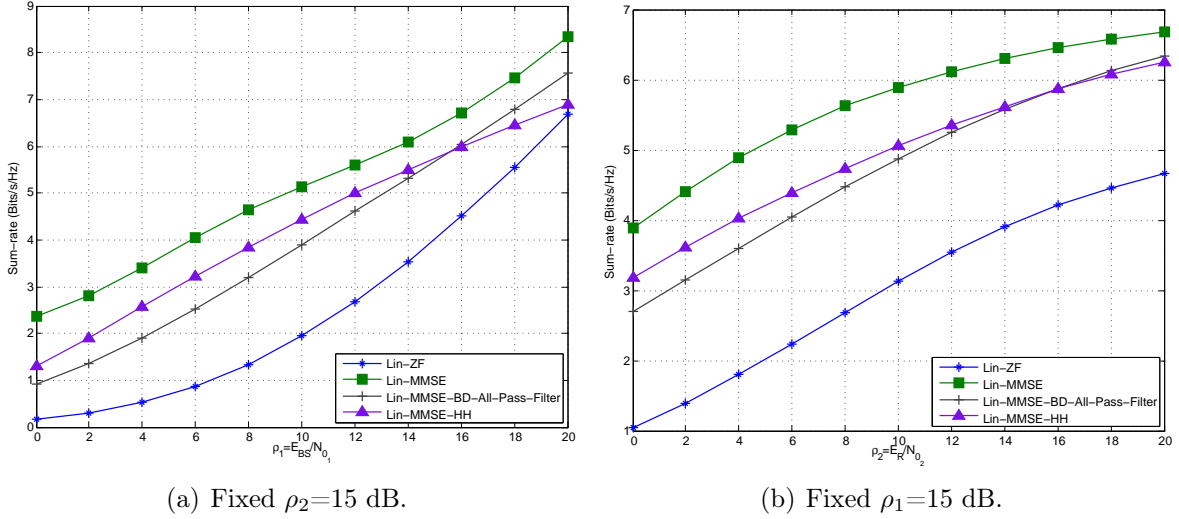


Figure 4.5: Comparison of sum-rate versus SNR for *Lin-MMSE-HH* approach for the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

plotted constant for all iterations.

In order to show the error obtained by *Lin-MMSE-HH*, Figure 4.6 shows the MSE error for a fixed number of iterations for $4 \times 4 \times \{4 \times 1\}$ set-up when $\rho_1 = 10$ dB and $\rho_2 = 15$ dB.

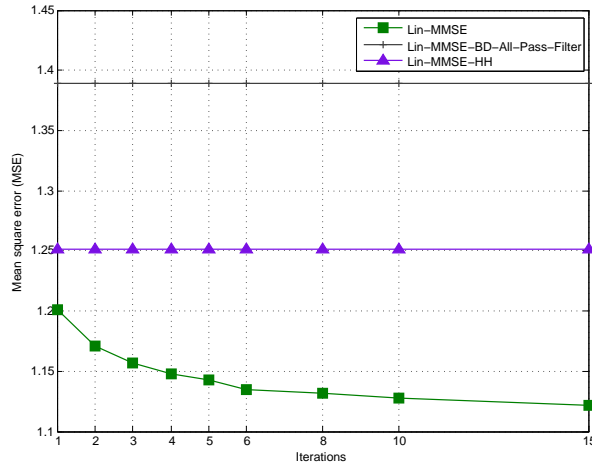


Figure 4.6: Convergence analysis by means of MSE error versus numbers of iterations for *Lin-MMSE-HH* approach for the $4 \times 4 \times \{4 \times 1\}$ set-up (A) when $\rho_1 = 10$ dB and $\rho_2 = 15$ dB.

Thus, as expected, the error is constant in comparison with the convergence achieved by the optimal *Lin-MMSE*. Apart from this, it should be pointed out that the error of *Lin-MMSE-HH* is over *Lin-MMSE*'s. Oppositely, *Lin-MMSE-HH* achieves a lower error compared with *Lin-MMSE-BD-All-Pass-Filter*, outperforming this suboptimal approach.

4.3 System Diagonalization Based Approaches

In this section, system diagonalization based algorithms will be analysed, creating multiple parallel subchannels per stream from the BS to the end users. Throughout this algorithms the equivalent channel between the BS and the end users becomes diagonal, simplifying the analysis. The main aim of the approaches proposed here is the reduction of the computational cost, which, as it was introduced in the Chapter 3 is excessively high due to the application of an iterative approach.

The use of singular value decomposition is very extended in the literature for the joint design of the transceivers. For example, in point-to-point SISO systems [Berger67], SVD is used to obtain the transmit-receive filters for frequency selective SISO channels, designing jointly the filters for the minimization of the MSE, being an iterative water-filling algorithm applied for the optimum energy distribution. In [Amitay84] the previous solution was extended to 2×2 MIMO channels.

Taking into account the effectiveness and simplicity of the singular value decomposition, it has been applied to single user MIMO relaying systems in [Mo09] [Guan08] [Tseng10] [Zhang11]. For example, in [Mo09] the precoding and the relaying processing matrices are jointly optimized under the MMSE criterion. Due to the non-convexity of the optimization problem, the joint design is not possible. A solution is obtained through primal decomposition [Zschau67], which first performs the optimization over some variables and then over the remaining. In [Guan08], the optimal relay precoder that minimizes the arithmetic MSE is obtained under the assumption that the source precoder is a scaled identity. Finally, SVD decomposition is used for the diagonalization of the MSE considering the direct link in [Tseng10], while [Zhang11] applies a grid search based algorithm to find a global solution for a multicarrier SISO relaying system for the maximization of the mutual information, increasing the complexity.

Singular value decomposition has been also considered for the joint transceiver design in multiuser MIMO AF relaying systems. In [Yu10], the optimal power allocation is studied for multiuser MIMO wireless relay systems, being each node equipped with multiple antennas. In this research work, several algorithms are proposed for both uplink and downlink schemes to compute relaying and precoding matrices. The proposals are highly non-convex, which makes the solution difficult, being finally obtained through the joint gradient search, logarithmic barrier or interior point methods [Boyd04]. Héliot proposes in [Héliot10] three novel power allocation methods for a multiuser MIMO AF downlink relaying system, where the solution is obtained under the maximum weighted sum-rate criterion, taking into account different levels of channel state information (CSI). Finally, the precoding and relaying filters are optimized for sum-capacity maximization in [Xu11], being the non-convex problem made convex through quadratic programming.

Throughout this section, diagonalization approaches will be analysed. Firstly, the system model is introduced, continuing with the design approaches proposed for the mean square error minimization and sum-rate maximization. Due to the effectiveness of singular value decomposition, this technique has been chosen for the joint design of the matrices.

4.3.1 System Diagonalization Model

The system model employed for the analysis based on diagonalization is the one depicted in Figure 3.1. Throughout this analysis, the system described in Section 3.3.3 is considered, being linear precoding applied for user separation while block diagonalization, established at the relay, cancels the interferences between users at the second hop.

With the help of SVD decomposition, the channel from the BS to the end users is diagonalized getting multiple parallel interference-free subchannels. The main aim of the proposed algorithm is the joint design of the precoding and relaying matrices for the optimal power allocation. It is known that in order to obtain a capacity-achieving design, the channel matrix must be diagonal and then water-filling power allocation algorithm can be applied [Mo09][Guan08].

As analysed in Section 3.3.3, thanks to BD [Spencer04b] the interferences created in the second hop can be suppressed fixing the relaying matrix to lie in the null space of the equivalent channel created by the interfering users. Lets define the relaying matrix as $\mathbf{W} = \mathbf{W}_2 \mathbf{W}_1$, being $\mathbf{W}_2 = [\mathbf{W}_{2,1}, \dots, \mathbf{W}_{2,k}, \dots, \mathbf{W}_{2,K}] \in \mathbb{C}^{R \times N}$ the interference canceller matrix which allows the mapping at the null space and $\mathbf{W}_1 = [\mathbf{W}_{1,1}^T, \dots, \mathbf{W}_{1,k}^T, \dots, \mathbf{W}_{1,K}^T]^T \in \mathbb{C}^{N \times R}$ the signal processing matrix, respectively. The matrices $\mathbf{W}_{2,k} \in \mathbb{C}^{R \times N_k}$ and $\mathbf{W}_{1,k} \in \mathbb{C}^{N_k \times R}$ denote the matrices related to user k .

For block diagonalization approach, the constraint $\mathbf{H}_{2,k} \mathbf{W}_{2,j} = \mathbf{0}_{N_k}$ has to be defined for all $j \neq k \quad 1 \leq j, k \leq K$. After carrying out the same procedure introduced in Section 3.3.3, the received signal at each user is represented by (3.19). Additionally, the following mean square error expression is obtained:

$$\begin{aligned}
 \xi &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{s}_k\|_2^2 \right], \\
 &= \sum_{k=1}^K \text{Tr}(\mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{D}_k^H + N_{01} \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{D}_k^H \\
 &+ N_{02} \mathbf{D}_k \mathbf{D}_k^H - 2\Re(\mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F}_k) + \mathbf{I}_{L_k}), \tag{4.6}
 \end{aligned}$$

where N_{01} and N_{02} are the noise spectral power densities of the first and second hops, respectively. Moreover, $\beta_1 = \beta_2 = 1$ is considered due to the presence of power allocation mechanisms.

In order to obtain the equalizers located at each user, it has to be set the partial derivative

of ξ respect to \mathbf{D}_k and must be equalled to zero, obtaining the following expression:

$$\mathbf{D}_k = \mathbf{F}_k^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H (\mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H + N_{01} \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H + N_{02} \mathbf{I}_{N_k})^{-1} \quad (4.7)$$

4.3.2 Optimal Precoder Structure

The main aim on the design entails the diagonalization of the system, getting a more tractable solution. Apart from that, the design of the power allocation matrix has to be carried out. In order to diagonalize the system, the singular value decomposition of the channels is accomplished as

$$\begin{aligned} \mathbf{H}_1 &= \mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^H, \quad \text{and} \\ \mathbf{H}_{e,k} &= \mathbf{U}_{e,k} \mathbf{\Lambda}_{e,k} \mathbf{V}_{e,k}^H, \end{aligned} \quad (4.8)$$

where $\mathbf{U}_1 \in \mathbb{C}^{R \times R}$ and $\mathbf{U}_{e,k} \in \mathbb{C}^{N_k \times N_k}$ are the unitary matrices containing the left singular vectors of the first and second hops' channels, whereas $\mathbf{V}_1 \in \mathbb{C}^{M \times M}$ and $\mathbf{V}_{e,k} \in \mathbb{C}^{N_k \times N_k}$ represent the unitary matrices containing the right singular vectors of the first and second hops, respectively. Additionally, $\mathbf{\Lambda}_1 \in \mathbb{R}^{R \times N}$ and $\mathbf{\Lambda}_{e,k} \in \mathbb{R}^{N_k \times N_k}$ are the diagonal matrices containing the singular values of the first and second hop channels in descending order, being the lowest singular values located at the last positions of the matrices. From now on, the optimal structure will be defined as

$$\begin{aligned} \mathbf{F} &= \mathbf{\Theta} \mathbf{\Omega}, \quad \text{and} \\ \mathbf{W}_{1,k} &= \mathbf{\Xi}_k \mathbf{\Phi}_k \mathbf{\Psi}_k, \end{aligned}$$

being $\mathbf{\Omega} \in \mathbb{C}^{M \times L}$, $\mathbf{\Xi}_k \in \mathbb{C}^{N_k \times L_k}$ and $\mathbf{\Psi}_k \in \mathbb{C}^{L_k \times R}$ unitary matrices while $\mathbf{\Omega} \in \mathbb{R}^{L \times L}$ and $\mathbf{\Phi}_k \in \mathbb{R}^{L_k \times L_k}$ are diagonal matrices containing the power allocation values at the BS and RS. For the optimum diagonalization of the channel from the BS to the end users, we assume

$$\begin{aligned} \mathbf{\Theta} &= \overline{\mathbf{V}}_1, \\ \mathbf{\Xi}_k &= \overline{\mathbf{V}}_{e,k}, \quad \text{and} \\ \mathbf{\Psi}_k &= \overline{\mathbf{U}}_{1,k}^H, \end{aligned}$$

where $\overline{\mathbf{V}}_1 = [\overline{\mathbf{V}}_{1,1}, \dots, \overline{\mathbf{V}}_{1,k}, \dots, \overline{\mathbf{V}}_{1,K}] \in \mathbb{C}^{M \times L}$ takes the first L columns from \mathbf{V}_1 , being $\overline{\mathbf{V}}_{1,k} \in \mathbb{C}^{M \times L_k}$ the columns related to the k^{th} user. In the same way, $\overline{\mathbf{V}}_{e,k}$ takes the first L_k columns from $\mathbf{V}_{e,k}$, while $\overline{\mathbf{U}}_{1,k}$ represents the L_k columns related to the user k grouped into the matrix $\overline{\mathbf{U}}_1 = [\overline{\mathbf{U}}_{1,1}, \dots, \overline{\mathbf{U}}_{1,k}, \dots, \overline{\mathbf{U}}_{1,K}] \in \mathbb{C}^{R \times L}$, which takes the first L columns from \mathbf{U}_1 . In the following sections, the diagonal matrices $\mathbf{\Omega}$ and $\mathbf{\Phi}$ containing the optimal power

allocation values will be computed under different cost functions.

4.3.3 Sum-MSE Minimization

In this section, the optimal power allocation matrices will be determined for mean square error minimization defined in (4.6), which after applying the equalizer defined in (4.7) becomes

$$\begin{aligned} \xi &= \sum_{k=1}^K \text{Tr}(\mathbf{I}_{L_k} - \mathbf{F}_k^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H (\mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H + \mathbf{N}_{0_1} \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \\ &+ \mathbf{N}_{0_2} \mathbf{I}_{N_k})^{-1} \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F}_k). \end{aligned}$$

After the application of the optimal precoder structures, we get

$$\begin{aligned} \xi &= \sum_{k=1}^K \text{Tr}(\mathbf{I}_{L_k} - \mathbf{\Omega}_k^H \bar{\mathbf{\Lambda}}_{1,k}^H \mathbf{\Phi}_k^H \bar{\mathbf{\Lambda}}_{e,k}^H (\bar{\mathbf{\Lambda}}_{e,k} \mathbf{\Phi}_k \bar{\mathbf{\Lambda}}_{1,k} \mathbf{\Omega}_k \mathbf{\Omega}_k^H \bar{\mathbf{\Lambda}}_{1,k}^H \mathbf{\Phi}_k^H \bar{\mathbf{\Lambda}}_{e,k}^H + \mathbf{N}_{0_1} \bar{\mathbf{\Lambda}}_{e,k} \mathbf{\Phi}_k \mathbf{\Phi}_k^H \bar{\mathbf{\Lambda}}_{e,k}^H \\ &+ \mathbf{N}_{0_2} \mathbf{I}_{L_k})^{-1} \bar{\mathbf{\Lambda}}_{e,k} \mathbf{\Phi}_k \bar{\mathbf{\Lambda}}_{1,k} \mathbf{\Omega}_k), \end{aligned}$$

where $\mathbf{\Omega} = \text{blkdiag}(\mathbf{\Omega}_1, \dots, \mathbf{\Omega}_k, \dots, \mathbf{\Omega}_K)$ and $\mathbf{\Phi} = \text{blkdiag}(\mathbf{\Phi}_1, \dots, \mathbf{\Phi}_k, \dots, \mathbf{\Phi}_K)$ represent the $L \times L$ real diagonal matrices containing the optimum power allocation values for the first and second hops, respectively. In addition, $\bar{\mathbf{\Lambda}}_1 = \text{blkdiag}(\bar{\mathbf{\Lambda}}_{1,1}, \dots, \bar{\mathbf{\Lambda}}_{1,k}, \dots, \bar{\mathbf{\Lambda}}_{1,K}) \in \mathbb{R}^{L \times L}$ contains K $L_k \times L_k$ matrices of the first hop singular values in descending order. Due to the Woodbury matrix identity, shown as $(\mathbf{A} + \mathbf{BCE})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{EA}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{EA}^{-1}$, the MSE is reduced to

$$\xi = \sum_{k=1}^K \text{Tr} \left(\left(\mathbf{I}_{L_k} + \mathbf{\Omega}_k^2 \bar{\mathbf{\Lambda}}_{1,k}^2 \mathbf{\Phi}_k^2 \bar{\mathbf{\Lambda}}_{e,k}^2 \left(\mathbf{N}_{0_1} \bar{\mathbf{\Lambda}}_{e,k}^2 \mathbf{\Phi}_k^2 + \mathbf{N}_{0_2} \mathbf{I}_{L_k} \right)^{-1} \right)^{-1} \right).$$

Analysing thoroughly the mean square error defined above, we realized that, after the application of the optimal power structure the MSE function is also diagonalized. Taking into account the definition of the the trace, it can be further simplified to

$$\xi = \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{\mathbf{N}_{0_1} \lambda_{e,k,i}^2 \phi_{k,i}^2 + \mathbf{N}_{0_2}}{\lambda_{e,k,i}^2 \phi_{k,i}^2 (\lambda_{1,k,i}^2 \omega_{k,i}^2 + \mathbf{N}_{0_1}) + \mathbf{N}_{0_2}}, \quad (4.9)$$

where $\lambda_{e,k,i}$, $\phi_{k,i}$, $\lambda_{1,k,i}$ and $\omega_{k,i}$ are the i^{th} diagonal elements of the matrices $\Lambda_{e,k}$, Φ_k , $\Lambda_{1,k}$ and Ω_k , respectively. Therefore, the scalarization of the MSE cost function has been achieved, getting a expression composed by multiple scalars instead of matrices.

The design based on the minimization of the sum-MSE is going to be subject to two power constraints that limit the power at the BS and the relay to E_{BS} and E_R , respectively.

Taking into account the previous expressions, these restrictions can be expressed as:

$$\text{Tr}(\mathbf{F}\mathbf{F}^H) = \sum_{k=1}^K \sum_{i=1}^{L_k} \omega_{k,i}^2 = E_{BS}, \quad (4.10)$$

$$\sum_{k=1}^K \text{Tr}(\mathbf{W}_{1,k} (\mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H + N_{01} \mathbf{I}_R) \mathbf{W}_{1,k}^H) = \sum_{k=1}^K \sum_{i=1}^{L_k} \phi_{k,i}^2 (\lambda_{1,k,i}^2 \omega_{k,i}^2 + N_{01}) = E_R.$$

In order to obtain matrices $\mathbf{\Omega}$ and $\mathbf{\Phi}$ the optimization problem is defined as

$$\begin{aligned} \{\omega_{k,i}, \phi_{k,i}\} &= \underset{\{\omega_{k,i}, \phi_{k,i}\}}{\text{argmin}} \quad \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01} \lambda_{e,k,i}^2 \phi_{k,i}^2 + N_{02}}{\lambda_{e,k,i}^2 \phi_{k,i}^2 (\lambda_{1,k,i}^2 \omega_{k,i}^2 + N_{01}) + N_{02}} \\ \text{s.t.} \quad &\sum_{k=1}^K \sum_{i=1}^{L_k} \omega_{k,i}^2 = E_{BS}, \quad \omega_{k,i} \geq 0, \\ &\sum_{k=1}^K \sum_{i=1}^{L_k} \phi_{k,i}^2 (\lambda_{1,k,i}^2 \omega_{k,i}^2 + N_{01}) = E_R, \quad \phi_{k,i} \geq 0. \end{aligned} \quad (4.11)$$

Since (4.11) is non-convex, the derivation of a global solution is intractable. In order to get a local optimal solution, the numerical method employed in [Fang06] is used. The main reason for this choice is the reduced computational cost in comparison to other solutions like grid search [Zhang11], gradient search, logarithmic barrier or interior point methods [Yu10]. The numerical method used for the optimization considers $\omega_{k,i}$ constant and optimizes $\phi_{k,i}$ directly, and vice versa. So, in order to as to simplify (4.11), the following variable change is carried out:

$$a_{k,i} = \lambda_{1,k,i}^2, \quad b_{k,i} = \lambda_{e,k,i}^2, \quad x_{k,i} = \omega_{k,i}^2, \quad \text{and} \quad y_{k,i} = \phi_{k,i}^2 (\lambda_{1,k,i}^2 \omega_{k,i}^2 + N_{01}). \quad (4.12)$$

After the application of (4.12) into (4.11), the optimization problem is reduced to

$$\begin{aligned} \{x_{k,i}, y_{k,i}\} &= \underset{\{x_{k,i}, y_{k,i}\}}{\text{argmin}} \quad \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} x_{k,i} + N_{01})}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})} \\ \text{s.t.} \quad &\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} = E_{BS}, \quad x_{k,i} \geq 0, \\ &\sum_{k=1}^K \sum_{i=1}^{L_k} y_{k,i} = E_R, \quad y_{k,i} \geq 0. \end{aligned} \quad (4.13)$$

Going into detail in (4.13), it is noteworthy that $x_{k,i}$ and $y_{k,i}$ are symmetric. Moreover, the power constraints defined for the optimization have been decomposed, breaking the

dependence between them. As pointed before, the solution is going to be obtained in an alternating way [Fang06], optimizing $x_{k,i}$ considering $y_{k,i}$ constant and vice versa. Thus, the optimization problem to get $x_{k,i}$ is defined as

$$\begin{aligned} \{x_{k,i}\} &= \underset{\{x_{k,i}\}}{\operatorname{argmin}} \quad \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} x_{k,i} + N_{01})}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})} \\ \text{s.t} \quad & \sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} = E_{BS}, \quad x_{k,i} \geq 0, \end{aligned}$$

while for getting $y_{k,i}$ the optimization problem is set as

$$\begin{aligned} \{y_{k,i}\} &= \underset{\{y_{k,i}\}}{\operatorname{argmin}} \quad \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} x_{k,i} + N_{01})}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})} \\ \text{s.t} \quad & \sum_{k=1}^K \sum_{i=1}^{L_k} y_{k,i} = E_R, \quad y_{k,i} \geq 0. \end{aligned} \quad (4.14)$$

The solutions subject to equality and inequality constraints can be obtained throughout Karush-Kuhn-Tucker (KKT) conditions [Boyd04]. By means of the procedure depicted in Appendix C, the following solution is obtained:

$$x_{k,i} = \frac{1}{a_{k,i}} \left[\sqrt{\frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{\mu_1 (b_{k,i} y_{k,i} + N_{02})}} - N_{01} \right]^+, \quad (4.15)$$

where $[p]^+ = \max(p, 0)$ or in other words, the power allocation will be equal to zero if a negative result is obtained. The solution is obtained through the water-filling algorithm, being μ_1 the water-level that meets E_{BS} at the base station terminal, which is given by

$$\mu_1 = \left[\frac{\sum_{k=1}^K \sum_{i=1}^{L_k} \frac{1}{a_{k,i}} \sqrt{\frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(b_{k,i} y_{k,i} + N_{02})}}}{E_{BS} + \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01}}{a_{k,i}}} \right]^2. \quad (4.16)$$

In the same fashion, $y_{k,i}$ is get, being the optimization problem established in (4.14). After some manipulations and carrying out the procedure described in Appendix C for $x_{k,i}$, the solution is

$$y_{k,i} = \frac{1}{b_{k,i}} \left[\sqrt{\frac{N_{02} a_{k,i} b_{k,i} x_{k,i}}{\mu_2 (a_{k,i} x_{k,i} + N_{01})}} - N_{02} \right]^+, \quad (4.17)$$

being μ_2 the water-level that limits the power at the relay to E_R , getting

$$\mu_2 = \left[\frac{\sum_{k=1}^K \sum_{i=1}^{L_k} \frac{1}{b_{k,i}} \sqrt{\frac{N_{0_2} a_{k,i} b_{k,i} x_{k,i}}{(a_{k,i} x_{k,i} + N_{0_1})}}}{E_R + \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{0_2}}{b_{k,i}}} \right]^2. \quad (4.18)$$

The power allocation matrices are computed through the iterative algorithm described in Algorithm 3. Note that l denotes the current iteration. Apart from this, $\mathbf{X} = \text{blkdiag}(\mathbf{X}_1, \dots, \mathbf{X}_K)$ and $\mathbf{Y} = \text{blkdiag}(\mathbf{Y}_1, \dots, \mathbf{Y}_K)$ are the total power allocation matrices at the BS and RS, respectively.

Algorithm 3 $x_{k,i}$ and $y_{k,i}$ derivation for the MSE minimization.

- 1: Perform block diagonalization at the second hop.
 - 2: Singular value decomposition of \mathbf{H}_1 and \mathbf{H}_e .
 - 3: Initialize the variables to fulfill with: $\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i}^0 = E_{BS}$ and $\sum_{k=1}^K \sum_{i=1}^{L_k} y_{k,i}^0 = E_R$.
 - 4: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 5: Calculate μ_1 as in (4.16) for a fixed $y_{k,i}^l$.
 - 6: Computation of $x_{k,i}^{l+1}$ (4.15).
 - 7: Calculate μ_2 (4.18) for a fixed $x_{k,i}^{l+1}$.
 - 8: Computation of $y_{k,i}^{l+1}$ (4.17).
 - 9: Error calculation: $\epsilon = \|\mathbf{X}^{l+1} - \mathbf{X}^l\|^2 + \|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|^2$.
 - 10: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 11: $\mathbf{X}^l = \mathbf{X}^{l+1}$.
 - 12: $\mathbf{Y}^l = \mathbf{Y}^{l+1}$.
 - 13: **end if**
 - 14: **end while**
 - 15: Compute $\omega_{k,i}$ and $\phi_{k,i}$.
-

A monotonic convergence of $x_{k,i}$ and $y_{k,i}$ is achieved because of the conditional updates of $x_{k,i}$ and $y_{k,i}$, which may either decrease or maintain their values but cannot increase the objective function in (4.9).

It should be underlined that not all the users will receive data, because it may occur that $x_{k,i} < 0$ or $y_{k,i} < 0$. Hence, the power allocation for that subchannel will be equal to zero. In other words, if the power allocation value is zero, it means that the subchannel k, i is over the water-level and cannot be served. SVD decomposition is effective for the diagonalization of the system but the main drawback resides on the impossibility to serve always all the users.

Figure 4.7 shows the percentage of negative $x_{k,i}$ and $y_{k,i}$ values that are obtained in $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ set-ups. When $\rho_2 = 15$ dB, shown in Figure 4.7(a),

at low ρ_1 ($\rho_1 = 0$ dB), the 90 % of channel realizations have at least a negative value. The same conclusion holds for $6 \times 6 \times \{6 \times 1\}$ set-up, being this percentage increased to 100 % of channel realizations. We observe that the negative values decrease to 2-3 % at high ρ_1 for the system $4 \times 4 \times \{4 \times 1\}$, while $6 \times 6 \times \{6 \times 1\}$ set-up presents also a lower percentage, although greater than for the previous set-up.

Additionally, Figure 4.7(b) shows the percentage of negative values that are obtained for fixed $\rho_1 = 15$ dB. We can state that less negative values are obtained at low ρ_2 , but at high ρ_2 the negatives values get the 3 % and 5% for $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$, respectively.

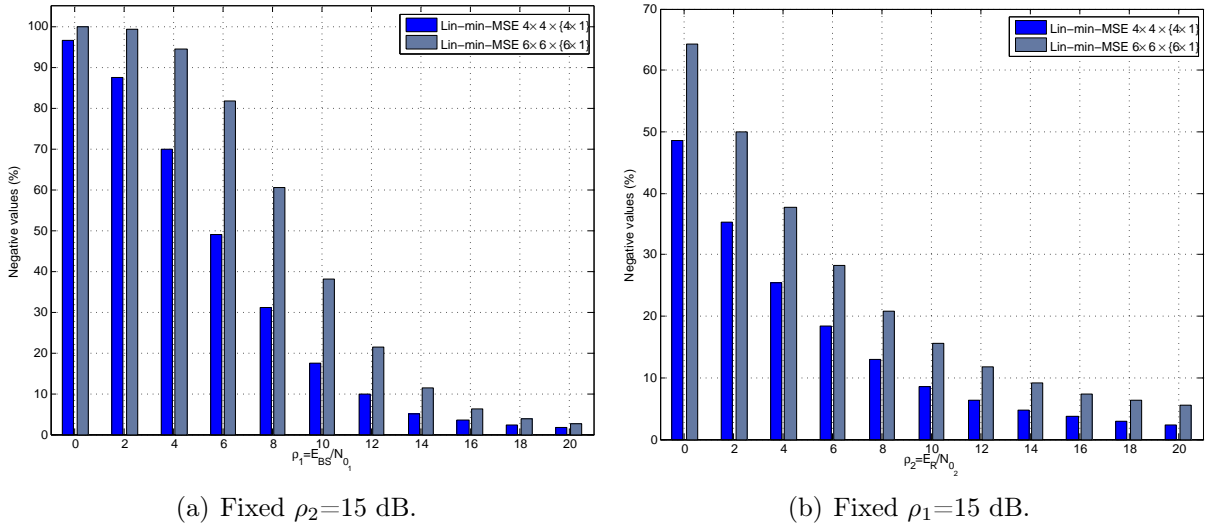


Figure 4.7: Percentage of negative values obtained for the sum-MSE diagonalized system for $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ set-ups (A and B).

After several simulations, we observe that increasing the fixed SNR, ρ_2 in Figure 4.7(a) and ρ_1 in 4.7(b), the number of negatives obtained is reduced considerably. However, the fixed SNR value has to be really high to make all the modes transmit over the water-level.

For the case where water-filling cannot serve all the users, two different design approaches are presented in the following sections, each one with different computational cost. The first resides on the equal power allocation of each subchannel, whereas the second is based on user selection.

4.3.3.1 Equally Distributed Power Allocation

The main aim of this strategy is to give a power allocation value when any of the updates of $x_{k,i}$ or $y_{k,i}$ becomes negative, serving in this way all the users. When a power allocation value is negative, which means that the transmission is over the water-level, the power at the first and second hops will be equally distributed among all the streams as

$$x_{k,i} = \frac{E_{BS}}{L} \quad \text{and} \quad y_{k,i} = \frac{E_R}{L},$$

in order to fulfil with $\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} = E_{BS}$ and $\sum_{k=1}^K \sum_{i=1}^{L_k} y_{k,i} = E_R$, being $x_{k,i}$ and $y_{k,i}$ the diagonal elements of the power allocation matrices $\mathbf{X}_k \in \mathbb{R}^{L_k \times L_k}$ and $\mathbf{Y}_k \in \mathbb{R}^{L_k \times L_k}$ related to each users and grouped into the total power allocation matrices $\mathbf{X} = \text{blkdiag}(\mathbf{X}_1, \dots, \mathbf{X}_K)$ and $\mathbf{Y} = \text{blkdiag}(\mathbf{Y}_1, \dots, \mathbf{Y}_K)$, respectively.

The algorithm of this equally distributed power allocation approach is summarized in Algorithm 4, where if a power allocation value is negative, the algorithm is stopped and the power is distributed equally for all the user streams.

Algorithm 4 Computation of $x_{k,i}$ and $y_{k,i}$ under the MMSE criterion including the equally distribution of the power, when not all users can be served with the optimal approach.

- 1: Perform block diagonalization at the second hop.
 - 2: Computation the singular value decomposition of \mathbf{H}_1 and \mathbf{H}_e
 - 3: Variables initialization, being l the current iteration $\mathbf{X}^0 = \left(\frac{E_{BS}}{L}\right) \mathbf{I}_L$ and $\mathbf{Y}^0 = \left(\frac{E_R}{L}\right) \mathbf{I}_L$.
 - 4: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 5: Calculate μ_1 (4.16) for a fixed \mathbf{Y}^l .
 - 6: Calculate \mathbf{X} (4.15).
 - 7: Calculate μ_2 (4.18) for a fixed \mathbf{X}^{l+1} .
 - 8: Compute \mathbf{Y} (4.17).
 - 9: **if** Any of the diagonal values of \mathbf{X} or \mathbf{Y} are < 0 **then**
 - 10: Go out of the loop while.
 - 11: **else**
 - 12: Error calculation: $\epsilon = \|\mathbf{X}^{l+1} - \mathbf{X}^l\|^2 + \|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|^2$.
 - 13: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 14: $\mathbf{X}^l = \mathbf{X}^{l+1}$.
 - 15: $\mathbf{Y}^l = \mathbf{Y}^{l+1}$.
 - 16: **end if**
 - 17: **end if**
 - 18: **end while**
 - 19: **if** The loop has been broken for the negative value of $x_{k,i}$ or $y_{k,i}$ **then**
 - 20: $\mathbf{X} = \left(\frac{E_{BS}}{L}\right) \mathbf{I}_L$.
 - 21: $\mathbf{Y} = \left(\frac{E_{BS}}{L}\right) \mathbf{I}_L$.
 - 22: **end if**
 - 23: Compute Ω and Φ .
-

As it will be seen in the following section, the equally distributed power allocation algorithm ensures the transmission of data to all the users. However, the performance obtained is degraded in comparison to the user selection algorithms. Nevertheless, this solution presents a lower cost in complexity terms.

4.3.3.2 Sum-MSE Minimization with User Selection

When not all the users need to be served, user selection algorithms can be used. Firstly, it should be underlined that a brute-force search over all possible user sets guarantees that the

total MSE is minimized.

The complexity, however, is prohibitive if the number of users is large. For example, if K_{max} is the maximum number of users that can be served simultaneously and K is the total number of users, the complete search of the optimal user set has a combinatoric complexity because every k ($1 \leq k \leq K_{max}$) out of K users must be searched. It is noteworthy that the exhaustive search method needs to consider $O(K^{K_{max}})$ possible user sets. For simplicity, we will assume that $K_{max} = \min(M, R)/N_k$, considering the same number of antennas at each user.

In order to reduce the complexity of the brute-force algorithm, an alternative algorithm, described in Algorithm 5, is here proposed that simplifies the search. Following this approach, the users that minimize the error most are chosen first. After that, from the remaining unselected users, the user that provides the lowest MSE in combination with the already selected user is chosen.

4.3.4 Sum-Rate Maximization Based Approach

Applying the system diagonalization approach to the sum-rate function depicted in [Wang05], the achievable sum-rate expression becomes

$$C = \frac{1}{2} \sum_{k=1}^K \log_2 \left(|\mathbf{I}_{L_k} + \mathbf{F}_k^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{R}_{\mathbf{n}_k}^{-1} \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{F}_k| \right),$$

where $\mathbf{R}_{\mathbf{n}_k} = (N_{0_1} \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H + N_{0_2} \mathbf{I}_{N_k})$. The term $\frac{1}{2}$ is included due to the half-duplex transmission. After the singular value decomposition of matrices \mathbf{H}_1 and \mathbf{H}_e , the achievable sum-rate expression is reduced to

$$C = \frac{1}{2} \sum_{k=1}^K \log_2 \left| \mathbf{I}_{L_k} + \Omega_k^2 \bar{\mathbf{\Lambda}}_{1,k}^2 \Phi_k^2 \bar{\mathbf{\Lambda}}_{e,k}^2 \left(N_{0_1} \bar{\mathbf{\Lambda}}_{e,k}^2 \Phi_k^2 + N_{0_2} \mathbf{I}_{N_k} \right)^{-1} \right|, \quad (4.19)$$

being $\bar{\mathbf{\Lambda}}_{1,k}^2$ and $\bar{\mathbf{\Lambda}}_{e,k}^2$ previously defined in (4.8). Analysing (4.19) we realize that all the matrices are related to each user k and that the problem has been diagonalized. Comparing (4.19) with (4.9), we perceive that the capacity and the MSE are related by means of this expression:

$$C = \frac{1}{2} \sum_{k=1}^K \log_2 |\mathbf{R}_{e,k}^{-1}|,$$

where $\mathbf{R}_{e,k} \in \mathbb{C}^{L_k \times L_k}$ is the error covariance matrix which takes part into the MSE definition as $\xi = \sum_{k=1}^K \text{Tr}(\mathbf{R}_{e,k})$. Since it is known that the determinant of a diagonal matrix is the

Algorithm 5 User selection algorithm for sum-MSE minimization.

```

1: if The first user has to be selected then
2:   Users subset definition  $U = \{1, 2, \dots, K\}$ .
3:   Initialization of  $\xi = \infty$ .
4:   for  $k = 1 \dots K$  do
5:     Calculate power allocation parameters as in Algorithm 3.
6:     Compute the sum-MSE defined in (4.9).
7:     if  $\xi^{l+1} < \xi^l$  then
8:        $U_{\text{selected}} = k$ .
9:        $\xi^l = \xi^{l+1}$ .
10:    end if
11:  end for
12:  Remove the user from the subset:  $U = U - U_{\text{selected}}$ ,  $K = K - 1$  and  $K_{\text{max}} = K_{\text{max}} - 1$ .
13: else
14:  Create a temporal subset:  $U_{\text{temp}} = U_{\text{selected}}$ .
15:  for  $kk = 1, \dots, K_{\text{max}}$  do
16:    Initialization of  $\xi = \infty$ .
17:    for  $k = 1 \dots K$  do
18:      Select the temporal subset adding the user  $k$  to the selected subset:  $U_{\text{temp}} =$ 
         $U_{\text{selected}} + k$ .
19:      Calculate power allocation parameters as in Algorithm 3.
20:      if Any of the power allocation values are negatives then
21:        Discard the  $k^{\text{th}}$  user.
22:      else
23:        Compute the sum-MSE defined in (4.9).
24:        if  $\xi^{l+1} < \xi^l$  then
25:           $U_{\text{selected}} = k$ .
26:           $\xi^l = \xi^{l+1}$ .
27:        end if
28:      end if
29:    end for
30:    Remove the user from the subset:  $U = U - U_{\text{selected}}$ .
31:  end for
32: end if
33: Compute  $\Omega$  and  $\Phi$ .
    
```

product of the diagonal elements, the sum-rate expression is simplified to

$$C = \frac{1}{2} \sum_{k=1}^K \log_2 \left(\prod_{i=1}^{L_k} 1 + \frac{\omega_{k,i}^2 \lambda_{k,i}^2 \phi_{k,i}^2 \lambda_{e,k,i}^2}{N_{01} \lambda_{e,k,i}^2 \phi_{k,i}^2 + N_{02}} \right).$$

Taking into account logarithm properties, this expression can be rewritten as:

$$C = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{L_k} \log_2 \left(1 + \frac{\omega_{k,i}^2 \lambda_{1,k,i}^2 \phi_{k,i}^2 \lambda_{e,k,i}^2}{N_{01} \lambda_{e,k,i}^2 \phi_{k,i}^2 + N_{02}} \right),$$

where $\omega_{k,i}$ and $\phi_{k,i}$ are the power allocation values for user k and stream i , whereas $\lambda_{1,k,i}$ and $\lambda_{e,k,i}$ are the i^{th} singular values related to the k^{th} user.

The aim of this approach is the joint design of the power allocation matrices for sum-rate maximization at both BS and RS terminals subject to two power constraints:

$$\begin{aligned} \{\omega_{k,i}, \phi_{k,i}\} &= \max_{\{\omega_{k,i}, \phi_{k,i}\}} \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{L_k} \log_2 \left(1 + \frac{\omega_{k,i}^2 \lambda_{1,k,i}^2 \phi_{k,i}^2 \lambda_{e,k,i}^2}{N_{01} \lambda_{e,k,i}^2 \phi_{k,i}^2 + N_{02}} \right) \\ \text{s.t.} \quad &\sum_{k=1}^K \sum_{i=1}^{L_k} \omega_{k,i}^2 = E_{BS}, \quad \omega_{k,i} \geq 0, \\ &\sum_{k=1}^K \sum_{i=1}^{L_k} \phi_{k,i}^2 = E_R, \quad \phi_{k,i} \geq 0. \end{aligned} \quad (4.20)$$

Once again, the optimization problem (4.20) is non-convex. As it happens with the optimization based on the minimum mean square error, if a fixed $\omega_{k,i}$ is assumed, the problem is convex with respect to $\phi_{k,i}$ and vice versa. Carrying out the same procedure used for the min-MSE analysis, the solution for power allocation at the BS is

$$x_{k,i} = \frac{N_{01}}{2a_{k,i}} \left[\sqrt{\frac{b_{k,i}^2 y_{k,i}^2}{N_{02}^2} + \frac{2b_{k,i} a_{k,i} y_{k,i} \mu_1^*}{\ln(2) N_{01} N_{02}}} - \left(\frac{b_{k,i} y_{k,i}}{N_{02}} + 2 \right) \right]^+.$$

The water-level $\mu_1^* = 1/\mu_1$ assures that the transmission power at the BS does not exceed E_{BS} , given by

$$\sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01}}{2a_{k,i}} \left[\sqrt{\frac{b_{k,i}^2 y_{k,i}^2}{N_{02}^2} + \frac{2b_{k,i} a_{k,i} y_{k,i} \mu_1^*}{\ln(2) N_{01} N_{02}}} - \left(\frac{b_{k,i} y_{k,i}}{N_{02}} + 2 \right) \right]^+ - E_{BS} = 0.$$

This non-linear equation can be solved by means of the bisection method [Boyd04], which is a root finding method that has been summarized at Appendix D (see Algorithm 16).

In the same fashion, in order to get the power allocation matrix Φ for the relay, the optimization problem is defined as

$$\begin{aligned} \{y_{k,i}\} &= \max_{\{y_{k,i}\}} \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{L_k} \log_2 \left(1 + \frac{\omega_{k,i}^2 \lambda_{1,k,i}^2 \phi_{k,i}^2 \lambda_{e,k,i}^2}{N_{01} \lambda_{e,k,i}^2 \phi_{k,i}^2 + N_{02}} \right) \\ \text{s.t.} \quad &\sum_{k=1}^K \sum_{i=1}^{L_k} y_{k,i} = E_R \quad y_{k,i} \geq 0. \end{aligned} \quad (4.21a)$$

Due to the convexity of the problem and after carrying out the steps described in Ap-

pendix D, we obtain

$$y_{k,i} = \frac{N_{02}}{2b_{k,i}} \left[\sqrt{\frac{a_{k,i}^2 x_{k,i}^2}{N_{01}^2} + \frac{2\mu_2^* a_{k,i} b_{k,i} x_{k,i}}{\ln(2) N_{01} N_{02}}} - \left(\frac{a_{k,i} x_{k,i}}{N_{01}} + 2 \right) \right]^+,$$

which can be computed by means of the water-filling algorithm, being $\mu_2^* = 1/\mu_2$ the water level that fulfils with the power constraint in (4.21a). Considering $\sum_{k=1}^K \sum_{i=1}^{L_k} y_{k,i} - E_R = 0$, the water-level μ_2^* can be obtained through:

$$\sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{02}}{2b_{k,i}} \left[\sqrt{\frac{a_{k,i}^2 x_{k,i}^2}{N_{01}^2} + \frac{2\mu_2^* a_{k,i} b_{k,i} x_{k,i}}{\ln(2) N_{01} N_{02}}} - \left(\frac{a_{k,i} x_{k,i}}{N_{01}} + 2 \right) \right]^+ - E_R = 0.$$

This equation can again be solved efficiently by means of the bisection method, described in Algorithm 16 of Appendix D.

The optimization under the sum-rate maximization criterion shows the same disadvantage as for the min-MSE case, where due to the weakest values of certain singular values, the power allocation value can sometimes be negative.

For example, Figure 4.8(a) shows the percentage of negative values obtained for a concrete number of channel realization when the $\rho_2=15$ dB. We observe that when $\rho_1=0$ dB, the 80 % of channel realizations are negative for $4 \times 4 \times \{4 \times 1\}$, while this quantity reach the 95 % when the set-up changes to $6 \times 6 \times \{6 \times 1\}$. For the latter, the negative values decrease until $\rho_1=12$ dB, being the negative value from 14 to 20 dB the 60 %. It happens the same for $4 \times 4 \times \{4 \times 1\}$ system, being 30 % the minimum negative percentage achieved.

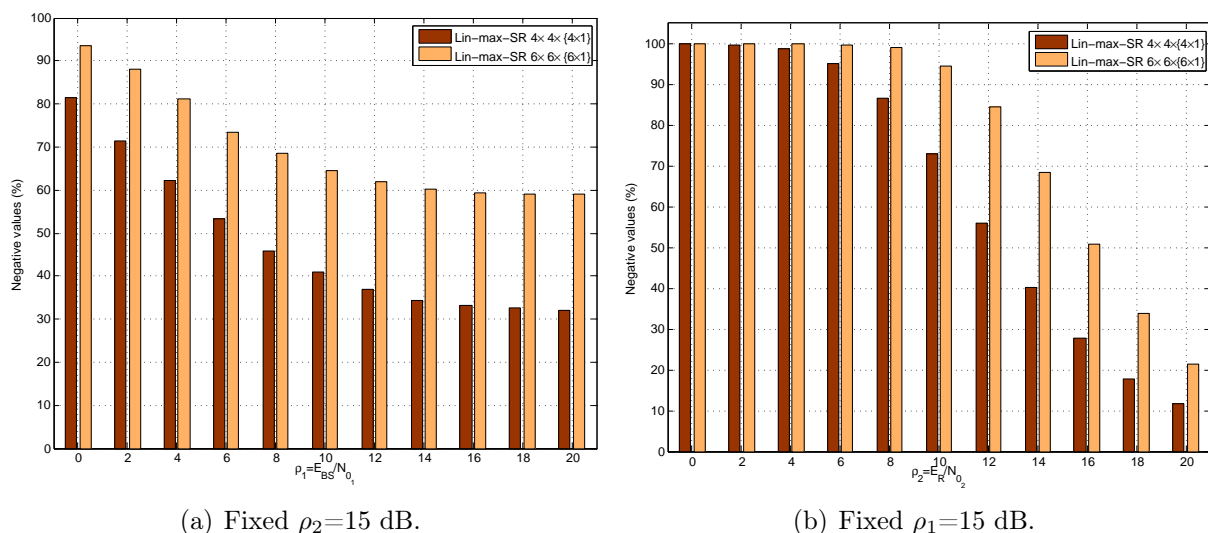


Figure 4.8: Percentage of negative values obtained for the sum-rate maximization system for $4 \times 4 \times \{4 \times 1\}$ and $6 \times 6 \times \{6 \times 1\}$ set-ups (A and B).

In the same way, for a fixed $\rho_1=15$ dB, for low ρ_2 (0-4 dB), the set-up $4 \times 4 \times \{4 \times 1\}$

gets 100 % of negatives values, whereas in $6 \times 6 \times \{6 \times 1\}$ this value is reached from 0 to 8 dB. For mid and high SNR ranges, the negative ratio decreases in both set-ups.

In order to serve all the users considered in the model, two algorithms are proposed here: the first one based on equally power distribution and the second one based on user selection.

4.3.4.1 Equally Distributed Power Allocation Approach

Basically the algorithm is described in Section 4.3.3.1, where the aim of this equally distributed power allocation proposal is to fulfil with

$$x_{k,i} = \frac{E_{BS}}{L} \quad \text{and} \quad y_{k,i} = \frac{E_R}{L}.$$

As pointed out before, if a negative value is obtained for power allocation, the loop is forced to stop, distributing the power equally among all the streams as depicted in Algorithm 4.

4.3.4.2 Sum-Rate Maximization Based User Selection Algorithm

Sum-rate maximization metric is a good choice for user selection as it can be seen in [Shen06]. Instead of choosing the users who minimize the mean square error most, the algorithm described in this section selects the users that improve the sum-rate jointly. In few words, we will choose the subset of users that achieves the maximum sum-rate.

Following the same strategy as described in Section 4.3.3.2 for MSE minimization, the user who achieves the best sum-rate is first selected. After this, we choose a second user, which in combination with the first one selected achieves the highest sum-rate. This procedure is executed until a number of K_{max} users are selected.

We assume that $N_{max} = \min(M, R)$ and $K_{max} = \frac{N_{max}}{N_k}$, considering the same number of receive antennas at each user. Before continuing, it should be taken into account that the power allocation matrices are grouped into the matrices $\mathbf{X} = \text{blkdiag}(\mathbf{X}_1, \dots, \mathbf{X}_K) \in \mathbb{R}^{L \times L}$ and $\mathbf{Y} = \text{blkdiag}(\mathbf{Y}_1, \dots, \mathbf{Y}_K) \in \mathbb{R}^{L \times L}$, being $\mathbf{X}_k = \text{diag}(x_{k,1}, \dots, x_{k,L_k}) \in \mathbb{R}^{L_k \times L_k}$ and $\mathbf{Y}_k = \text{diag}(y_{k,1}, \dots, y_{k,L_k}) \in \mathbb{R}^{L_k \times L_k}$ the power allocation matrices related to the user k .

Algorithm 5 sums-up the user selection algorithm based on sum-rate maximization, being the cost function change to the maximization of the sum-rate.

4.3.5 Simulation Results

In this section, the results obtained for the approaches based on system diagonalization will be presented, being the simulation parameters the ones defined in the Section 4.1.3. Simulations, accomplished for the set-up $4 \times 4 \times \{4 \times 1\}$ (set-up A) will compare the local

optimal *Lin-MMSE*, the equally power distribution solution for MSE minimization (*Lin-min-MSE-Naive*) and sum-rate maximization (*Lin-max-SR-Naive*), as well as user selection approaches for both metrics, called *Lin-min-MSE* and *Lin-max-SR*.

Before starting with the analysis of the results, the subset of user should be specified for user selection approaches when sum-MSE and sum-rate are minimized and maximized, respectively. In order to choose the number of users, we select subsets of 8, 12 and 16 users.

Figure 4.9 depicts the performance of the aforementioned algorithms for the set-up $4 \times 4 \times \{4 \times 1\}$ when $\rho_2=15$ dB. As it can be observed in the figure, *Lin-min-MSE-Naive* is the system that distributes the power equally when the optimal allocation cannot attend $K = 4$ users, while *Lin-min-MSE* is the name used for the user selection algorithm which ensures that users are always served.

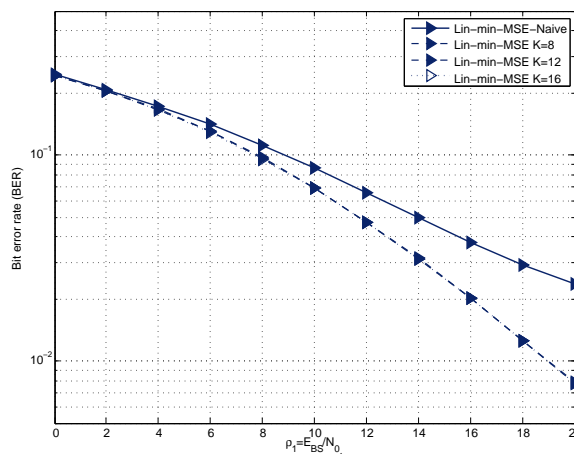


Figure 4.9: BER curves of sum-MSE diagonalized approaches for QPSK modulation and $4 \times 4 \times \{4 \times 1\}$ set-up (A) for different sets of users.

We observe that the performance is more or less the same when four users are selected from a subset of 8, 12 and 16 users, being the maximum gain that can be achieved over *Lin-min-MSE-Naive* of about 3 dB. Furthermore, we expect that increasing the number of users that form the subset the performance might be better.

In addition, Figure 4.10 shows the sum-rate performance of the equally distributed power allocation approach for the sum-rate maximization (*Lin-max-SR-Naive*) and the user selection algorithm based on sum-rate (SR) maximization for a subset of $K = 8$ (*Lin-max-SR K=8*), $K = 12$ (*Lin-max-SR K=12*) and $K = 16$ (*Lin-max-SR K=16*) for the set-up $4 \times 4 \times \{4 \times 1\}$ when $\rho_2 = 15$ dB.

Obviously and as expected, the naive solution is outperformed by the user selection approaches. However, for user selection approach, the difference gap in terms of BER performance between a subset to the other is low, whereas in sum-rate terms, the difference is significant.

It should be pointed out that a compromise between complexity and performance has

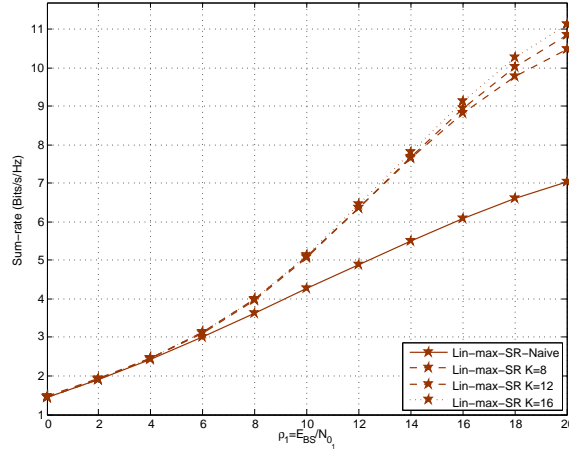


Figure 4.10: Sum-rate versus ρ_1 of diagonalization based approach for $4 \times 4 \times \{4 \times 1\}$ set-up (A) and $\rho_2 = 15$ dB for different sets of users.

to be taken. Thus, if the number of users that forms the subset is high, the complexity is also high. In contrast, if a reduced number of users are selected, the performance will not be near the optimal, but the complexity will be lower.

For now on, the simulations presented will be tested for a subset of $K = 12$ users, being the proposals renamed as *Lin-min-MSE K=12* and *Lin-max-SR K=12* for MSE minimization and SR maximization, respectively.

To conclude, it should be underlined that the comparison between the equally power distributed approaches and user selection approaches is not fair due to the diversity imposed by the latter on user selection. Hence, the best users are selected from a subset introducing a gain on performance.

4.3.5.1 BER Performance

Figure 4.11 depicts the BER performance for the set-up A when $\rho_1 = 15$ dB (Figure 4.11(b)) and $\rho_2 = 15$ dB (Figure 4.11(a)).

As expected, *Lin-min-MSE-Naive*, designed for the minimization of the mean square error, outperforms *Lin-max-SR-Naive*, which targets the sum-rate maximization. When $\rho_2 = 15$ dB, from 6 dB on *Lin-min-MSE-Naive* outperforms *Lin-max-SR-Naive* reaching a maximum gain of 4 dB. For the approaches based on user selection, it should be underlined that *Lin-min-MSE K=12* improves significantly the solution that distributes equally the power and also the solution based on the maximization of the sum-rate. Furthermore, *Lin-min-MSE K=12* performs near *Lin-MMSE*, being the gain obtained by *Lin-MMSE* over *Lin-min-MSE K=12* of 2 dB. From 17 dB on, the situation changes, being *Lin-MMSE* outperformed by *Lin-min-MSE K=12* due to the diversity provided by user selection. The same conclusions are obtained for $\rho_1 = 15$ dB depicted in the Figure 4.11(b) being *Lin-MMSE* outperformed by *Lin-min-MSE K=12* from $\rho_2 = 4$ dB to $\rho_2 = 14$ dB.

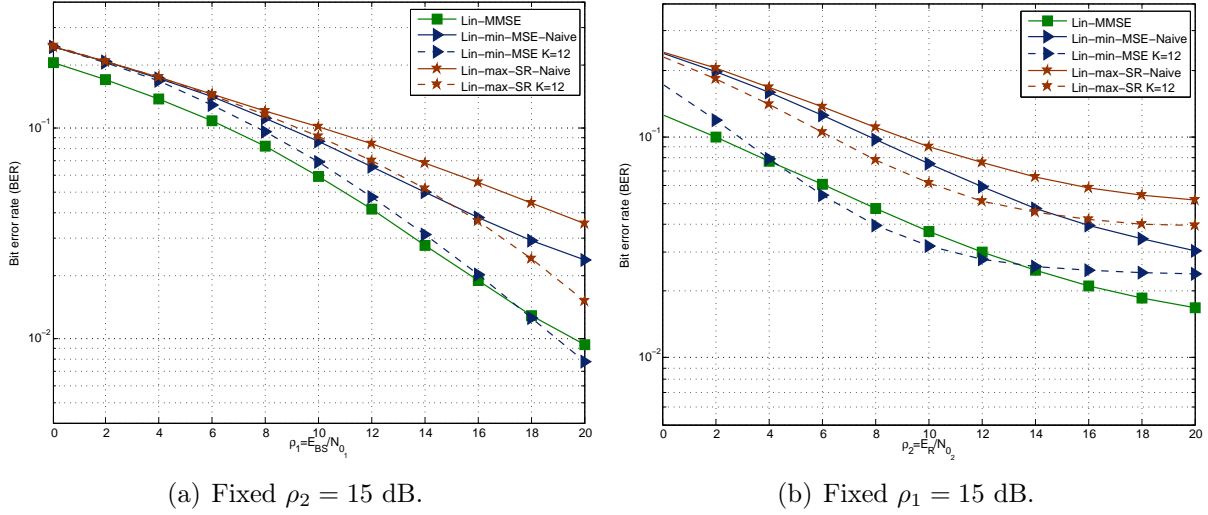


Figure 4.11: BER performance comparison for diagonalization based approaches with QPSK modulation for $4 \times 4 \times \{4 \times 1\}$ set-up (A).

4.3.5.2 Achievable Sum-Rate

In this section, the achievable sum-rate of the aforementioned systems is analysed. Figure 4.12 shows the achievable sum-rate when $\rho_1 = 15$ (Figure 4.12(b)) and $\rho_2 = 15$ (Figure 4.12(a)) for the set-up A ($4 \times 4 \times \{4 \times 1\}$). When $\rho_2 = 15$ dB, *Lin-max-SR K=12* outperforms all the systems based on diagonalization. As expected, naive or equally distributed power allocation algorithms obtain worst performance compared to user selection algorithms. Furthermore, as expected the solution based on sum-rate maximization achieve better capacity in comparison with the ones that minimize the MSE. What is more, *Lin-max-SR K=12* and *Lin-min-MSE K=12* achieve better sum-rate than *Lin-MMSE* from $\rho_1 = 12$ on.

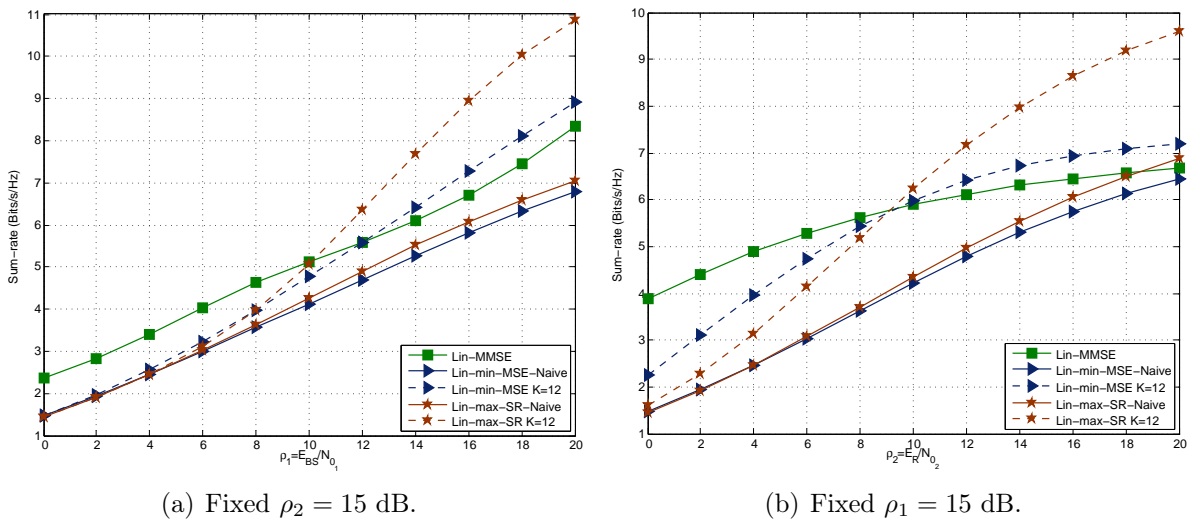


Figure 4.12: Achievable sum-rate for diagonalization based approaches for $4 \times 4 \times \{4 \times 1\}$ set-up (A).

The same conclusions hold when $\rho_1 = 15$ as depicted in Figure 4.12(b), being naive solutions outperformed by *Lin-min-MSE* $K=12$ and *Lin-max-SR* $K=12$. Until $\rho_2 = 9$ dB *Lin-min-MSE* $K=12$ performs better than *Lin-max-SR* $K=12$, but when $\rho_2 > 9$ the results show the opposite being the difference significant. For this configuration set-up, *Lin-max-SR* $K=12$ obtains a significant improvement respect to the rest of the systems, specially at high ρ_2 .

4.4 Computational Complexity

As pointed out at the beginning of this chapter, the main aim of the proposed suboptimal algorithms is the complexity reduction, in comparison to the iterative and BD approaches of Chapter 3.

Three different measurements associated with complexity will be provided in this section. Firstly, the number of FLOPs will be measured, after which, the complexity order will be given. Finally, to conclude with the analysis, algorithms' run time will be shown.

For simplicity, specific number of iterations for the iterative loop and the bisection method are assumed, called it and it_{bisec} , respectively. Provided simulations will prove the effectiveness of the suboptimal algorithms in terms of complexity.

4.4.1 Complexity Analysis for the BD All-Pass Filter Design

This section analyses the computational cost related to the proposal called *Lin-MMSE-BD-All-Pass-Filter* introduced in Section 4.1. Table 4.1 shows the amount of FLOPs that the algorithm executes for the computation of the precoding (\mathbf{F}) and relaying (\mathbf{W}) matrices.

| Algorithm | Operation | FLOPs |
|-----------------------------|--------------|---|
| Lin-MMSE | \mathbf{F} | $4M^3 + (6R+8)M^2 + (2N+6M+3)R^2 + (4R+3)M + (2R+1)MN + 1$ |
| | \mathbf{W} | $10R^3 + (19+4N+4M)R^2 + (6N+2)MR + 2NR + 1$ |
| Lin-MMSE-BD | \mathbf{F} | $2N^3 + 4M^3 + (2+M)N^2 + (2R+4N+7)M^2 + (4R+4)NM + (2R+3)M$ |
| | \mathbf{W} | $4R^3 + (2N+2M+8)R^2 + 10N^3 + (10+2M)N^2 + (3+6M)NR + (3+M)N + (3+M)R + 1$ |
| Lin-MMSE-BD-All-Pass-Filter | \mathbf{F} | $4N^3 + (4N+7)M^2 + (3+N)M$ |
| | \mathbf{W} | R^2 |

Table 4.1: Number of FLOPs for *Lin-MMSE-BD-All-Pass-Filter*.

Assuming $M = R = N = K = n$, the number of FLOPs carried out by the naive BD algorithm is summed up and compared with *Lin-MMSE* and *Lin-MMSE-BD* in Table 4.2, where apart from giving a general expression, the case $4 \times 4 \times \{4 \times 1\}$ has been analysed.

| Algorithm | Comp. order | FLOPs | |
|------------------------------------|--------------|--------------------------------|---------------------|
| | | M=R=N=n | M=R=N=4 |
| Lin-MMSE | $O(it\ n^3)$ | $it(52n^3 + 54n^2 + 5n + 16)$ | $4 \times 10^3\ it$ |
| Lin-MMSE-BD | $O(it\ n^3)$ | $it(71n^3 + 26n^2 + 13n + 10)$ | $5 \times 10^3\ it$ |
| Lin-MMSE-BD-All-Pass-Filter | $O(n^3)$ | $(32n^3 + 24n^2 + 1)$ | 2×10^3 |

Table 4.2: Complexity order and number of FLOPs when $n = 4$ for *Lin-MMSE-BD-All-Pass-Filter*.

As we can observe in Table 4.2, the number of operations required is lower for *Lin-MMSE-All-Pass-Filter* despite the block diagonalization process that executes. The complexity order for *Lin-MMSE* and *Lin-MMSE-BD* is cubic and depends on the number of iterations, whereas *Lin-MMSE-All-Pass-Filter* has only a cubic complexity cost.

In the same fashion, Figure 4.13 shows the time needed for the computation of the algorithm for set-up A ($4 \times 4 \times \{4 \times 1\}$). Obviously, in comparison to the joint linear approaches, *Lin-MMSE-All-Pass-Filter*'s run-time is negligible. For example, for each Monte Carlo simulation, 0.002 seconds are needed for set-up A.

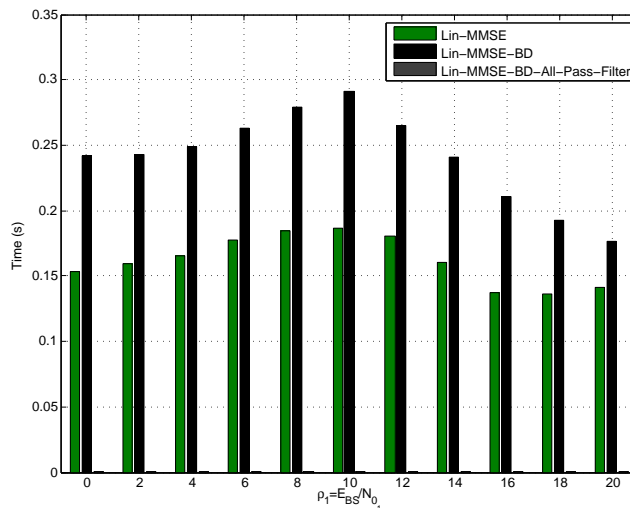


Figure 4.13: Run-time for *Lin-MMSE-All-Pass-Filter* for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with $\rho_2 = 15$ dB.

4.4.2 Complexity Analysis for Independent Hop-by-Hop Approach

The purpose of this section is to analyse the computational complexity of the *Lin-MMSE-HH* technique introduced in the Section 4.2, where \mathbf{F} and \mathbf{W} are obtained by means of the minimization of each hop independently.

Table 4.3 shows the number of operations, where the highest complexity of this algorithm resides on the computation of the pseudo-inverse matrices.

| Algorithm | Operation | FLOPs |
|-------------|--------------|--|
| Lin-MMSE-HH | \mathbf{F} | $2N^3 + 4M^3 + (4R + 8)M^2 + (3 + R)M + 2$ |
| | \mathbf{W} | $4R^3 + (4N + 8)R^2 + (3 + N)R + 2$ |

Table 4.3: Number of FLOPs for *Lin-MMSE-HH*.

Assuming $N = M = R = n$, as it can be seen at the Table 4.4, the expression is reduced being the complexity order $O(n^3)$ due to the inverse of the matrices. Obviously and as expected, this algorithm executes a lower number of operations compared with the suboptimal approach *Lin-MMSE-BD-All-Pass-Filter*.

Running-time, depicted in Figure 4.14 and measured as the addition of computation times at three terminals, is only shown for *Lin-MMSE-BD-All-Pass-Filter* and *Lin-MMSE-HH* in order to show the reduction of the computational cost. Both approaches, which do not need any kind of iterative process, do not require more than 5×10^{-4} seconds per Monte Carlo simulation.

4.4.3 Complexity Analysis for System Diagonalization Based Approaches

This section compares the computational cost of the system diagonalization based approaches named *Lin-min-MSE-Naive*, *Lin-min-MSE K=12*, *Lin-max-SR-Naive* and *Lin-max-SR K=12* introduced in Section 4.3.

| Algorithm | Comp. order | FLOPs | |
|-----------------------------|-------------|----------------------------|-----------------|
| | | M=R=N=n | M=R=N=4 |
| Lin-MMSE-BD-All-Pass-Filter | $O(n^3)$ | $(32n^3 + 24n^2 + 1)$ | 2×10^3 |
| Lin-MMSE-HH | $O(n^3)$ | $(16n^3 + 18n^2 + 6n + 4)$ | 1×10^3 |

Table 4.4: Number of FLOPs when $n = 4$ and complexity order for *Lin-MMSE-HH* approach.

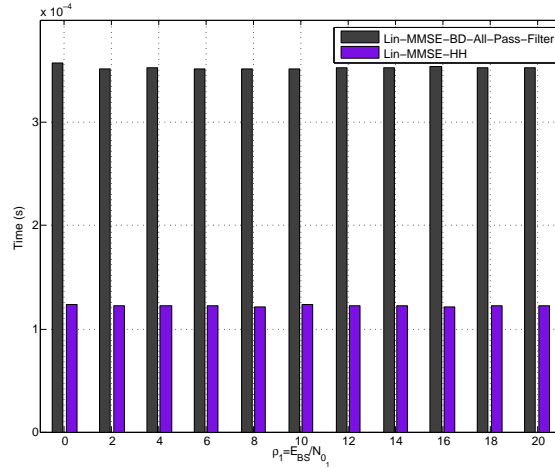


Figure 4.14: Execution-time for *Lin-MMSE-HH* for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with $\rho_2 = 15$ dB.

| Algorithm | Comp. order | FLOPs | |
|--------------------------|---------------------|--|--------------------------------------|
| | | M=R=N=n | M=R=N=4 |
| Lin-min-MSE-Naive | $O(it n^2)$ | $it (8n+7+n^2)$ | 55 it |
| Lin-min-MSE K=12 | $O(it K_{max} n^2)$ | $it K_{max} (9n^2+11n)$ | 188 $K_{max} it$ |
| Lin-max-SR-Naive | $O(it n^2)$ | $it (20n+17+it_{bisecc} (4n+7)+n^2)$ | (113 + 31 $iter_{bisecc}$) it |
| Lin-min-MSE K=12 | $O(it K_{max} n^2)$ | $K_{max} it (20n+17+it_{bisecc} (4n+7)+n^2)$ | $K_{max} (113 + 31 iter_{bisecc})it$ |

Table 4.5: Complexity order for *Lin-min-MSE-Naive*, *Lin-min-MSE K=12*, *Lin-max-SR-Naive* and *Lin-max-SR K=12* approaches.

Unsurprisingly, the number of FLOPs obtained for the computation of \mathbf{F} and \mathbf{W} is the same for all the proposals, needing ML and RL FLOPs for \mathbf{F} and \mathbf{W} computation.

For user selection algorithm, $N = M = R = n$ is assumed. Furthermore, taking into account that $L = \min(N, M, R)$, the number of operations is reduced obtaining a expression which depends on n , $iter$ and K_{max} as can be viewed at Table 4.5. $iter_{bisecc}$, which denotes the run time of the bisection method will add complexity to *Lin-max-SR-K* and *Lin-max-SR-K* proposals. It should be underlined that the complexity order of user selection algorithms will depend on the number of iterations performed and the users that form the subset denoted K_{max} . If $iter$ is considered fixed the complexity will be quadratic due to the scalarization of the problem.

As expected and can be seen in Figure 4.15, *Lin-min-MSE-Naive* and *Lin-max-SR-Naive*

present the lower run time being lower than 0.01 seconds. Furthermore the run time of *Lin-max-SR-K* increases gradually being the minimum 0.07 seconds and the maximum 0.12 seconds. Due to the *scalarization* of the system the complexity of the previous proposals is reduced considerably. Despite the low run-time, this is not comparable with the one obtained in the execution of the other suboptimal approaches, which do not reach 5×10^{-4} seconds.

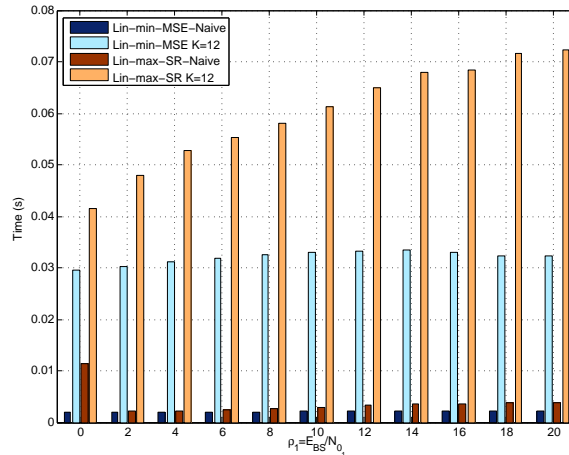


Figure 4.15: Running-time for system diagonalization based algorithms for $4 \times 4 \times \{4 \times 1\}$ set-up (A) with $\rho_2 = 15$ dB.

4.4.4 Complexity Performance Trade-Off

Throughout this chapter, some suboptimal algorithms have been proposed with the aim of minimizing the complexity of the joint linear optimal design. Furthermore, by means of simulations, we have showed that the performance of this suboptimal algorithms was not near the optimal's. Nevertheless, a complete complexity analysis has showed that this suboptimal approaches have fulfilled beyond expectations, reducing the cost considerably. Therefore, a trade-off between performance and complexity has been showed necessary for the design.

Suboptimal approaches include the ones that diagonalized the channel. Moreover the computational cost has been considerably reduced as a consequence of the scalarization. Thus, the optimization problem, composed by matrices, has been converted to scalars. Even so, this approach still iterative due to the dependence of the scalars.

On the other hand, *Lin-MMSE-BD-All-Pass-Filter* has reduced the number of floating point operations from 5×10^7 to 2×10^3 , being the number of operations carried out by the iterative algorithm. Obviously, this suboptimal approach has avoided the iterative search, obtaining the matrices directly. Furthermore, although the poor results on BER performance, at high SNR this suboptimal approach obtains a sum-rate near the local optimal.

Finally, the same has happened with *Lin-MMSE-HH*, avoiding the use of an iterative

process due to the interdependence of the matrices and requiring only 1×10^3 FLOPs. Thanks to this direct search the precoding and relaying matrices can be searched in 1.25×10^{-4} seconds whilst this time increases to 0.2 seconds for joint iterative linear systems. As well as *Lin-MMSE-BD-All-Pass-Filter*, this suboptimal solution has presented a great capacity and a BER near the optimal at high SNR.

4.5 Chapter Summary

In this chapter several suboptimal approaches have been proposed for the design of the precoding and relaying matrices. As it has been shown in previous chapters, as a consequence of the non-convexity of the optimization problem, an iterative algorithm is required to get a local optimal solution, increasing the computational complexity heavily. In order to avoid the iterative process, three novel design strategies have been proposed in this chapter.

Firstly, a suboptimal all pass filtering approach has been introduced in Section 4.1, where no processing is carried out at the relay. This has been only considered for power-limiting, focusing on the optimization of the systems on the precoding matrix.

Afterwards, an independent hop-by-hop suboptimal linear solution has been presented in Section 4.2. This design strategy focuses on the independent minimization of the mean square error of each hop, obtaining independent precoding and relaying matrices in a simple efficient manner.

The last proposal has been centered around the diagonalization of the channel. Thus, by means of singular value decomposition, the equivalent channel created from the base station to the end users becomes diagonal, simplifying the analysis. As it has been described in Section 4.3, water-filling has been also applied after diagonalization for power allocation. Furthermore, it has been observed that the optimization problem has been scalarized, making its solution simpler.

Nevertheless, diagonalization based approaches, designed for MSE minimization and sum-rate maximization, have presented a disadvantage. As a consequence of the nature of SVD, not all the users can always be served. Therefore, two novel algorithms have been proposed to overcome this problem. The first has been designed for equal power distribution, whereas the second has been based on user selection, bringing much more diversity.

Provided simulation results have shown the degradation of the performance of the aforementioned suboptimal approaches compared to the optimal ones. However, the main advantage of this suboptimal algorithms is presented in Section 4.4, showing the considerable reduction of the complexity.

The main contributions and results of this chapter are the proposal of three novel design strategies for complexity reduction. The first one, which combines BD and all pass filtering, has been proposed as a good candidate for the design of the precoder. Independent hop-

by-hop optimization has been proposed for the independent optimization of the matrices, whilst finally system diagonalization based approach have been suggested to create multiple parallel subchannels from the BS to the end users. Several simulation results have been given to demonstrate the effectiveness of these suboptimal systems, including a complete complexity analysis.

Non-Linear Precoding in Multiuser MIMO AF Relaying Systems

As underlined before, due to the lack of cooperation between receiver terminals, a precoding stage has to be included at the BS. Furthermore, as introduced in Chapter 2, non-linear precoding techniques, such as THP and VP, outperform linear precoding in multiuser MIMO systems. Taking into consideration the outstanding results obtained by non-linear precoding techniques in multiuser MIMO systems, this section analyses the application of such techniques in combination with the joint design of precoding and relaying matrices.

In this chapter, firstly a design based on THP precoding at the BS is performed, cancelling the interferences from the base station to the end users successively. Secondly, VP is applied to multiuser MIMO relaying systems and the joint design of the matrices is conducted. Finally, block diagonalization is applied combined with VP precoding. Whilst BD is employed at the relay for interference cancellation, VP is used at the BS to minimize the overall error from the BS to the end users while limiting the power at base station and relaying terminal.

Despite its performance benefits, the complexity of the system increases due to VP processing at the BS. In fact, the complexity required for the search of the precoding vector has to be added to the total cost of the computation of precoding and relaying matrices. Furthermore, it should be pointed out that the statistics of the perturbation vector are unknown and approximations cannot be employed for the analysis. Moreover, the perturbation vector has to be searched in each iteration, increasing considerably the complexity of the system.

Finally, simulation results will show that non-linear precoding techniques outperform the already shown linear approaches, being the proposal based on the combination of BD and VP the one that performs best in the considered scenarios.

5.1 Local Optimal Design with Tomlinson-Harashima Precoding

THP, which has been briefly introduced in Section 2.5.2.1, processes user streams sequentially, being its main objective the successive interference cancellation of the streams. Basically, THP is composed by two filters, a feedback filter used for interference cancellation and a feedforward filter employed to shape the transmitted signal, adapting it to the channel.

THP research in multiuser MIMO systems has focused on the joint design of the filters under ZF criterion in [Fischer02a] [Joham04] [Joham06] and MMSE counterpart in [Joham04] [Kusume05] [Joham06]. Taking into consideration the effectiveness of THP for interference mitigation, this precoding technique has been also considered in single-user MIMO relaying schemes. For example, THP is considered in single-user MIMO relaying networks in [Millar11], where an MMSE solution is given for the minimization of the arithmetic mean square error, being the arithmetic-geometric inequality applied for the solution. THP is combined with GMD in [Jiang05a], ensuring the same SNR for each antenna.

Unsurprisingly, THP has been also applied to multiuser MIMO relaying systems. In [Zhou11], direct and relaying links are combined to increase diversity gain. Whilst maximum ratio combining (MRC) weights are derived, V-BLAST ordering algorithm is considered for interference cancellation and order selection. THP, achieved by LQ decomposition, is applied at the RS, whereas data is processed linearly at the BS. Furthermore, [Chae08b] proposes upper and lower bounds on the achievable sum-rate assuming ZF-DPC at the BS, neglecting the direct link. Since DPC is theoretical, THP and adaptive stream selection [Catreux02] are applied at the BS. By means of the latter, the sum-rate is maximized adaptively, determining the number of streams in each case. Finally, in [Kim09] filters are designed under ZF and MMSE criteria, assuming no processing at the relay and employing the direct link.

In this section, THP transmission for MSE minimization is considered, being the filters at the BS and RS designed jointly.

5.1.1 System Model

Figure 5.1 depicts a multiuser MIMO-AF relaying system where THP precoding is applied at the BS for interference cancellation. The framework described in Section 3.1 is used for the analysis, where K N_k -antenna users are targeted, receiving data from a BS equipped with M antennas through an R -antenna relay. In this case, $L = N$ and no receiver equalization is assumed.

As pointed out before, the purpose of THP is the successive cancellation of interference. The precoding stage is composed by two filters: on one hand, the feedforward filter $\mathbf{F} \in \mathbb{C}^{M \times N}$ and on the other hand, the feedback filter $\mathbf{B} \in \mathbb{C}^{N \times N}$, which is a lower triangular

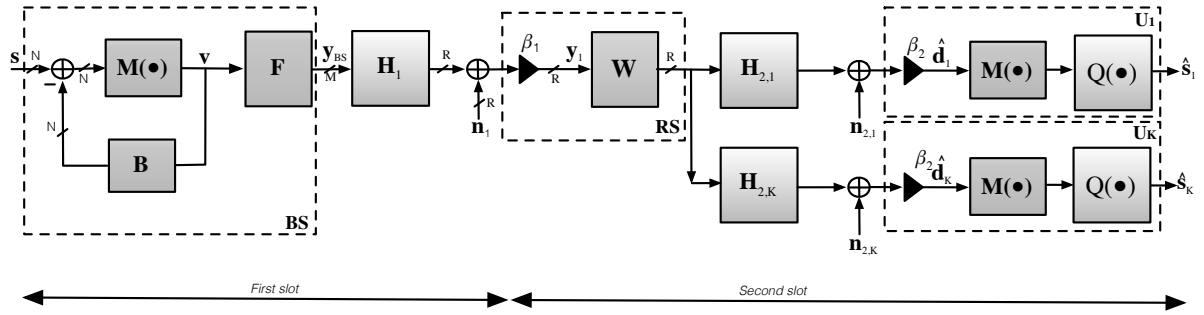


Figure 5.1: Multiuser MIMO downlink AF relaying block diagram for THP precoding.

matrix with zeros in the main diagonal for cancelling the interference sequentially. Thereby, symbols previously processed are fed back, following the property known as *spatial causality* [Joham04].

Note that, since the feedback loop is now located at the transmitter, its summation suggests that the transmitted signal can be unboundedly increased beyond its original constellation, leading to a total transmit power boost, or equivalently, noise enhancement at the receiver.

In order to limit the transmission power, the modulo operation is introduced, which is depicted in Figure 5.1 as the block $\mathbf{M}(\bullet)$. Throughout the modulo operation, the signal is mapped onto a region where the total transmission power is lower. Briefly, the modulo operator can be viewed as the addition of integer multiples of the modulo constant τ [Dietrich07]. The term τ depends on the modulation employed, computed as

$$\tau = C_{max} + \Delta,$$

where C_{max} is the absolute value of the constellation symbol with the largest magnitude in real or imaginary axis, whereas Δ stands for the minimum spacing between constellation points. For example, $\tau = 2\sqrt{2}$ for QPSK modulation.

Briefly, the signal is mapped in the fundamental Voronoi region of the lattice τ as $\mathbb{V} = \tau\mathbb{CZ}$, where $\mathbb{CZ} = \mathbb{Z} + j\mathbb{Z}$ represents a set of Gaussian integers. The Voronoi region is defined as $\nu_M \triangleq \{x + jy | x, y \in [-\tau/2, \tau/2)\}$, which is a square of size τ centered in the origin that ensures the minimization of power consumption, improving in this way the detection.

In addition, the modulo operation works independently on the real and imaginary parts of the signal, i.e.,

$$\mathbf{M}(d) = d - \lfloor \frac{\Re(d)}{2} + \frac{1}{2} \rfloor \tau - j \lfloor \frac{\Im(d)}{2} + \frac{1}{2} \rfloor \tau,$$

where $\Re(d)$ and $\Im(d)$ are the real and imaginary parts of symbol d , respectively, and $\lfloor \bullet \rfloor$ denotes the floor operator.

Considering matrix \mathbf{B} lower triangular, the feedback symbols are defined as

$$v_j = \mathbf{M} \left(s_j + \mathbf{e}_j^T \sum_{i=1}^{j-1} \mathbf{b}_i v_i \right),$$

where s_j is the j^{th} symbol of vector \mathbf{s} , while \mathbf{e}_j and \mathbf{b}_i are the j^{th} and i^{th} column vector of matrices \mathbf{e} and \mathbf{B} , respectively, being \mathbf{e} an identity matrix. The processed and feedback symbols v_j are grouped into the vector $\mathbf{v} = [v_1^T, \dots, v_j^T, \dots, v_N^T]^T \in \mathbb{C}^N$. Obviously, the first symbol is transmitted unaltered. Before the transmission of the second one, it is modified considering the interference created by the first user. This procedure is repeated by all streams until the last one is reached, for which the interference created by all the users is considered.

Since the symbols contained in \mathbf{v} are uncorrelated, the input covariance matrix is $\mathbf{R}_{\mathbf{v}} = \text{diag}(\sigma_{v_1}^2, \dots, \sigma_{v_j}^2, \dots, \sigma_{v_N}^2)$, being $\sigma_{v_j}^2 = \mathbb{E}[|v_j|^2]$. Since the first symbol is sent unaltered, $\sigma_{v_1}^2 = 1$ is taken, whereas, for $k = 2, \dots, K$, $\sigma_{v_j}^2 = \sigma_v^2 = \tau^2/6$ [Kusume05] is considered.

Once the feedback loop is terminated, symbols are sent to the feedforward filter in order to precode the transmitted symbols, which are computed as:

$$\mathbf{y}_{BS} = \mathbf{F}\mathbf{v} \in \mathbb{C}^M,$$

being $\mathbf{F} \in \mathbb{C}^{M \times N}$ the feedforward filter. The precoded symbols are limited in power by the scaling factor $1/\beta_1$, included in the precoded, ensuring that the power at the BS is E_{BS} , i.e:

$$\text{Tr}(\mathbf{F}\mathbf{R}_{\mathbf{v}}\mathbf{F}^H) = E_{BS}.$$

The precoded symbols are transmitted through the channel $\mathbf{H}_1 \in \mathbb{C}^{R \times M}$, where after re-scaling the signal the relay receives:

$$\mathbf{y}_1 = \beta_1 \mathbf{H}_1 \mathbf{F}\mathbf{v} + \beta_1 \mathbf{n}_1 \in \mathbb{C}^R,$$

where $\mathbf{n}_1 \in \mathbb{C}^R$ is the first hop noise with covariance matrix $\mathbf{R}_{\mathbf{n}_1} = N_{01} \mathbf{I}_R$. In the second time slot, after processing the signal at the relay, this forwards:

$$\mathbf{y}_R = \beta_1 \mathbf{W}\mathbf{H}_1 \mathbf{F}\mathbf{v} + \beta_1 \mathbf{W}\mathbf{n}_1 \in \mathbb{C}^R,$$

being $\mathbf{W} \in \mathbb{C}^{R \times R}$ the relaying matrix. As it happens at the BS, the signal at the relay is scaled by the scaling factor $1/\beta_2$, to fulfil with $\|\mathbf{y}_R\|_2^2 = E_R$, which is included in \mathbf{W} . Finally

the total received signal vector after re-scaling is

$$\begin{aligned}\hat{\mathbf{d}} &= \left[\hat{\mathbf{d}}_1^T, \dots, \hat{\mathbf{d}}_K^T \right]^T \\ &= \beta_2 \beta_1 \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{v} + \beta_2 \beta_1 \mathbf{H}_2 \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{n}_2 \in \mathbb{C}^N.\end{aligned}\quad (5.1)$$

The matrix $\mathbf{H}_2 = [\mathbf{H}_{2,1}^T, \dots, \mathbf{H}_{2,K}^T]^T \in \mathbb{C}^{N \times R}$ is the matrix containing the channel gains at the second hop, formed by the channel gains of the users that compose the system. The vector $\mathbf{n}_2 = [\mathbf{n}_{2,1}^T, \dots, \mathbf{n}_{2,K}^T]^T \in \mathbb{C}^K$ stands for the second hop noise with covariance $\mathbf{R}_{\mathbf{n}_2} = N_{0_2} \mathbf{I}_N$. Before detection, the modulo operation has to be carried out at the receiver to revert the modulo operation applied at the transmitter as shown in Figure 5.1.

5.1.2 Joint Filter Desing with Tomlinson-Harashima Precoding

In this section, the joint design of the precoding and relaying matrices is derived under MMSE criterion. The precoding THP strategy will help on the performance improvement due to the interference cancellation from the BS to the end multiple multi-antenna users.

Despite the performance improvement, the computation of this algorithm is not as simple as for the linear ones, where apart from the cost of the filters, the complexity related to the feedback loop has to be computed.

Before continuing a transformation has to be carried out at the transmitter side in order to simplify the analysis. Instead of executing the modulo operation at feedback loop, this can be viewed as the addition of an integer to the transmitted symbols, being the perturbed symbols reduced to:

$$\mathbf{d} = (\mathbf{I}_N - \mathbf{B}) \mathbf{v}, \quad (5.2)$$

where $\mathbf{d} = \mathbf{s} + \mathbf{a}$ is the vector that groups the addition between the symbols and an integer. The Wiener solution will be given, where the overall MSE computed from the BS to the end users is minimized subject to the two power constraints.

Due to the lack of correlation between user stream, the mean square error is defined as:

$$\begin{aligned}\xi &= \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{d}\|_2^2 \right], \\ &= \text{Tr} \left(\beta_1^2 \beta_2^2 \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{R}_{\mathbf{v}} \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{H}_2^H + \beta_2^2 \beta_1^2 N_{0_1} \mathbf{H}_2 \mathbf{W} \mathbf{W}^H \mathbf{H}_2^H + \beta_2^2 N_{0_2} \mathbf{I}_N \right. \\ &\quad \left. - 2\Re \left(\beta_2 \beta_1 \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{R}_{\mathbf{v}} (\mathbf{I}_N - \mathbf{B})^H \right) + (\mathbf{I}_N - \mathbf{B}) \mathbf{R}_{\mathbf{v}} (\mathbf{I}_N - \mathbf{B})^H \right),\end{aligned}$$

being the vectors $\hat{\mathbf{d}}$ and \mathbf{d} defined in (5.1) and (5.2), respectively. Considering the mean

square error as optimization metric, the optimization problem is formulated as

$$\begin{aligned}
 \{\mathbf{F}, \mathbf{W}, \mathbf{B}, \beta_1, \beta_2\} &= \underset{\{\mathbf{F}, \mathbf{W}, \mathbf{B}, \beta_1, \beta_2\}}{\operatorname{argmin}} \quad \xi \\
 \text{s.t.} \quad & \operatorname{Tr}(\mathbf{F}\mathbf{R}_v\mathbf{F}^H) = E_{BS}, \\
 & \beta_1^2 \operatorname{Tr}(\mathbf{W}(\mathbf{H}_1\mathbf{F}\mathbf{R}_v\mathbf{F}^H\mathbf{H}_1^H + N_{0_1}\mathbf{I}_R)\mathbf{W}^H) = E_R, \\
 & \mathbf{S}_j\mathbf{B}\mathbf{e}_j = \mathbf{0} \quad \text{for } j = 1, \dots, N,
 \end{aligned} \tag{5.3a}$$

where the constraint (5.3a) is derived from the triangular nature of \mathbf{B} . Matrix \mathbf{S} denotes a projection used to null-out the rows of the equivalent matrix related to the previously precoded users. Thus, $\mathbf{S}_j = [\mathbf{I}_j, \mathbf{0}_{j \times N-j}] \in \{0, 1\}^{j \times N}$ selects the first j elements from an N -dimensional vector.

Nevertheless, due to the non-convexity of the optimization problem defined in (5.3), getting a global solution is unfeasible. Hence, the iterative method employed for the joint linear design described in Chapter 3 will be applied. The solution is obtained by means of the Lagrangian multipliers. Carrying the procedure described in Section B.2 and continuing with the steps described in [Joham04], we obtain the next expressions for the precoding matrix and the scaling factor β_1

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} N}{E_R} \mathbf{I}_R \right) \overline{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{B}, \quad \text{and} \tag{5.4}$$

$$\beta_1 = \sqrt{\frac{\operatorname{Tr} \left(\mathbf{R}_v \mathbf{B}^H \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} N}{E_R} \mathbf{I}_R \right) \overline{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-2} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{B} \right)}{E_{BS}}},$$

where $\alpha_1 = \frac{N_{0_1}}{E_{BS}} \operatorname{Tr} \left(\overline{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} N}{E_R} \mathbf{I}_R \right) \overline{\mathbf{W}} \right)$. Furthermore, $\overline{\mathbf{W}} = \beta_2 \mathbf{W}$ is the unscaled version of the relaying matrix. In the same fashion, the relaying matrix and the scaling factor β_2 are defined as:

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} N}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H \mathbf{B} \mathbf{R}_v \overline{\mathbf{F}}^H \mathbf{H}_1^H \left(\mathbf{H}_1 \overline{\mathbf{F}} \mathbf{R}_v \overline{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{0_1} \mathbf{I}_R \right)^{-1} \tag{5.5}$$

$$\beta_2 = \sqrt{\frac{\operatorname{Tr} \left(\mathbf{B} \mathbf{R}_v \overline{\mathbf{F}}^H \mathbf{H}_1^H \mathbf{A}^{-1} \mathbf{H}_1 \overline{\mathbf{F}} \mathbf{R}_v \mathbf{B}^H \mathbf{H}_2 \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} N}{E_R} \mathbf{I}_R \right)^{-2} \mathbf{H}_2^H \right)}{E_R}},$$

where $\mathbf{A} = \left(\mathbf{H}_1 \overline{\mathbf{F}} \mathbf{R}_v \overline{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{0_1} \mathbf{I}_R \right)$ and $\overline{\mathbf{F}} = \beta_1 \mathbf{F}$. Furthermore, we recall that β_1 can be

replaced following $\beta_1^2 = \frac{\text{Tr}(\bar{\mathbf{F}}\mathbf{R}_v\bar{\mathbf{F}}^H)}{E_{BS}}$ for simplicity. Finally the feedback filter \mathbf{B} is computed as

$$\mathbf{B} = \sum_{j=1}^N (\mathbf{S}_j^T \mathbf{S}_j - \mathbf{I}_N) \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \bar{\mathbf{F}} \mathbf{e}_j \mathbf{e}_j^T.$$

As it happens with the linear analysis, as a consequence of the interconnection of the matrices, an iterative process has to be accomplished whose complexity increases due to the computation of \mathbf{B} . Algorithms 6, 7 and 8, described in the following lines, depict the general iterative process, the computation of \mathbf{B} and the feedback loop, respectively. Since no ordering is assumed, l stands for the current iteration and $\|\mathbf{A}\|^2$ for the norm operation.

Algorithm 6 Estimation of $\bar{\mathbf{F}}$, $\bar{\mathbf{W}}$ and \mathbf{B} for THP MMSE approach.

- 1: Initialize the variables: $\bar{\mathbf{F}}^0 = \mathbf{I}_{M \times N}$, $\bar{\mathbf{W}}^0 = \mathbf{I}_R$ and $\epsilon = \infty$.
 - 2: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 3: Compute \mathbf{B}^{l+1} following Algorithm 7.
 - 4: Calculate $\bar{\mathbf{F}}^{l+1}$ (5.4) for fixed $\bar{\mathbf{W}}^l$ and \mathbf{B}^{l+1} .
 - 5: Computation of $\bar{\mathbf{W}}^{l+1}$ (5.5) for updated $\bar{\mathbf{F}}^{l+1}$ and \mathbf{B}^{l+1} .
 - 6: Error calculation: $\epsilon = \|\bar{\mathbf{F}}^{l+1} - \bar{\mathbf{F}}^l\|^2 + \|\bar{\mathbf{W}}^{l+1} - \bar{\mathbf{W}}^l\|^2$.
 - 7: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 8: $\bar{\mathbf{F}}^l = \bar{\mathbf{F}}^{l+1}$
 - 9: $\bar{\mathbf{W}}^l = \bar{\mathbf{W}}^{l+1}$.
 - 10: **end if**
 - 11: **end while**
-

Algorithm 7 Computation of \mathbf{B} under MMSE criterion.

- 1: **for** $j = 1, \dots, N$ **do**
 - 2: $\mathbf{b}_j = (\mathbf{S}_j^T \mathbf{S}_j - \mathbf{I}_N) \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{f}_j$.
 - 3: **end for**
-

Algorithm 8 Feedback loop.

- 1: **for** $j = 1, \dots, N$ **do**
 - 2: $v_j = \mathbf{M} \left(s_j + \sum_{i=1}^{j-1} \mathbf{e}_j^T \mathbf{b}_i v_i \right)$.
 - 3: **end for**
 - 4: $\mathbf{v} = [v_1^T, \dots, v_N^T]^T$.
 - 5: $\mathbf{y}_{BS} = \mathbf{F}\mathbf{v}$.
-

Provided simulation results will show the effectiveness of THP over the linear approaches, ensuring the cancellation of the interference from the BS to the end users. The main drawback of this design based on THP is its computational cost, increased due to the modulo operation at both transmitter and receiver, and specially because of the feedback loop.

5.2 Joint Linear Design with VP Transmission

The non-linear VP precoding technique was previously introduced in Section 2.5.2.2. Despite its better performance in comparison to THP, vector precoding has not been considered as precoding option for multiuser MIMO relaying environments due to the encompassed computational complexity. Apart from the search of precoding and relaying matrices, the optimum precoding vector has to be found, increasing considerably the computational cost of the system.

What is more, vector precoding presents another disadvantage. The perturbation vector has to be searched in an ideally infinite lattice, which leads to a non-deterministic NP hard problem, which is solved employing lattice search techniques. In [Hochwald05] SE is introduced, translating the idea of sphere decoding to the transmitter. Throughout this encoding algorithm, the exhaustive search within all the integers that forms the lattice is avoided, limiting the search to a hypersphere of radius R centered around a starting point.

Furthermore, lattice reduction techniques have been widely used to search the precoding vector with a moderate complexity. The main aim of these suboptimal techniques is the employment of an equivalent and more advantageous set of basis vectors to allow the suboptimal resolution of the problem. For example, Windpassinger et al [Windpassinger04] find the perturbation vector throughout Babai's approximations. In the same way, Meurer et al propose to split the symbols in groups to reduce the dimensionality of the problem for the closest point search in [Meurer04].

However, the statistics of the perturbation vector are unknown and really complex to obtain. Contrary to THP, there are not analytical expressions or approximations for vector precoding, being this problem widely studied in multiuser MIMO system. In [Damen03] and [Barbero08] the inferior power limit is obtained for lattice theory, allowing to set a maximum limit for the sum-capacity term. Finally, Barrenechea introduces several bounds on the unscaled power, giving lower and upper bounds on the data rate of VP in multiuser systems in [Barrenechea10]. Nevertheless, the statistics of the precoded symbols are not obtained for multiuser MIMO relaying systems, largely because the transmission in two hops and the complexity that it involves.

In this section, multiuser MIMO-AF relaying is analysed under VP transmission. The precoding and relaying matrices are jointly obtained under MMSE criterion. In fact, VP is used for error minimization while limiting the total transmission power at the BS and RS, respectively.

Apart from this, the consideration of VP and block diagonalization is proposed. Due to the non-convergence of BD from the BS to the end users (see Chapter 3), VP is applied at the BS to limit the power at both BS and RS terminals, whereas BD is applied at the RS for block diagonalization and interference reduction.

5.2.1 System Model

The system is the one employed in Section 3.1, where three different terminals take part in the communication: a base station equipped with M antennas, an R -antenna relay and K multi-antenna users.

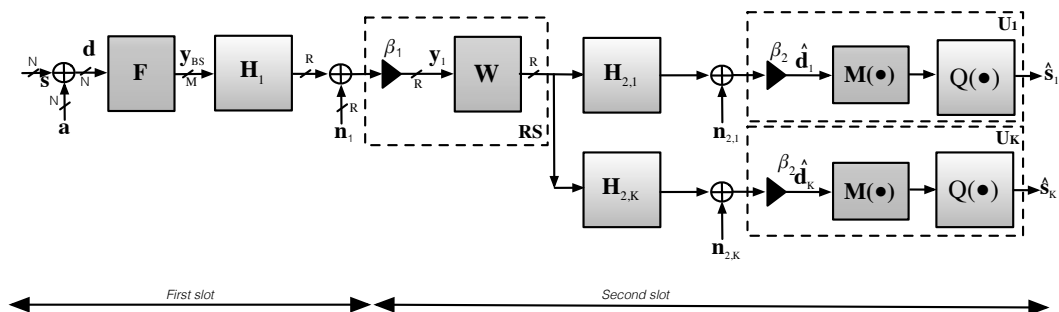


Figure 5.2: Multiuser MIMO downlink AF relaying block diagram for VP precoding.

The main difference between the system analysed in Section 3.1 and the one that will be used for VP analysis accomplished here is that the symbols are perturbed before the transmission, so that

$$\mathbf{d} = \mathbf{s} + \mathbf{a},$$

being $\mathbf{a} \in \mathbb{C}^N$ the perturbation vector and $\mathbf{d} \in \mathbb{C}^N$ the perturbed symbols. As pointed out before, vector precoding was inspired by the idea of THP. The modulo operation is replaced by the addition of a perturbation vector to the transmitted signal. Furthermore, feedback and feedforward filters used by THP processing can be combined in filter \mathbf{F} , reducing the number of parameters to be optimized.

Remembering that the total received signal is

$$\hat{\mathbf{d}} = \left[\hat{\mathbf{d}}_1^T, \dots, \hat{\mathbf{d}}_k^T, \dots, \hat{\mathbf{d}}_K^T \right]^T = \beta_2 \beta_1 \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2 \beta_1 \mathbf{H}_1 \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{n}_2 \in \mathbb{C}^K, \quad (5.6)$$

where $\hat{\mathbf{d}}_k \in \mathbb{C}^{N_k}$ is the signal received by user k , which can be grouped into the vector $\hat{\mathbf{d}}$, while $\mathbf{y}_{BS} = \mathbf{F} \mathbf{d} \in \mathbb{C}^M$ and \mathbf{W} are the precoded symbols and relaying matrix, respectively.

Finally, before detection, the modulo operation is carried out at each receiver. By means of this, the perturbation vector is removed from the signal without any knowledge of the perturbation vector \mathbf{a} that has been previously chosen at the transmitter. Hence, VP's detection complexity is not increased with respect to THP, because it carries out only a re-scale process, a modulo operation and a quantization process. In the following section, the local optimal design approaches based on VP and BD-VP are studied, deriving the optimal

perturbation vector as well as the local optimal precoding and relaying matrices.

5.2.2 Joint Design of the Filters

In this section, precoding and relaying matrices are jointly designed under MMSE criterion. Furthermore, vector precoding is added at the base station to improve the performance of the system. In this case, the perturbation vector will be computed to minimize the MSE as:

$$\mathbf{a} = \underset{\mathbf{a}' \in \tau\mathbb{Z}^N + j\tau\mathbb{Z}^N}{\operatorname{argmin}} \quad \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{d}\|_2^2 \right],$$

where $\hat{\mathbf{d}}$ is the total received signal defined in (5.6). Obviously, the selection of the perturbation vector will affect the mean square error, defined as the difference between the data vector \mathbf{d} and the received symbols vector $\hat{\mathbf{d}}$ before the modulo operation. When the vector $\hat{\mathbf{d}}$ is a good approximation of \mathbf{d} , the estimated symbols $\hat{\mathbf{s}}$ will be also near the unperturbed symbol vector \mathbf{s} [Hochwald05].

Similarly to the linear analysis accomplished in Chapters 3 and 4, the optimization will be carried out in two steps. Firstly, the vector \mathbf{y}_{BS} , the matrix \mathbf{W} and the scaling factors β_1 and β_2 are optimized, and then the perturbation vector \mathbf{a} is searched for the minimization of the quadratic error under the knowledge of \mathbf{y}_{BS} , \mathbf{W} and the scaling factors β_1 and β_2 . Considering this, the mean square error is defined as:

$$\begin{aligned} \xi &= \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{d}\|_2^2 \right], \\ &= \beta_2^2 \beta_1^2 \mathbf{y}_{BS}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2^2 \beta_1^2 N_{01} \operatorname{Tr} (\mathbf{W}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{W}) + \beta_2^2 N_{02} N \\ &\quad - 2\beta_2 \beta_1 \Re (\mathbf{d}^H \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS}) + \mathbf{d}^H \mathbf{d}. \end{aligned} \quad (5.7)$$

In order to get the complete MMSE solution, we define the optimization problem as:

$$\begin{aligned} \{\mathbf{a}, \mathbf{F}, \mathbf{W}, \beta_1, \beta_2\} &= \underset{\{\mathbf{a}, \mathbf{F}, \mathbf{W}, \beta_1, \beta_2\}}{\operatorname{argmin}} \quad \xi \\ \text{s.t.} \quad &\|\mathbf{y}_{BS}\|_2^2 = E_{BS}, \end{aligned} \quad (5.8a)$$

$$\|\mathbf{y}_R\|_2^2 = E_R, \quad (5.8b)$$

being E_{BS} and E_R the total transmission power at the BS and RS, respectively. The optimization problem is highly non-convex. Therefore, instead of applying quadratic programming algorithms, a simpler numerical method is proposed to get a close solution as in [Fang06].

Briefly, this numerical method, previously employed for THP based precoding, obtains a local solution in an iterative fashion. Hence, for fixed \mathbf{W} and β_2 , \mathbf{y}_{BS} and β_1 are optimized and vice versa. After setting the Lagrangian function as in Appendix B.2, and carrying out the procedure described there, the expressions for the precoded symbols and the scaling

factor β_1 are computed as

$$\mathbf{y}_{BS} = \frac{1}{\beta_1} \underbrace{\left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02}N}{E_R} \mathbf{I}_R \right) \overline{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{d}}_{\mathbf{F}} \quad (5.9)$$

and

$$\beta_1 = \sqrt{\frac{\mathbf{d}^H \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02}N}{E_R} \mathbf{I}_R \right) \overline{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-2} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{d}}{E_{BS}}},$$

where $\alpha_1 = \frac{N_{01}}{E_{BS}} \text{Tr} \left(\overline{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02}N}{E_R} \mathbf{I}_R \right) \overline{\mathbf{W}} \right)$ and $\overline{\mathbf{W}} = \beta_2 \mathbf{W}$.

In the same fashion, the relaying matrix \mathbf{W} and the scaling factor β_2 are found, getting

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02}N}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H \mathbf{R}_d \overline{\mathbf{F}}^H \mathbf{H}_1^H \left(\mathbf{H}_1 \overline{\mathbf{F}} \mathbf{R}_d \overline{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R \right)^{-1} \quad (5.10)$$

and

$$\beta_2 = \sqrt{\frac{\text{Tr} \left(\overline{\mathbf{F}}^H \mathbf{H}_1^H \mathbf{A}^{-1} \mathbf{H}_1 \overline{\mathbf{F}} \mathbf{R}_d \mathbf{H}_2 \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02}N}{E_R} \mathbf{I}_R \right)^{-2} \mathbf{H}_2^H \mathbf{R}_d \right)}{E_R}},$$

where $\mathbf{A} = \left(\mathbf{H}_1 \overline{\mathbf{F}} \mathbf{R}_d \overline{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R \right)$ and $\mathbf{R}_d = \text{E} [\mathbf{d} \mathbf{d}^H]$ is the precoded signal covariance matrix, while $\overline{\mathbf{F}} = \beta_1 \mathbf{F}$. It should be pointed out that \mathbf{R}_d is unknown and approximations cannot be used for matrix derivation. Thus, the perturbation vector has to be searched in each iteration.

It is important to point out that the optimum transmit vector at the BS is a linearly filtered version of the symbol vector \mathbf{d} . Assuming that the local precoding and relaying matrices have been found, in order to get the optimal precoding vector, the expression for \mathbf{y}_{BS} is introduced in (5.7) and with the help of the matrix inversion lemma, we get:

$$\xi = \mathbf{d}^H \left(\mathbf{I}_N - \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}} \right) \mathbf{d}. \quad (5.11)$$

By means of the procedure depicted in [Schmidt05], using the Cholesky factorization of the inner term in (5.11) as follows,

$$\left(\mathbf{I}_N - \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}} \right) = \mathbf{L}^H \mathbf{L},$$

the computation of the perturbing signal vector \mathbf{a} can be simplified to

$$\mathbf{a} = \underset{\mathbf{a}' \in \tau\mathbb{Z}^N + j\tau\mathbb{Z}^N}{\operatorname{argmin}} \quad \|\mathbf{L}(\mathbf{s} + \mathbf{a}')\|_2^2, \quad (5.12)$$

being $\mathbf{L} \in \mathbb{C}^{N \times N}$ a lower triangular matrix. Obviously, the optimum choice of the perturbation vector \mathbf{a} is the solution to a closest point in a lattice. In this case, the lattice is generated by $\tau\mathbf{L}$, and we are looking for the integer vector that corresponds to the lattice point closest to $-\mathbf{L}\mathbf{s}$.

Finally, it should be pointed out that an iterative process has to be implemented to get a local optimal solution for VP approach, which includes the search for the optimum precoding vector in each iteration, increasing the cost considerably.

Algorithm 9 Computation of $\overline{\mathbf{F}}$, $\overline{\mathbf{W}}$ and \mathbf{a} under the MMSE criterion.

- 1: Initialize the variables: $\overline{\mathbf{F}}^0 = \mathbf{I}_{M \times N}$, $\overline{\mathbf{W}}^0 = \mathbf{I}_R$ and $\epsilon = \infty$.
 - 2: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 3: Calculate $\overline{\mathbf{F}}^{l+1}$ (5.9) for a fixed $\overline{\mathbf{W}}^l$.
 - 4: Cholesky factorization of the inner term of the mean square error (5.11) with $\overline{\mathbf{W}}^l$ and $\overline{\mathbf{F}}^{l+1}$.
 - 5: Sphere encoder algorithm execution for the search of the perturbing vector \mathbf{a}^{l+1} .
 - 6: Computation of $\overline{\mathbf{W}}^{l+1}$ (5.10) for the updated $\overline{\mathbf{F}}^{l+1}$ and \mathbf{a}^{l+1} .
 - 7: Error calculation: $\epsilon = \|\overline{\mathbf{F}}^{l+1} - \overline{\mathbf{F}}^l\|^2 + \|\overline{\mathbf{W}}^{l+1} - \overline{\mathbf{W}}^l\|^2$.
 - 8: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 9: $\overline{\mathbf{F}}^l = \overline{\mathbf{F}}^{l+1}$.
 - 10: $\overline{\mathbf{W}}^l = \overline{\mathbf{W}}^{l+1}$.
 - 11: **end if**
 - 12: **end while**
-

As expected, provided simulation results will show the performance improvement of VP when it is applied to multiuser MIMO relaying networks. Nevertheless, the computational cost increases considerably with the execution of the sphere encoding algorithm in each iteration. In the following chapter, suboptimal novel VP solutions are proposed which reduce considerably the cost as a consequence of signal degradation.

5.2.3 Vector Precoding with Block Diagonalization

In this section, a novel approach is proposed for multiuser MIMO relaying networks with VP precoding. Under the knowledge that BD can be employed combined with VP for MUI and ISI cancellation in multiuser MIMO networks, BD-VP is here proposed for multiuser MIMO relaying networks. Furthermore, BD will be used for MUI cancellation and VP for error minimization and power limitation, instead of ISI cancellation as in [Chae08a].

Summing up, the proposal described in this section encompasses three different ideas: BD at the relay for interference cancellation, precoding at the BS for users' signal separation and, finally, VP for error minimization and power limitation. Provided simulation results will show the effectiveness of this proposal, which also improves the joint local optimal VP design introduced in Section 5.2.2. Considering again the model in Figure 5.2, the total received signal by the user k is defined as:

$$\hat{\mathbf{d}}_k = \beta_2\beta_1\mathbf{H}_{2,k}\mathbf{W}\mathbf{H}_1\mathbf{y}_{BS} + \beta_2\beta_1\mathbf{H}_{2,k}\mathbf{W}\mathbf{n}_1 + \beta_2\mathbf{n}_{2,k},$$

where $\mathbf{H}_{2,k} \in \mathbb{C}^{N_k \times R}$ denotes the channel vector from the relay to the k^{th} user, grouped into the matrix $\mathbf{H}_2 = [\mathbf{H}_{2,1}^T, \dots, \mathbf{H}_{2,k}^T, \dots, \mathbf{H}_{2,K}^T]^T \in \mathbb{C}^{N \times R}$. For BD application at the relay, the relaying matrix has to be defined such as $\mathbf{W} = \mathbf{W}_2\mathbf{W}_1$, being $\mathbf{W}_2 = [\mathbf{W}_{2,1}, \dots, \mathbf{W}_{2,k}, \dots, \mathbf{W}_{2,K}] \in \mathbb{C}^{R \times N}$ for block diagonalization approach and $\mathbf{W}_1 = [\mathbf{W}_{1,1}^T, \dots, \mathbf{W}_{1,k}^T, \dots, \mathbf{W}_{1,K}^T]^T \in \mathbb{C}^{N \times R}$ used for signal processing. With this relaying matrix structure, the received signal at user k is defined as follows:

$$\begin{aligned} \hat{\mathbf{d}}_k &= \beta_2\beta_1\mathbf{H}_{2,k}\mathbf{W}_{2,k}\mathbf{W}_{1,k}\mathbf{H}_1\mathbf{y}_{BS} + \underbrace{\sum_{\substack{j=1 \\ j \neq k}}^K \beta_2\beta_1\mathbf{H}_{2,k}\mathbf{W}_{2,j}\mathbf{W}_{1,j}\mathbf{H}_1\mathbf{y}_{BS} + \beta_2\beta_1\mathbf{H}_{2,k}\mathbf{W}_{2,j}\mathbf{W}_{1,j}\mathbf{n}_1}_{\text{Interference}} \\ &+ \underbrace{\beta_2\beta_1\mathbf{H}_{2,k}\mathbf{W}_{2,k}\mathbf{W}_{1,k}\mathbf{n}_1 + \beta_2\mathbf{n}_{2,k}}_{\text{Noise}}. \end{aligned}$$

Owing to the interference suppression, the constraint $\mathbf{H}_{2,k}\mathbf{W}_{2,j} = \mathbf{0}_{N_k}$ has to be set. Carrying out the steps introduced in Section 3.3.3, we get

$$\hat{\mathbf{d}}_k = \beta_2\beta_1\mathbf{H}_{e,k}\mathbf{W}_{1,k}\mathbf{H}_1\mathbf{y}_{BS} + \beta_2\beta_1\mathbf{H}_{e,k}\mathbf{W}_{1,k}\mathbf{n}_1 + \beta_2\mathbf{n}_{2,k},$$

being $\mathbf{H}_{e,k} = \mathbf{H}_{2,k}\mathbf{W}_{2,k}$ the matrix that represents the equivalent channel created from the relay to the multi-antenna user k . The received signal of each user can be grouped into the data vector $\hat{\mathbf{d}} = [\hat{\mathbf{d}}_1^T, \dots, \hat{\mathbf{d}}_k^T, \dots, \hat{\mathbf{d}}_K^T]^T$ as

$$\hat{\mathbf{d}} = \beta_2\beta_1\mathbf{H}_e\mathbf{W}_1\mathbf{H}_1\mathbf{y}_{BS} + \beta_2\beta_1\mathbf{H}_e\mathbf{W}_1\mathbf{n}_1 + \mathbf{n}_2,$$

where the matrix $\mathbf{H}_e = \text{blkdiag}(\mathbf{H}_{e,1}, \dots, \mathbf{H}_{e,k}, \dots, \mathbf{H}_{e,K}) \in \mathbb{C}^{N \times N}$ collects the equivalent channel gains for the second hop.

5.2.3.1 Joint Design of Precoder and Relay under BD-VP Criterion

As described in Section 5.2, where the joint optimal design considers VP transmission at the base station, the optimization is carried out in two steps. Firstly, assuming the perturbation

vector \mathbf{a} , the vector \mathbf{y}_{BS} , the matrix \mathbf{W}_1 as well as the scaling factors β_1 and β_2 are optimized, whereas in the second step, once the previous parameters are set, the optimal precoding vector \mathbf{a} that minimizes the MSE is chosen. The MSE error, defined in (5.7), is reduced to:

$$\begin{aligned} \xi &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{d}_k\|_2^2 \right], \\ &= \sum_{k=1}^K \beta_2^2 \beta_1^2 \mathbf{y}_{BS}^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2^2 \beta_1^2 N_{01} \text{Tr} \left(\mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{H}_{e,k} \mathbf{W}_{1,k} \right) + N_{02} N \\ &\quad - 2\Re \left(\mathbf{d}_k^H \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS} \right) + \mathbf{d}^H \mathbf{d}, \end{aligned} \quad (5.13)$$

being $\mathbf{H}_{e,k}$ the equivalent channel matrix. The perturbation vector will be chosen as the closest point in the lattice as

$$\mathbf{a} = \underset{\mathbf{a}' \in \tau\mathbb{Z}^N + j\tau\mathbb{Z}^N}{\text{argmin}} \quad \xi,$$

where ξ stands for the mean square error defined in (5.13). The optimization problem is again defined as

$$\{\mathbf{a}, \mathbf{F}, \mathbf{W}_1, \beta_1, \beta_2\} = \underset{\{\mathbf{a}, \mathbf{F}, \mathbf{W}_1, \beta_1, \beta_2\}}{\text{argmin}} \quad \xi$$

$$\text{s.t.} \quad \|\mathbf{y}_{BS}\|_2^2 = E_{BS}, \quad (5.14a)$$

$$\|\mathbf{y}_R\|_2^2 = E_R. \quad (5.14b)$$

Once more, the alternating numerical method introduced in [Fang06] has been chosen for the optimization. It should be pointed out that the term $\mathbf{W}_{2,k}^H \mathbf{W}_{2,k}$ has been omitted due to the orthogonality of the matrix $\mathbf{W}_{2,k}$. Carrying out the procedure accomplished in Section B.2, the expression for the precoded symbols and the scaling factor at the first hop are

$$\mathbf{y}_{BS} = \frac{1}{\beta_1} \underbrace{\left(\mathbf{H}_1^H \overline{\mathbf{W}}_1^H \left(\mathbf{H}_e^H \mathbf{H}_e + \frac{N_{02} N}{E_R} \mathbf{I}_N \right) \overline{\mathbf{W}}_1 \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}_1^H \mathbf{H}_e^H \mathbf{d}}_{\mathbf{F}}, \quad (5.15)$$

and

$$\beta_1 = \sqrt{\frac{\mathbf{d}^H \mathbf{H}_e \overline{\mathbf{W}}_1 \mathbf{H}_1 \left(\mathbf{H}_1^H \overline{\mathbf{W}}_1^H \left(\mathbf{H}_e^H \mathbf{H}_e + \frac{N_{02} N}{E_R} \mathbf{I}_N \right) \overline{\mathbf{W}}_1 \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-2} \mathbf{H}_1^H \overline{\mathbf{W}}_1^H \mathbf{H}_e^H \mathbf{d}}{E_{BS}}},$$

being $\alpha_1 = \frac{N_{01}}{E_{BS}} \text{Tr} \left(\overline{\mathbf{W}}_1 \left(\mathbf{H}_e^H \mathbf{H}_e + \frac{N_{02} N}{E_R} \mathbf{I}_N \right) \overline{\mathbf{W}}_1 \right)$ and $\overline{\mathbf{W}}_1 = \beta_2 \mathbf{W}_1$. The precoding matrix defined in (5.15) denotes the pseudo inverse of the equivalent channel created from the BS to the end users. In the same fashion, the relaying matrix related to user k and the scaling

factor β_2 are

$$\mathbf{W}_{1,k} = \frac{1}{\beta_2} \left(\mathbf{H}_{e,k}^H \mathbf{H}_{e,k} + \frac{N_{0_2} N}{E_R} \mathbf{I}_{N_k} \right)^{-1} \mathbf{H}_{e,k}^H \mathbf{R}_{\mathbf{d},k} \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \left(\mathbf{H}_1 \bar{\mathbf{F}} \mathbf{R}_{\mathbf{d}} \bar{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{0_1} \mathbf{I}_R \right)^{-1} \quad (5.16)$$

where $\bar{\mathbf{F}} = [\bar{\mathbf{F}}_1, \dots, \bar{\mathbf{F}}_k, \dots, \bar{\mathbf{F}}_K] = \beta_1 \mathbf{F}$ is the unscaled precoding matrix at the BS and $\mathbf{R}_{\mathbf{d}} = \mathbb{E}[\mathbf{d}\mathbf{d}^H]$ the covariance matrix. The scaling factor β_2 that limits the power at the relay to E_R is

$$\beta_2 = \sqrt{\frac{\sum_{k=1}^K \text{Tr} \left(\mathbf{R}_{\mathbf{d},k} \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \mathbf{A}^{-1} \mathbf{H}_1 \bar{\mathbf{F}}_k \mathbf{R}_{\mathbf{d},k} \mathbf{H}_{e,k} \left(\mathbf{H}_{e,k}^H \mathbf{H}_{e,k} + \frac{N_{0_2} N}{E_R} \mathbf{I}_{N_k} \right)^{-2} \mathbf{H}_{e,k}^H \right)}{E_R}},$$

being $\mathbf{A} = \left(\mathbf{H}_1 \bar{\mathbf{F}} \mathbf{R}_{\mathbf{d}} \bar{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{0_1} \mathbf{I}_R \right)$.

Once the precoding (5.15) and relaying (5.16) matrices are derived, the next step will address the computation of the perturbation vector. Following [Schmidt05], the expression for the precoded symbols \mathbf{y}_{BS} (5.15) is substituted in (5.13) and with the help of the matrix inversion lemma, we get

$$\boldsymbol{\xi} = \mathbf{d}^H \left(\mathbf{I}_N - \mathbf{H}_e \bar{\mathbf{W}}_1 \mathbf{H}_1 \bar{\mathbf{y}}_{BS} \right) \mathbf{d}. \quad (5.17)$$

Carrying out the Cholesky factorization of the inner term of (5.17)

$$\left(\mathbf{I}_N - \mathbf{H}_e \bar{\mathbf{W}}_1 \mathbf{H}_1 \bar{\mathbf{y}}_{BS} \right) = \mathbf{L}^H \mathbf{L},$$

where $\mathbf{L} \in \mathbb{C}^{N \times N}$ stands for a lower triangular matrix. The cost function for the search of the vector \mathbf{a} is again obtained as in (5.12) being \mathbf{s} the symbol data vector transmitted by the BS.

Provided simulation results will show the effectiveness of BD combined with VP, improving the performance of the local optimal VP approach introduced in the previous section due to the interference cancellation provided by BD at the relay.

5.3 Simulation Results

In this section, the performance of the non-linear approaches will be analysed and compared to the linear approach introduced in Chapter 3. Three different systems have been tested and compared with *Lin-MMSE: THP-MMSE* (Section 5.1.2), *VP-MMSE* (Section 5.2.2) and *VP-MMSE-BD* (Section 5.2.3).

We assume the parameters defined in Section 3.2.4, where M transmission antennas are used at the BS, while an R -antenna relay and K N_k -antenna users take part in the

framework.

Apart from $4 \times 4 \times \{4 \times 1\}$ set-up (A), the results will be also given for multi-antenna set-up E ($8 \times 8 \times \{2 \times 4\}$) in order to show the effectiveness of BD-VP approach when multi-antenna users are included.

An ensemble of 10^4 Rayleigh channel matrix realizations \mathbf{H}_k for $k = 1, 2$ have been evaluated, being \mathbf{H}_1 and \mathbf{H}_2 the matrices containing the channel matrix gains for first and second hops, respectively. The modulation chosen is QPSK where $N_B = 100$ block of symbols are sent. Moreover, the signal-to-noise ratio of each hop have been assumed independent, being the first hop SNR ρ_1 and the second hop SNR ρ_2 . Simulations have been carried out fixing one of the SNRs.

In order to reduce the computational complexity of the search, [Barrenechea12] proposes to limit the lattice to certain values. As analysed in [Barrenechea12] and depicted in Figure 5.3, the search of the perturbation vector is limited to the set $\boldsymbol{\nu}$ of the closest points to the origin, making the computational complexity lower. The region $\boldsymbol{\nu}$ is then defined as $\boldsymbol{\nu} \triangleq \{x + yj : |x|, |y| \leq (\sqrt{|\boldsymbol{\nu}|} - 1) / 2\}$. Unless otherwise stated, $|\boldsymbol{\nu}| = 25$ is assumed from now on.

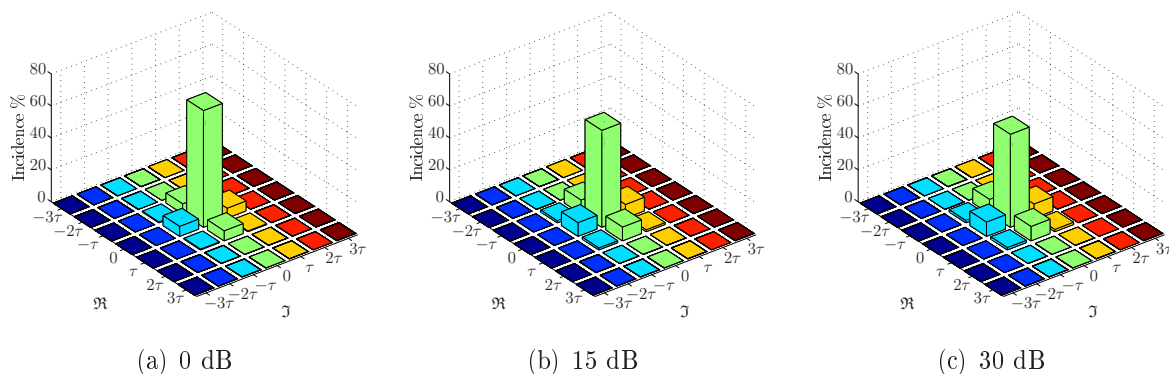


Figure 5.3: Incidence of the lattice values in the optimum solution vector for different SNR regions [Barrenechea12].

Due to the iterative process that has to be accomplished to obtain the precoding and relaying matrices, the convergence error is set to $\epsilon_{\min} = 0.001$. The performance of the aforementioned systems have been measured for BER, achievable sum-rate and convergence analysis.

5.3.1 BER Performance

In this section the BER performance of non-linear systems *THP-MMSE*, *VP-MMSE* and *VP-MMSE-BD* has been compared with the performance of linear approach. Figure 5.4 shows BER performance when $\rho_2 = 15$ dB. As expected, VP based approaches outperform the rest

of the presented proposals, including *THP-MMSE* and the linear one. Furthermore, the best performance is achieved by *VP-MMSE-BD* for both configurations, being the maximum gain with respect to *VP-MMSE* 0.5 dB and 3 dB when $\rho_2 = 15$ dB and $\rho_1 = 15$ dB, respectively. As pointed out in previous chapters, this difference might be greater when users are equipped with multiple antennas instead of one, which is the case analysed here. When ρ_2 is set to 15 dB, the maximum gain that *VP-MMSE-BD* achieves over *THP-MMSE* is 5 dB.

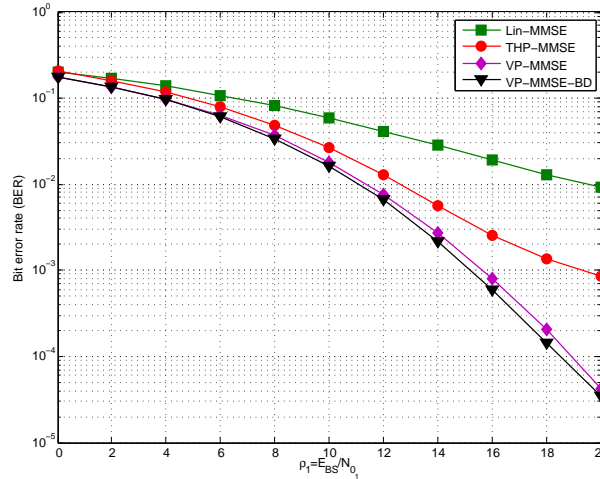


Figure 5.4: BER performance for non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_2 = 15$ dB.

The same conclusions are obtained for the system with a fixed $\rho_1 = 15$ dB as shown in Figure 5.5. Here the percentage of erroneous bits is greater compared with the previous case, as well as the difference between VP based designs and *THP-MMSE*, being the maximum gain of 6 dB.

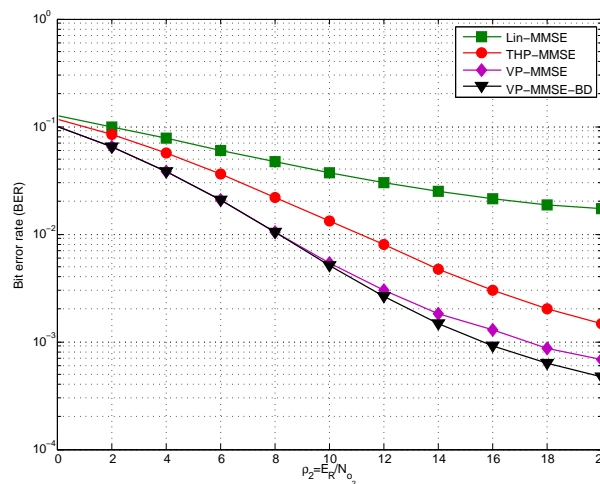


Figure 5.5: BER performance for non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_1 = 15$ dB.

Figure 5.6(a) shows the BER performance of set-up E ($8 \times 8 \times \{2 \times 4\}$) when $\rho_2 = 15$

dB, which depicts the effectiveness of BD in multi-antenna set-ups, getting a difference gap of 2 dB at high SNRs with respect to *VP-MMSE*. Nevertheless, the difference between *VP-MMSE* and *THP-MMSE* is lower. It should be underlined that the gain achieved by non-linear systems over the linear is greater, comparing it with set-up A's results.

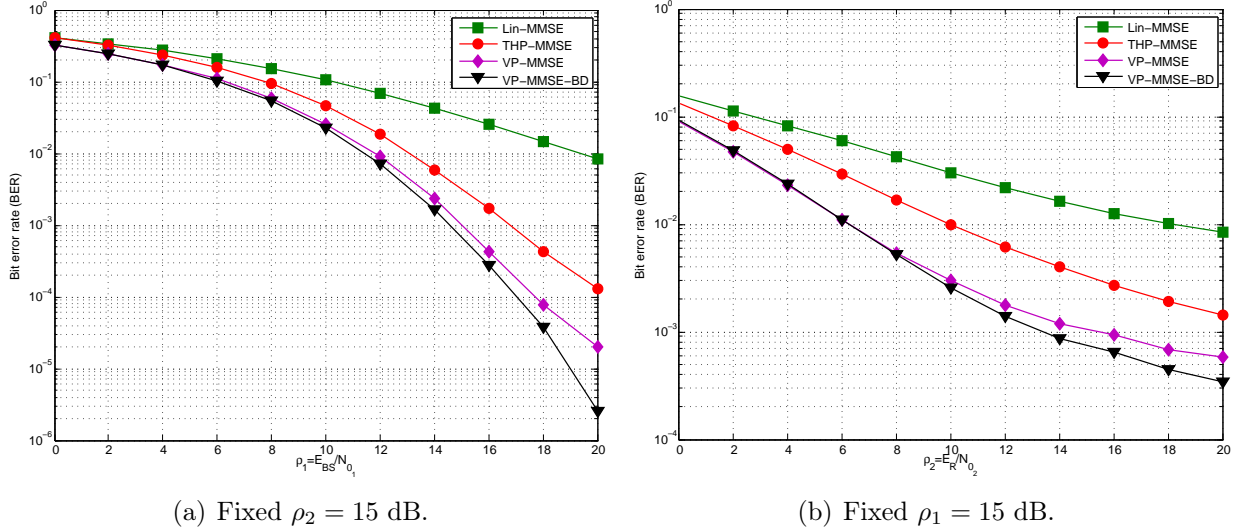


Figure 5.6: BER performance curves for non-linear precoding techniques in the $8 \times 8 \times \{2 \times 4\}$ set-up (E) with QPSK modulation.

In the same way, Figure 5.6(b) depicts the BER performance of the aforementioned systems when $\rho_1 = 15$ dB. The same conclusions hold for this framework where we can state that the difference between non-linear systems in comparison with *Lin-MMSE* is lower. Thus, ρ_1 has more impact over non-linear precoding approaches.

5.3.2 Achievable Sum-Rate

The expressions for achievable sum-rate can be easily adapted to multiuser MIMO relaying with non-linear precoding techniques such as

$$C = \frac{1}{2} \sum_{k=1}^K \log_2 \left(|\mathbf{I}_{N_k} + \mathbf{R}_{\mathbf{n}_k}^{-1} \mathbf{H}_{2,k} \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}} \mathbf{R}_{\mathbf{x}} \overline{\mathbf{F}}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_{2,k}^H| \right), \quad (5.18)$$

where $\mathbf{H}_{2,k}$ stands for the channel related to the k^{th} user while $\overline{\mathbf{W}} = \beta_2 \mathbf{W}$ and $\overline{\mathbf{F}} = \beta_1 \mathbf{F}$ denote the unscaled relaying and precoding matrices, respectively. The matrix $\mathbf{R}_{\mathbf{x}} \in \mathbb{C}^{N \times N}$ is the precoded signal covariance matrix. Whilst for THP precoding the approximation $\mathbf{R}_{\mathbf{x}} = \mathbf{R}_{\mathbf{v}} = \text{diag}(1, \tau^2/6, \dots, \tau^2/6)$ is employed, for VP non-linear precoding $\mathbf{R}_{\mathbf{x}} = \mathbf{R}_{\mathbf{v}} = \mathbb{E}[\mathbf{d}\mathbf{d}^H]$ is unknown and approximations cannot be used for this proposal. Once the perturbation vector is obtained, the input covariance matrix is calculated by simulations. The term $1/2$

comes from the half-duplex communication while $\mathbf{R}_{n_k} = \beta_1^2 N_{01} \mathbf{H}_{2,k} \overline{\mathbf{W}\mathbf{W}^H} \mathbf{H}_{2,k}^H + \beta_2^2 N_{02} \mathbf{I}_{N_k}$ stands for the noise covariance matrix.

The achievable sum-rate, obtained by simulations, has been tested for fixed $\rho_2=15$ dB and $\rho_1 = 15$ dB as shown in Figures 5.7(a) and 5.7(b), respectively for set-up A. The results for set-up E, have been omitted owing to the conclusions obtained, which match up with the ones for set-up A.

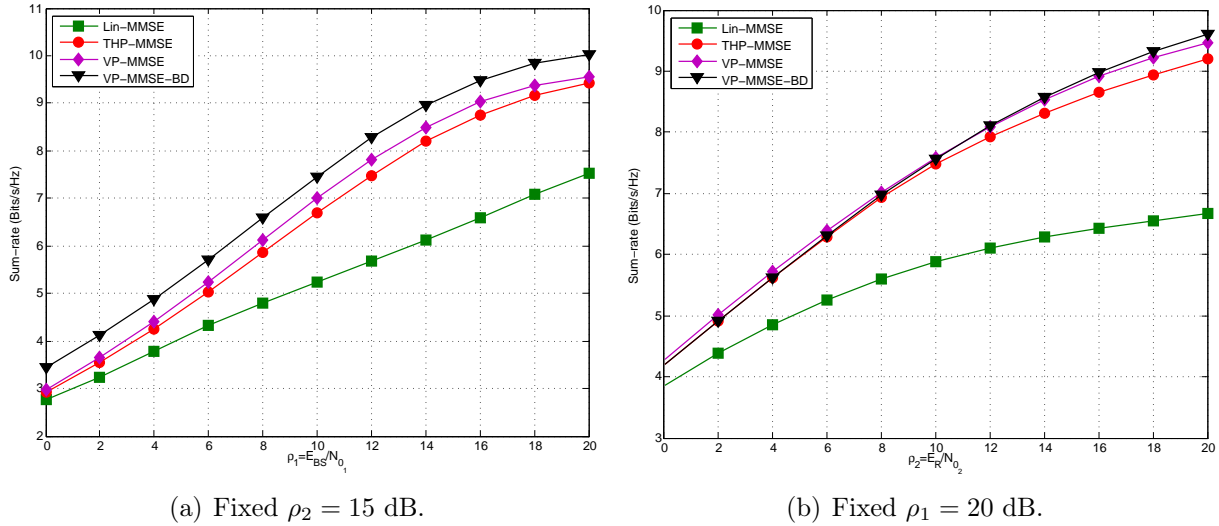


Figure 5.7: Sum-rate achieved by non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

As expected, non-linear systems *THP-MMSE*, *VP-MMSE* and *VP-MMSE-BD* achieved a better sum-rate compared with the linear approach, improving the performance considerably. Moreover, the best sum-rate is achieved by *VP-MMSE-BD*. Figure 5.7(a) shows that the capacity grows with the SNR, being the maximum sum-rate achieved at $\rho_1=20$ dB with 9.3 Bits/s/Hz for *THP-MMSE*, 9.5 Bits/s/Hz for *VP-MMSE* and finally 10 Bits/s/Hz for *VP-MMSE-BD*.

Same conclusions hold for a fixed $\rho_1 = 15$ dB, being the difference between *VP-MMSE-BD* and *VP-MMSE* lower. Until $\rho_2 = 10$ dB is reached *VP-MMSE* performs slightly better than *VP-MMSE-BD*. As it happens with the previous figure, the sum-rate increases with SNR being the maximum achieved by *VP-MMSE-BD* 9.75 Bits/s/Hz at $\rho_2 = 20$ dB. It should be pointed out that at low SNR, from $\rho_2 = 0$ to $\rho_2 = 6$ dB, the achieved sum-rate by *VP-MMSE-BD* matches up with *THP-MMSE*.

When the noise of the first and second hop are low, the achievable sum-rate is worst compared with the performance obtained when the first hop noise is low and the second hop noise takes a medium value. On the contrary, the performance is similar when ρ_1 takes a medium value and ρ_2 is high. Hence, the noise of the first hop limits the performance of the system.

5.3.3 Convergence

In this section, the convergence analysis is carried out. As described previously, due to the dependence of the precoding and relaying matrices, an iterative process is executed to get a joint local optimal solution. The convergence will give us an idea of the required number of iterations or the convergence speed of the algorithms.

Two different tests have been carried out. On one hand, bit-error-rate convergence has been tested subject to a fixed number of iterations. Therefore, the local optimal solution has been compared with the performance of the system forced to stop after a concrete number of iterations. Hence, after 1, 3, 6 and 10 iterations BER performance is measured. On the other hand, MSE convergence is evaluated considering the SNR in pairs, being the pairs:

- $\rho_1 = 10$ dB and $\rho_2 = 15$ dB ($\rho_1 < \rho_2$).
- $\rho_1 = 20$ dB and $\rho_2 = 15$ dB ($\rho_1 > \rho_2$).

Figures 5.8(a), 5.8(b) and 5.8(c) show the performance of *THP-MMSE*, *VP-MMSE* and *VP-MMSE-BD* approaches, respectively, when ρ_2 is set to 15 dB and the iterative algorithm is forced to stop after 1, 3, 6 and 10 iterations (*it*).

The figures depicted in Figure 5.8 show how increasing the number of iterations improves the BER performance. Furthermore, it can be observed that oppositely to *Lin-MMSE*'s convergence (see Figure 3.6), more than 10 iterations are required to fall near the local optimal. Thus, this clearly might increase the complexity of the system.

On the other hand, Figures 5.9(a) and 5.9(b) show the MSE convergence of the aforementioned systems when $\{ \rho_1 = 10$ dB and $\rho_2 = 20$ dB ($\rho_1 < \rho_2$) $\}$ and $\{ \rho_1 = 20$ dB and $\rho_2 = 15$ dB ($\rho_1 > \rho_2$) $\}$. By means of these results we will be able to stop the the algorithm after a certain number of iterations when the convergence is reached.

As depicted in Figure 5.9, all the iterative non-linear systems converge in a few number of iterations being *THP-MMSE* the one that reduces more the error from iteration 1 to 2. The difference gap between iterations of VP systems is less pronounced. Furthermore, when $\rho_1 < \rho_2$, which means that the first hop noise is greater than the second hop noise, the error obtained is greater than in the opposite case, i.e the maximum error obtained for *THP-MMSE* system is around 1.1 for $\rho_1 = 10$ dB, while for $\rho_1=20$ dB it decreases to 0.43.

5.4 Complexity Analysis

The complexity of the local optimal non-linear systems will be here analysed using two measurements: complexity order and computational complexity. The former, introduced in the complexity analysis at the third chapter, will show the complexity order of the algorithm. The latter, measured in FLOPs, defines the number of operations.

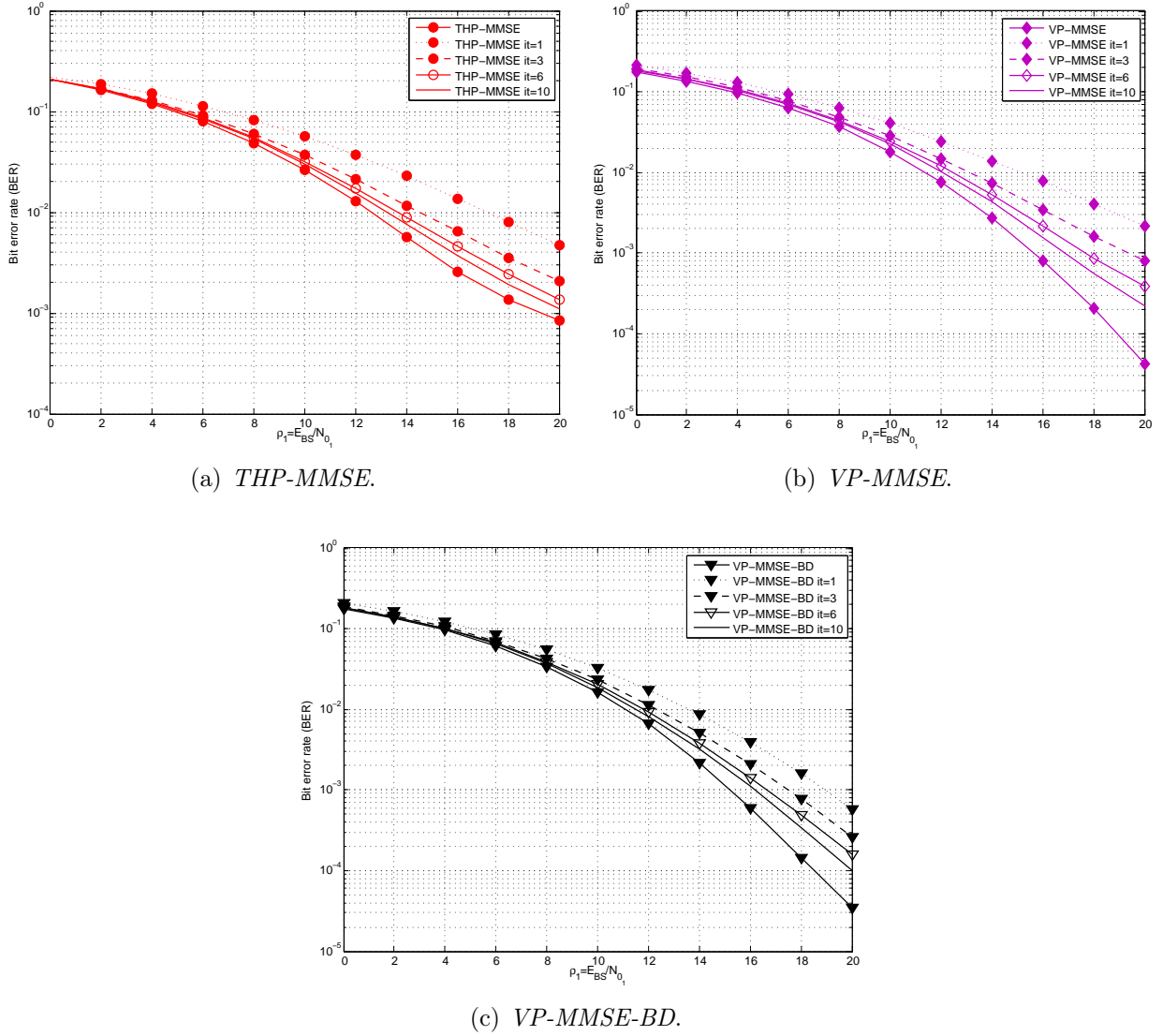


Figure 5.8: Comparison of BER vs. SNR for non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when they are forced to stop after a certain number of iterations.

As shown in Chapter 3, the complexity of the algorithms will also be given in terms of running-time, which is given as the addition of the time needed for the computation at the BS, RS and the users. Thus, it includes the addition of the time needed for the computation of the precoding matrix at the BS, the search of the perturbation vector, the computation of the relaying matrix, and finally, the modulo operation carried out at the end user terminals.

In order to simplify the analysis, $M = R = N = n$ is assumed. The term N_B stands for the length of transmitted data block and $T = |\mathcal{V}|$ denotes the total number of points that form the lattice. For simplicity, a fixed number of iterations (*it*) is assumed.

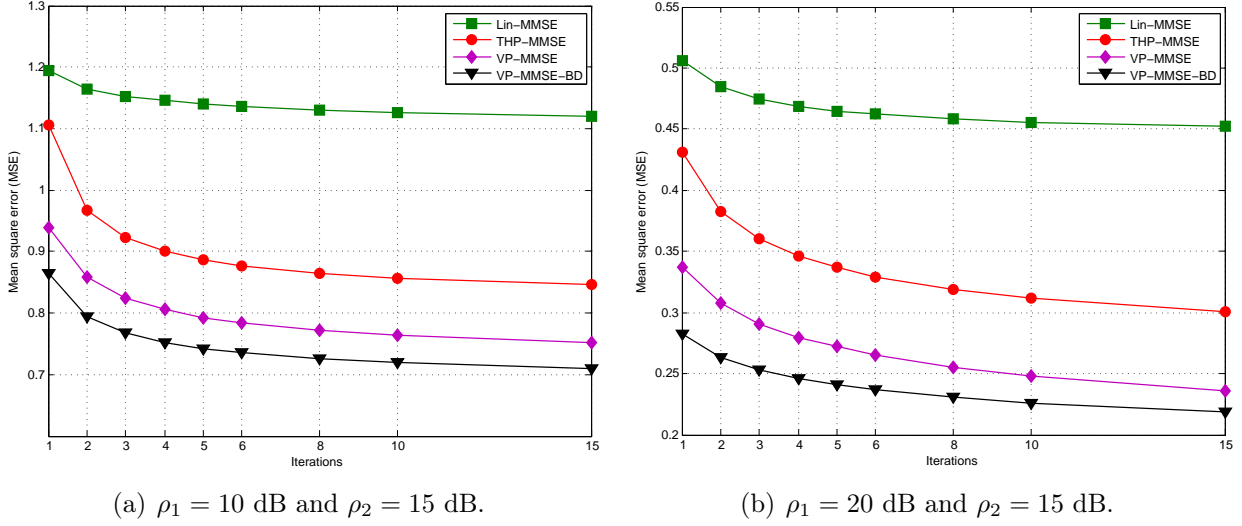


Figure 5.9: Error obtained by non-linear precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

5.4.1 Number of FLOPs and Complexity Order Analysis

In this section the complexity of the aforementioned non-linear approaches is analysed in terms of complexity order and FLOPs. Table 5.1 represents the number of FLOPs required for the computation of the precoding and relaying matrices, as well as for the computation of the feedback filter employed in *THP-MMSE*.

| Algorithm | Operation | FLOPs |
|-------------------|----------------------|---|
| THP-MMSE | F | $4M^3 + (3+2N+8M)R^2 + (1+2M)N^2 + (4R+9)M^2 + (2R+1)MN + (3+5R)M$ |
| | B | $4R^3N + (5K+2N^2)R^2 + (20+8M)N^3 + (2M^2 + 6RM + 4M-5)N^2 + 4M^3N + (2RN+12N)M^2 + 8N + 3MN + 3RMN + 1$ |
| | W | $10R^3 + (6N+4M+1)R^2 + (2R+1)N^2 + (3+6M)NR + 2MR + 1$ |
| VP-MMSE | F | $4M^3 + (4R+9)M^2 + (3+2N+8M)R^2 + (5+N)RM + (3+N)M$ |
| | W | $10R^3 + (4N+4M)R^2 + (6N+2)MR + 2NR + 1$ |
| VP-MMSE-BD | F | $2N^3 + 4M^3 + (2+M)N^2 + (2R+4N+7)M^2 + (4R+4)NM + (3+2R)M$ |
| | W₁ | $4R^3 + (4N+2M+8)R^2 + (10+2M)N^2 + (3+8M)NR + (3+M+2R)N + (3+M)R + 1$ |

Table 5.1: Number of FLOPs necessary for matrix computation by *THP-MMSE*, *VP-MMSE* and *VP-MMSE-BD* approaches.

In the same way, Table 5.2 shows the number of FLOPs necessary for the computation of the SE algorithm [Mohaisen11] and the modulo operation executed at each user terminal. In [Mohaisen11] the complexity of SE is analysed for multiuser MIMO multi-antenna systems.

| Operation | FLOPs |
|------------------|---|
| Sphere encoder | $\left(\frac{T^{N+1}-T}{T-1} + N\right) N_B$ [Mohaisen11] |
| Modulo Operation | NN_B |

Table 5.2: Number of FLOPs for sphere encoder and modulo operation.

Watching carefully Tables 5.1 and 5.2, the number of FLOPs depend on the number of transmit and relaying antennas (M and R), the total number of receive antennas (N), the amount of points of the lattice (T) as well as the length of the transmitted data block.

In addition, Table 5.3 depicts the complexity order and the total number of FLOPs executed for the convergence of each algorithm. As described in Table 5.3, VP-based systems obtain the highest complexity order, which depends on the number of iterations executed and the number of points that form the lattice.

| Algorithm | Comp. order | FLOPs | | |
|-------------------|-----------------|---|---------------------|--------------------------|
| | | M=R=N=n | M=R=N=4 | M=R=N=8 |
| THP-MMSE | $O(it n^4)$ | $it (28n^4 + 101n^3 + 40n^2 + 11n + 9 + N_B (6n^2 - 4n + 1))$ | $22 \times 10^3 it$ | $207 \times 10^3 it$ |
| VP-MMSE | $O(it T^{n+1})$ | $it (56n^3 + 35n^2 + 3n + 10 + N_B (\frac{T^{n+1}-T}{T-1} + n))$ | $40 \times 10^6 it$ | $1.58 \times 10^{13} it$ |
| VP-MMSE-BD | $O(it T^{n+1})$ | $it (71n^3 + 26n^2 + 13n + 10 + N_B (\frac{T^{n+1}-T}{T-1} + n))$ | $40 \times 10^6 it$ | $1.58 \times 10^{13} it$ |

Table 5.3: Complexity order of *THP-MMSE*, *VP-MMSE* and *VP-MMSE-BD*.

Finally, it should be pointed out that the complexity of the non-linear systems is greater compared with the linear approach introduced in Chapter 3 and, obviously, grows with the number of antennas, being the most complex approach *VP-MMSE-BD* system due to block diagonalization process.

5.4.2 Running-Time Analysis

Figure 5.10 shows the run-time needed for the computation of the non-linear algorithms when ρ_2 is fixed to 15 dB. It can be shown that the highest run-time stands for the *VP-MMSE-BD*

set-up, largely due to the block diagonalization process and the vector perturbation search. Obviously all the systems require larger run-times than the linear MMSE approach.

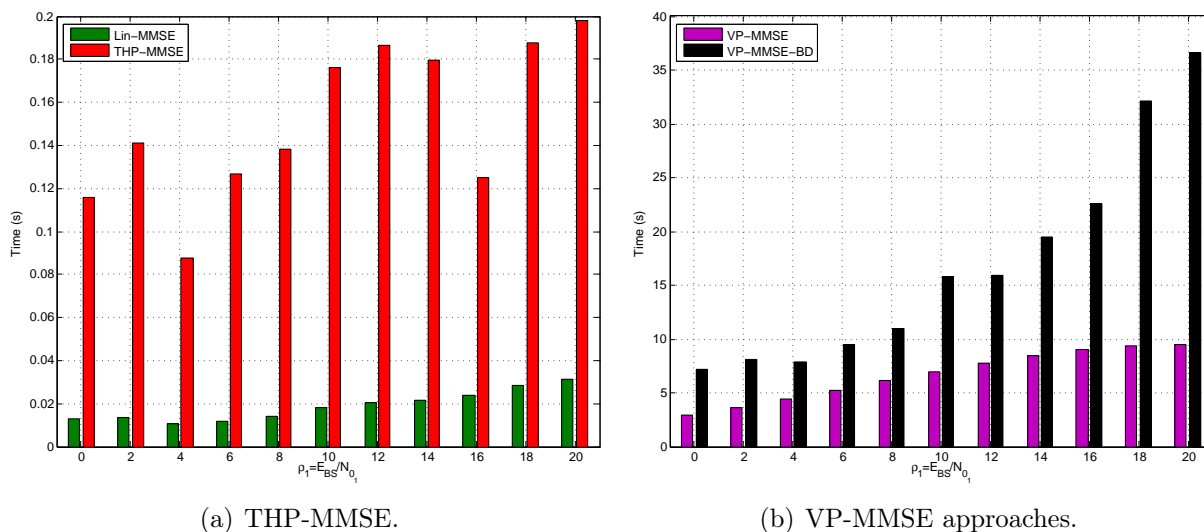


Figure 5.10: Simulation time required for *THP-MMSE*, *VP-MMSE* and *VP-MMSE-BD* computation for $4 \times 4 \times \{4 \times 1\}$ set-up when $\rho_2 = 15$ dB.

As shown in Figure 5.10(b) the run-time related to the joint local optimal BD-VP, called *VP-MMSE-BD* grows with ρ_1 . Therefore, the maximum run-time is obtained when $\rho_1 = 20$ dB, needing 37 seconds per Monte Carlo simulation and the minimum is got when $\rho_1 = 0$ dB, requiring 7 seconds per Monte Carlo simulation. The same conclusion holds for the case *VP-MMSE* where the run-time increases with the variable signal-to-noise ratio needing a maximum of 9 seconds when $\rho_1 = 20$ dB and a minimum of 3 seconds when $\rho_1 = 0$ dB. However, for the *THP-MMSE* approach the time does not grow with ρ_1 . For example this algorithm needs 0.2 seconds per Monte Carlo simulation when $\rho_1 = 20$ dB and a minimum of 0.09 seconds for $\rho_1 = 4$ dB.

Unsurprisingly, the performance obtained for VP based designs is better compared to *THP-MMSE* approach. However, the time required for some SNR values is sometimes intractable. For example, as a simple comparison, the period of time needed to get the maximum value of *VP-MMSE-BD* approach in set-up A is 4 days whereas the maximum time needed for the execution of 10^4 Monte Carlo simulations for *THP-MMSE* is 35 minutes.

5.5 Chapter Summary

This chapter has focused on the combination of non-linear precoding and the optimal design of precoding and relaying matrices, proposing three different approaches under MMSE criterion.

The chapter starts proposing THP transmission and the joint design of the matrices. This precoding technique has been applied at the BS to cancel the interference created from the BS to the end users.

Despite its better performance, VP has not been considered for multiuser MIMO relaying systems, owing to the computational complexity that presents. The local optimal VP solution has been given for the joint design of the matrices, showing the effectiveness of this precoding technique over linear and THP precoding.

The third approach has combined BD and VP, where BD has been applied at the relaying terminal for interference cancellation, whilst VP has been employed at the BS for mean square error minimization and power limiting.

Provided simulation results have shown the effectiveness of non-linear precoding in multiuser MIMO relaying systems. Linear precoding solutions, presented in previous chapters, have been outperformed by the novel proposals introduced in this chapter. However, in the end of the chapter, a complete complexity analysis has shown the main drawback of the aforementioned novel proposals: the high computational complexity.

The contributions of this chapter have derived on the first application of VP technique to multiuser MIMO AF relaying systems, a novel proposal that combines VP and BD, as well as the joint design of the matrices for THP precoding. Several set-ups have tested these approaches, demonstrating the effectiveness of the systems by means of different simulations. Furthermore, a complete computational cost analysis is provided.

Suboptimal Non-Linear Approaches

The effectiveness of non-linear precoding techniques for multiuser MIMO environment has been shown in Chapter 5. Nevertheless, these non-linear precoding techniques present a great disadvantage: the computational cost, which is higher on one hand, due to the feedback loop and the matrix optimization accomplished in each iteration in THP precoding, or on the other hand, as a consequence of the search of the perturbation vector and the design of the matrices in VP. Furthermore, the search of the perturbation vector requires the execution of sphere encoding algorithm, adding more complexity to the system.

As a consequence of the increased complexity, which grows with the number of antennas and multi-antenna users, the implementation in real time systems is uneconomic or just prohibitive.

As carried out for linear precoding, suboptimal approaches are proposed in this chapter for AF relaying communication frameworks based on non-linear precoding techniques. The main purpose of these non-linear precoding proposals is the reduction of the computational cost presented by the optimal non-linear precoding approaches introduced in Chapter 5 with a low degradation in performance.

Three different approaches are proposed here: Firstly, suboptimal multiuser MIMO THP relaying is analysed, where the relaying matrix is designed for the minimization of the second hop MSE. Once the relaying and the scaling factor are obtained, feedback and feedforward filters are designed to minimize the overall error of the system.

Secondly, a suboptimal VP multiuser MIMO relaying system is proposed based on the independent optimization of each hop transmission. This way, by means of the reduction of the MSE of the first hop, the precoding matrix and the scaling factor at the BS are obtained, whereas throughout the reduction of the second hop error the relaying matrix and the scaling factor at the second hop are computed. Once the matrices are derived, the perturbation vector that minimizes the total error and constrain the power at BS and RS terminals is searched.

Finally, another suboptimal proposal is suggested based on VP precoding, where the optimization problem is divided into a master problem and subproblem, making the optimization of the complete problem easier. Whilst the precoding and relaying matrices are

derived from the solution of the subproblem, the perturbation vector is obtained from the optimization of the master problem.

6.1 Suboptimal Design for Tomlinson-Harashima Precoding

Chapter 5 has introduced the local optimal THP approach, outperforming the local linear optimal approach presented in Chapter 3 and showing the effectiveness of THP, making it a candidate for multiuser MIMO relaying systems.

As pointed out before, the computational complexity increases due to the iterative process that has to be applied for matrix derivation and the cost of the precoding process itself. The main aim of the suboptimal approach presented in this section is to divide the problem, avoiding in this way the iterative process.

The design is divided into two steps: Firstly, the second hop MSE is minimized under a power limit at the relay in order to obtain an independent solution for the relaying matrix. After that, the feedback and feedforward filters as well as the scaling factor at the first hop are computed by minimizing the overall MSE under the constraint of the output power at the BS.

By means of this suboptimal approach, THP can be easily implemented in multiuser MIMO relaying systems. Moreover, this proposal would be very well suited to multihop systems, where more than one relay are used from the transmitter to the receivers. For this configuration, relaying processing matrices can be easily obtained if the channel is known at the terminal. For example, the relaying matrix concerning to the r^{th} relaying terminal will be designed for the minimization of the MSE of the r^{th} hop and to limit the power at that terminal instead of considering the overall system.

The system is the one described in Section 5.1 and depicted in Figure 5.1, where an M -antenna BS sends data to K single-antenna users through a relaying terminal equipped with R antennas, where $R = K$ is here assumed for simplicity. Furthermore, equalization process is not assumed for this approach.

As it has been previously said, the relaying processing matrix is first obtained as depicted in the first part of Appendix E. Before starting with the analysis, it should be recalled that the relay receives the data vector \mathbf{y}_1 , which is filtered, transmitted and received at the final users as

$$\hat{\mathbf{d}} = \beta_2 \mathbf{H}_2 \mathbf{W} \mathbf{y}_1 + \beta_2 \mathbf{n}_2,$$

being $\mathbf{W} \in \mathbb{C}^{R \times R}$ and β_2 the relaying matrix and the power scaling factor, respectively. These parameters are optimized in order to minimize MSE of the second hop, being the

optimization problem the following:

$$\begin{aligned} \{\mathbf{W}, \beta_2\} &= \underset{\{\mathbf{W}, \beta_2\}}{\operatorname{argmin}} \quad \xi_2 \\ \text{s.t.} \quad &\operatorname{Tr}(\mathbf{W}\mathbf{R}_{\mathbf{y}_1}\mathbf{W}^H) = E_R, \end{aligned}$$

where $\mathbf{R}_{\mathbf{y}_1} = \mathbb{E}[\mathbf{y}_1\mathbf{y}_1^H] \in \mathbb{C}^{R \times R}$ denotes the covariance matrix of the signal received at the relay. On the other hand, ξ_2 stands for the MSE of the second hop, computed as

$$\begin{aligned} \xi_2 &= \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{y}_1\|_2^2 \right], \\ &= \operatorname{Tr} \left(\beta_2^2 \mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H \mathbf{H}_2^H + \beta_2^2 N_{0_2} \mathbf{I}_K - 2\beta_2 \Re(\mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1}) + \mathbf{R}_{\mathbf{y}_1} \right). \end{aligned}$$

After solving the optimization problem, for which a complete description is given in Appendix E, the following expression is obtained:

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} K}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H. \quad (6.1)$$

In the same way, the scaling factor that limits the power at the relay to E_R is

$$\beta_2 = \sqrt{\frac{\operatorname{Tr} \left(\left(\mathbf{H}_1 \bar{\mathbf{F}} \mathbf{R}_{\mathbf{v}} \bar{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{0_1} \mathbf{I}_R \right) \mathbf{H}_2 \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} K}{E_R} \mathbf{I}_R \right)^{-2} \mathbf{H}_2^H \right)}{E_R}},$$

being $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$ the unscaled version of the feedforward filter, β_1 the power scaling factor of the first hop and, finally, $\mathbf{R}_{\mathbf{v}} = \operatorname{diag}(1, \tau^2/6, \dots, \tau^2/6) \in \mathbb{R}^{K \times K}$ denotes the precoded symbols' covariance matrix, being τ the modulo constant.

Once the relaying matrix \mathbf{W} and the scaling factor β_2 are obtained, the parameters of the transmitter are set, being the optimization problem for the first hop parameters to be solved defined as

$$\begin{aligned} \{\mathbf{F}, \beta_1, \mathbf{B}\} &= \underset{\{\mathbf{F}, \mathbf{B}, \beta_1\}}{\operatorname{argmin}} \quad \xi \\ \text{s.t.} \quad &\operatorname{Tr}(\mathbf{F}\mathbf{R}_{\mathbf{v}}\mathbf{F}^H) = E_{BS}, \\ &\mathbf{S}_k \mathbf{B} \mathbf{e}_k = \mathbf{0}_k \quad k = 1, \dots, K, \end{aligned}$$

which is subject to the power constraint and the triangularization imposed by the THP feedback filter. As pointed out at THP background section (see Section 5.1), the feedback filter \mathbf{B} is a lower triangular matrix with zero elements in the main diagonal employed for interference cancellation at the feedback loop. Matrix $\mathbf{S}_k = \mathbf{S}_{(0,k,K-k)} = [\mathbf{I}_k, \mathbf{0}_{k \times K-k}] \in \{0, 1\}^{k \times K}$ denotes a selection matrix that cuts out the first k elements of a K -dimensional

vector, whereas the vector \mathbf{e}_k stands for the k^{th} column of an identity matrix. The criterion used for matrix derivation is again MSE minimization, defined in this case as

$$\begin{aligned}\xi &= \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{d}\|_2^2 \right] \\ &= \text{Tr} \left(\beta_1^2 \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{R}_v \mathbf{F}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H + \beta_1^2 N_{0_1} \mathbf{H}_2 \overline{\mathbf{W}} \overline{\mathbf{W}}^H \mathbf{H}_2^H + \beta_2^2 N_{0_2} \mathbf{I}_K \right. \\ &\quad \left. - 2\beta_1 \Re \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{R}_v (\mathbf{I}_K - \mathbf{B})^H \right) + (\mathbf{I}_K - \mathbf{B}) \mathbf{R}_v (\mathbf{I}_K - \mathbf{B})^H \right),\end{aligned}$$

where $\mathbf{R}_v = \text{diag}(1, \sigma_v^2, \dots, \sigma_v^2) \in \mathbb{R}^{K \times K}$ is the covariance matrix of the precoded symbols, assuming $\sigma_v^2 = \tau^2/6$ [Joham06] and $\overline{\mathbf{W}} = \beta_2 \mathbf{W}$.

By means of the procedure depicted in Appendix E, the expression for the precoding matrices and the scaling factor are obtained, which are the following:

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 + \frac{N_{0_1}}{E_{BS}} \text{Tr} \left(\overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \overline{\mathbf{W}} \right) \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{e}_k \mathbf{e}_k^T,$$

$$\mathbf{B} = \sum_{k=1}^K (\mathbf{S}_k^T \mathbf{S}_k - \mathbf{I}_K) \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}} \mathbf{e}_k \mathbf{e}_k^T,$$

and finally

$$\beta_1 = \sqrt{\frac{\sum_{k=1}^K \sigma_v^2 \mathbf{e}_k^T \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{A}^{-2} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{e}_k}{E_{BS}}},$$

being $\mathbf{A} = \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 + \frac{N_{0_1}}{E_{BS}} \text{Tr} \left(\overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \overline{\mathbf{W}} \right) \mathbf{I}_M$.

By means of this novel suboptimal design, the dependence of the filters in the joint local optimal THP design is broken. Hence, the iterative algorithm needed for matrix computation is avoided, reducing in this way the computational cost. Algorithm 10 summarizes the proposed suboptimal THP algorithm, being the precoding algorithm described in Algorithm 8.

6.1.1 Simulation Results

The performance of the suboptimal THP approach (*THP-MMSE-Sub*) is here evaluated, comparing it with *Lin-MMSE* and *THP-MMSE* approaches. Simulation parameters employed are specified in Section 5.3. It should be underlined that only the set-up $4 \times 4 \times \{4 \times 1\}$ is tested as it has been made for the optimal approaches in Section 5.3. It should be recalled that the simulation parameters are specified at the section cited previously.

Algorithm 10 Computation of $\overline{\mathbf{F}}$, $\overline{\mathbf{W}}$ and \mathbf{B} under MMSE criterion for suboptimal THP approaches.

- 1: Compute $\overline{\mathbf{W}}$ defined in (6.1).
 - 2: Definition of the equivalent channel $\mathbf{H}_e = \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1$.
 - 3: Computation of α_1 .
 - 4: Feedforward filter \mathbf{F} derivation:
 - 5: **for** $k = 4$ to 1 **do**
 - 6: $\mathbf{F} = (\mathbf{H}_e \mathbf{H}_e^H + \alpha_1 \mathbf{I}_K)^{-1}$.
 - 7: Definition of the precoder related to the user $k \implies \mathbf{f}_k = \mathbf{H}_e^H \mathbf{F} \mathbf{e}_k$.
 - 8: Remove user k from the subset of users $\longrightarrow \mathbb{O} \longleftarrow \mathbb{O}/\{k\}$.
 - 9: Equivalent channel computation with the removed user: $\mathbf{H}_e = (\mathbf{I}_K - \mathbf{e}_k \mathbf{e}_k) \mathbf{H}_e$.
 - 10: **end for**
 - 11: Feedback filter \mathbf{B} computation:
 - 12: **for** $k = 1$ to K **do**
 - 13: $\mathbf{b}_k = (\mathbf{S}_k^T \mathbf{S}_k - \mathbf{I}_K) \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{f}_k$.
 - 14: **end for**
 - 15: Obtain β_1 .
 - 16: $\mathbf{F} = (1/\beta_1) [\mathbf{f}_1, \dots, \mathbf{f}_K]$.
-

To conclude, in order to show the effectiveness of the suboptimal *THP-MMSE-Sub* approach, BER performance, the achievable sum-rate and the MSE error have been measured to show the effectiveness of the suboptimal *THP-MMSE-Sub* approach.

6.1.1.1 BER Performance

In this section, the BER performance of the proposed suboptimal THP approach named *THP-MMSE-Sub* is given and compared with the joint local optimal approaches *Lin-MMSE* and *THP-MMSE*. Figures 6.1(a) and 6.1(b) show bit error rate performance when ρ_2 and ρ_1 are set to 15 dB, respectively.

Briefly, Figure 6.1(a) shows that *THP-MMSE-Sub* improves *Lin-MMSE*'s performance, being the maximum gain about 2.5 dB. Obviously, as expected, *THP-MMSE* obtains the better performance being the difference respect to *THP-MMSE-Sub* about 4.5 dB. Until 10 dB, *Lin-MMSE* performs slightly better compared with *THP-MMSE-Sub*. It should be underlined that the most important feature of *THP-MMSE-Sub* is its reduced computational complexity instead of its performance, which also improves linear's. It should be underlined that *THP-MMSE-Sub* does not execute any iterative process to obtain precoding and relaying matrices.

In addition, ρ_1 is fixed to 15 dB as depicted in Figure 6.1(b), obtaining the same conclusions. In this case, up to 8 dB, *Lin-MMSE* performs slightly better compared with *THP-MMSE-Sub*. After 8 dB, the gain obtained by *THP-MMSE-Sub* respect to the linear approach is greater, being the maximum gain of about 5 dB. Provided simulation results show that the first hop noise has a greater effect on the system limiting the performance of

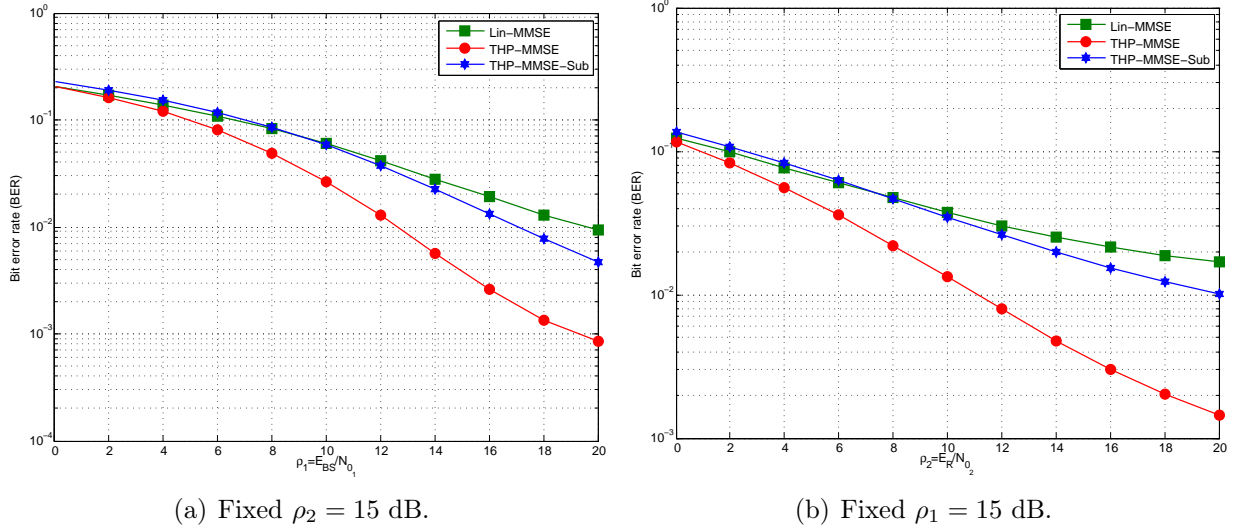


Figure 6.1: BER performance versus SNR for MMSE non-linear suboptimal THP precoding technique in the $4 \times 4 \times \{4 \times 1\}$ with QPSK modulation.

the system.

6.1.1.2 Achievable Sum-Rate

This section shows the sum-rate achieved for the set-up $4 \times 4 \times \{4 \times 1\}$ computed using Monte Carlo simulations. Firstly the capacity is tested when $\rho_2 = 15$ dB (Figure 6.2(a)) and $\rho_1 = 15$ dB (Figure 6.2(b)). For fixed $\rho_2 = 15$ dB, *THP-MMSE-Sub* proposal performs better than the local optimal approach *Lin-MMSE*, specially at high SNR. At low ρ_1 , from 0 to 6 dB, *Lin-MMSE* performs slightly better compared with *THP-MMSE-Sub*.

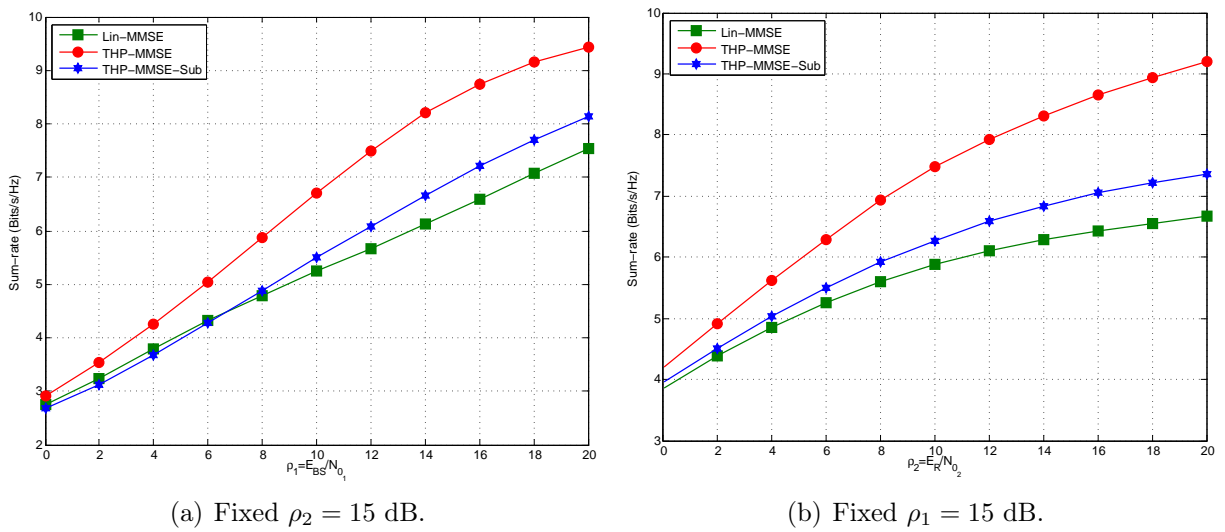


Figure 6.2: Sum-rate versus SNR of *THP-MMSE-Sub* approach in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

Figure 6.1(b) depicts the sum-rate achieved when ρ_1 is fixed to 15 dB, where the same conclusions hold. Furthermore, the difference between *THP-MMSE* and *THP-MMSE-Sub* is greater.

6.1.1.3 Mean Square Error

Once the results in terms of BER and SR performance are given, the overall MSE error of the system is measured. Briefly, the system is forced to stop after carrying out a certain number of iterations, measuring the obtained error. *THP-MMSE-Sub* proposal, designed to reduce the computational complexity of *THP-MMSE*, does not execute any iterative process. Hence, the error is plotted constant over all the iterations.

Figure 6.3(a) and 6.3(b) show the overall error obtained when $\{\rho_1 = 10 \text{ dB and } \rho_2 = 15 \text{ dB}\}$, and $\{\rho_1 = 20 \text{ dB and } \rho_2 = 15 \text{ dB}\}$, respectively.

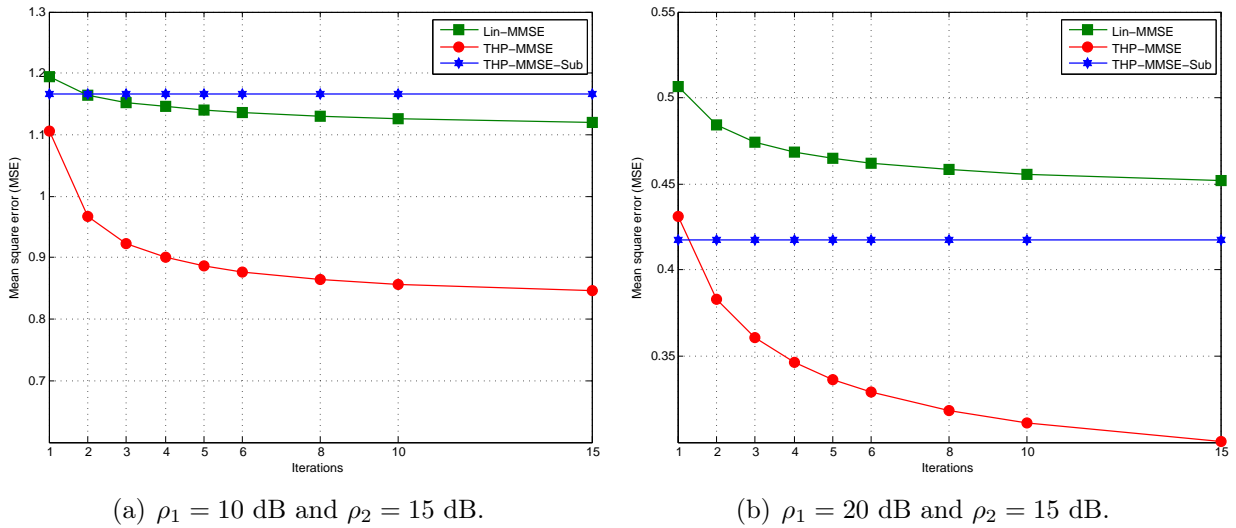


Figure 6.3: Mean square error of MMSE non-linear suboptimal THP precoding technique in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

As it is shown in Figure 6.3(a), when $\rho_1 < \rho_2$, after 2 iterations the error obtained with *Lin-MMSE* is lower compared with the one obtained for *THP-MMSE-Sub* approach. Nevertheless the gap is not remarkable.

Finally Figure 6.3(b) shows the error obtained when $\rho_1 = 20 \text{ dB} > \rho_2 = 15 \text{ dB}$ where the MSE is reduced considerably by means of *THP-MMSE-Sub* approach, improving linear's performance. We observe that the error is reduced significantly when the SNR increases.

6.2 Suboptimal Vector Precoding Designs

Non-linear VP precoding strategy has shown its effectiveness in both multiuser MIMO and multiuser MIMO relaying systems adding a perturbation vector to the transmitted symbols.

Despite its better performance, VP presents a huge disadvantage. For iterative approaches, the perturbation vector has to be derived in each iteration, i.e. a sphere encoder is executed in every iteration, increasing considerably the computational cost of the system.

In order to avoid the complexity of the optimal VP-based design introduced in Section 5.2, two suboptimal VP solutions are proposed in this section. Firstly, a simple hop-by-hop MMSE optimization is proposed where the MSE of each hop is minimized independently and VP is applied at the BS. On the other hand, the second proposal simplifies the MSE optimization problem by dividing it into a master problem and subproblem. While \mathbf{F} and \mathbf{W} are optimized jointly for MSE minimization in the subproblem, the master problem searches the optimal perturbation vector for the already derived matrices.

Simulation results will show the effectiveness of the proposed suboptimal approaches, which outperform the linear joint local optimal and THP suboptimal approaches. In spite of the penalty that this suboptimal approaches suffer in performance, the computational complexity is considerably reduced in comparison to the joint optimal VP approach.

6.2.1 Hop-by-Hop MMSE with VP Transmission

The proposed hop-by-hop approach, named *VP-MMSE-HH*, searches the precoding and relaying matrices minimizing the mean square error of each hop independently. This approach, also applied in Section 4.2 for suboptimal design with linear transmission, here includes a perturbation vector which minimizes the overall MSE while ensuring that the output power at the BS and RS terminals is E_{BS} and E_R , respectively.

As opposed to the joint iterative approach introduced in Section 5.2, this approach obtains a simple solution, where precoding and relaying matrices are independent, at the cost of introducing some performance degradation. Firstly, for the precoding matrix derivation, the optimization problem is defined as

$$\begin{aligned} \{\mathbf{F}, \beta_1\} &= \underset{\{\mathbf{F}, \beta_1\}}{\operatorname{argmin}} \quad \xi_1 \\ \text{s.t.} \quad & \|\mathbf{y}_{BS}\|_2^2 = E_{BS}, \end{aligned}$$

being \mathbf{y}_{BS} and β_1 the precoded symbols and the power scaling factor that limits the power at the BS to E_{BS} , respectively. The term ξ_1 denotes the MSE of the first hop, defined as:

$$\begin{aligned} \xi &= \mathbb{E} [\|\mathbf{y}_1 - \mathbf{d}\|_2^2] \\ &= \beta_1^2 \mathbf{y}_{BS}^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{y}_{BS} + \beta_1^2 N_{01} R - 2\Re(\beta_1 \mathbf{d}^H \mathbf{H}_1 \mathbf{y}_{BS}) + \mathbf{d}^H \mathbf{d}, \end{aligned}$$

where \mathbf{y}_1 stands for the received signal at the relay and $\mathbf{d} = \mathbf{s} + \mathbf{a}$ denotes the perturbed symbols, being \mathbf{a} the perturbation vector.

Carrying out the procedure described in Section B.2 and assuming $R = K$, the expres-

sions for the precoding matrix and scaling factor β_1 are the following:

$$\mathbf{y}_{BS} = \frac{1}{\beta_1} \underbrace{\left(\mathbf{H}_1^H \mathbf{H}_1 + \frac{N_{0_1} R}{E_{BS}} \mathbf{I}_M \right)^{-1}}_{\mathbf{F}} \mathbf{H}_1^H \mathbf{d}, \quad (6.2)$$

and

$$\beta_1 = \sqrt{\frac{\mathbf{d}^H \mathbf{H}_1^H \left(\mathbf{H}_1 \mathbf{H}_1^H + \frac{N_{0_1} R}{E_{BS}} \mathbf{I}_M \right)^{-2} \mathbf{H}_1 \mathbf{d}}{E_{BS}}}.$$

The precoding matrix expression defined in (6.2) is a regularized version of the pseudo-inverse of the first hop channel \mathbf{H}_1 , where the regularized factor is $\frac{N_{0_1} R}{E_{BS}}$.

In the same fashion, the relaying processing matrix and the scalar factor β_2 that limits the power at the relay are derived for the second hop error minimization. To get a solution, the optimization problem is defined such as:

$$\begin{aligned} \{\mathbf{W}, \beta_2\} &= \underset{\{\mathbf{w}, \beta_2\}}{\operatorname{argmin}} \quad \xi_2 \\ \text{s.t.} \quad &\|\mathbf{y}_R\|_2^2 = E_R, \end{aligned}$$

where \mathbf{y}_R and \mathbf{W} denote the transmitted data vector by the relay and the relaying matrix, respectively. The term ξ_2 stands for the MSE of the second hop defined as

$$\begin{aligned} \xi_2 &= \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{y}_1\|_2^2 \right] \\ &= \beta_2^2 \mathbf{y}_1^H \mathbf{W}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{W} \mathbf{y}_1 + \beta_2^2 N_{0_2} K - 2\Re \left(\beta_2 \mathbf{y}_1^H \mathbf{H}_2 \mathbf{W} \mathbf{y}_1 \right) + \mathbf{y}_1^H \mathbf{y}_1, \end{aligned}$$

being \mathbf{H}_2 the second hop channel and $\hat{\mathbf{d}} = \beta_2 \mathbf{H}_2 \mathbf{W} \mathbf{y}_1 + \beta_2 \mathbf{n}_2$ the received symbols by the end users. The solution leads us to the next expressions for the relaying matrix and the scaling factor β_2 after following the steps described in Section B.2

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} K}{E_R} \mathbf{I}_K \right)^{-1} \mathbf{H}_2^H, \quad (6.3)$$

and

$$\beta_2 = \sqrt{\frac{\mathbf{d}^H \bar{\mathbf{F}}^H \mathbf{H}_1^H \mathbf{H}_2^H \mathbf{A}^{-2} \mathbf{H}_2 \mathbf{H}_1 \bar{\mathbf{F}} \mathbf{d} + \beta_1^2 N_{0_1} \operatorname{Tr} (\mathbf{H}_2 \mathbf{A}^{-2} \mathbf{H}_2)}{E_R}},$$

being $\mathbf{A} = \mathbf{H}_2 \mathbf{H}_2^H + \frac{N_{0_2} K}{E_R} \mathbf{I}_K$ and $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$. Once again, the relaying matrix defined in (6.3) denotes the regularized pseudo inverse of the second hop channel.

Once the precoding and relaying matrices are obtained, the precoding vector that limits

the power at the BS and the RS is selected as the nearest point in a lattice that minimizes the MSE between the source and the end users. After applying (6.2) and (6.3) into the end-to-end MSE defined as $\xi = \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{d}\|_2^2 \right]$, we get

$$\xi = \mathbf{d}^H \left(\mathbf{I}_K - \mathbf{H}_2 \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} K}{E_R} \mathbf{I}_K \right)^{-1} \mathbf{H}_2^H \mathbf{H}_1 \left(\mathbf{H}_1^H \mathbf{H}_1 + \frac{N_{0_1} R}{E_{BS}} \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \right) \mathbf{d}. \quad (6.4)$$

Comparing the MSE expression defined in (6.4) with the one employed for joint local optimal VP design presented in previous chapter, we observe that the MSE function is simpler, since it only depends on channel matrices \mathbf{H}_1 and \mathbf{H}_2 .

Carrying out the steps described in Section 5.2.2 and introduced in [Schmidt05], by means of the Cholesky factorization of the inner term of (6.4) we get:

$$\left(\mathbf{I}_K - \mathbf{H}_2 \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2} K}{E_R} \mathbf{I}_K \right)^{-1} \mathbf{H}_2^H \mathbf{H}_1 \left(\mathbf{H}_1^H \mathbf{H}_1 + \frac{N_{0_1} R}{E_{BS}} \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \right) = \mathbf{L}^H \mathbf{L}.$$

The perturbation vector is searched according to [Schmidt05] such as

$$\mathbf{a} = \underset{\mathbf{a}' \in \tau \mathbb{Z}^K + j\tau \mathbb{Z}^K}{\operatorname{argmin}} \quad \|\mathbf{L}(\mathbf{s} + \mathbf{a})\|_2^2.$$

As introduced previously in Section 5.2.2, the optimal precoding vector is obtained through a SE algorithm [Peel05], simplifying the complexity of the search. Provided simulation results will show the performance of the proposed algorithm. It should be pointed out that the performance is not near the local optimal VP but outperforms the linear local optimal approach and the suboptimal THP approach presented before. As it happens with the suboptimal proposal based on THP precoding, this simple solution based on hop by hop minimization will be specially useful in multihop systems where more than one relay are introduced between the BS and the end users.

6.2.2 Problem Decomposition based Suboptimal Approach

The main purpose of the proposal analysed in this section is again the reduction of the computational cost presented by the joint local optimal VP approach derived in Section 5.2, while maintaining a nice complexity-performance trade-off.

Briefly, the idea of this algorithm was inspired by *primal decomposition*, which was first introduced by Boyd in [Boyd04] and adapted to single-user MIMO relaying systems in [Tseng11] and [Mo09]. Since the MSE objective function employed as metric for precoders design is a complicated function of the precoding and relaying matrices, the optimization is difficult. By means of primal decomposition, the original optimization problem is de-

composed into a *master problem* and a *sub-problem*. Thus, primal decomposition performs optimization over some variables and then over the remaining variables, i.e.,

$$\min_{x,y} f(x,y) = \min_y \min_x f(x,y).$$

Using primal decomposition, a complex non-convex problem is converted to two solvable and simpler optimization problems. In our case, the original idea of primal decomposition is modified for computation cost reduction.

The system, depicted in Figure 5.2, applies VP non-linear precoding at the BS to limit the power at the BS and RS terminals while minimizing the overall mean square error, whereas linear processing is applied at the relay.

As shown in Section F, after some manipulations, the MSE in (5.7) can be expressed as

$$\xi = \mathbf{d}^H \left(\underbrace{\mathbf{I}_K - 2\Re(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}}) + \overline{\mathbf{F}}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{A} \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}} + \frac{N_{01}}{E_{BS}} \text{Tr}(\overline{\mathbf{W}}^H \mathbf{A} \overline{\mathbf{W}})}_{\mathbf{X}} \overline{\mathbf{F}}^H \overline{\mathbf{F}} \right) \mathbf{d}, \quad (6.5)$$

being $\overline{\mathbf{F}} = \beta_1 \mathbf{F}$ and $\overline{\mathbf{W}} = \beta_2 \mathbf{W}$ the unscaled versions of the precoding and relaying matrices and $\mathbf{A} = \mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02} K}{E_R} \mathbf{I}_R$. It can be seen in (6.5) that $\overline{\mathbf{F}}$ and $\overline{\mathbf{W}}$ are assumed constant over a channel realization.

As design strategy, applying the primal decomposition idea and in order to avoid the search of the perturbation vector in each iteration, we divided the optimization problem into a subproblem and a master problem. While the former searches the optimal precoding and relaying matrices for a fixed perturbation vector taking into account the power constraints, the latter finds the vector \mathbf{a} that minimizes the MSE for the previously selected local optimal $\overline{\mathbf{F}}$ and $\overline{\mathbf{W}}$. The optimization problem can be then defined as follows:

$$\{\mathbf{a}, \overline{\mathbf{F}}, \overline{\mathbf{W}}\} = \min_{\mathbf{a}' \in \tau \mathbb{Z}^K + j\tau \mathbb{Z}^K} \left(\overbrace{\mathbf{d}^H \left(\underbrace{\min_{\{\overline{\mathbf{F}}, \overline{\mathbf{W}}\}} \text{Tr}(\mathbf{X})}_{\text{subproblem}} \right)}^{\text{master problem}} \right) \mathbf{d},$$

where the power constraints have been incorporated to the objective function as described in Appendix F and \mathbf{X} is defined in (6.5) as the inner term of the mean square error function. Firstly, in order to obtain the precoding matrices, the sub-problem optimization problem is established as

$$\{\overline{\mathbf{F}}, \overline{\mathbf{W}}\} = \min_{\{\overline{\mathbf{F}}, \overline{\mathbf{W}}\}} \text{Tr}(\mathbf{X}).$$

The non-convex optimization problem, can be efficiently solved by the numerical method presented in [Fang06], where for a fixed $\overline{\mathbf{W}}$ the precoding matrix $\overline{\mathbf{F}}$ is derived and vice versa. By means of this solution, a local optimal is obtained, getting

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \mathbf{A} \mathbf{H}_1 + \frac{N_{01}}{E_{BS}} \text{Tr}(\mathbf{A}) \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H, \quad (6.6)$$

where $\mathbf{A} = \overline{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02}K}{E_R} \mathbf{I}_K \right) \overline{\mathbf{W}}$ and

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02}K}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H \overline{\mathbf{F}}^H \mathbf{H}_1^H \left(\mathbf{H}_1 \overline{\mathbf{F}} \overline{\mathbf{F}}^H \mathbf{H}_1^H + \frac{N_{01} \text{Tr}(\overline{\mathbf{F}} \overline{\mathbf{F}}^H)}{E_{BS}} \mathbf{I}_R \right)^{-1}, \quad (6.7)$$

being the relaying matrix composed by the regularized pseudo inverse of the channel \mathbf{H}_2 and $\mathbf{H}_1 \overline{\mathbf{F}}$. Once the matrices $\overline{\mathbf{F}}$ and $\overline{\mathbf{W}}$ are derived in the sub-problem, the master problem is solved under the assumption that the local optimal precoding and relaying matrices are employed.

In order to solve the master problem, after some manipulations the MSE function defined in (6.5) is reduced to

$$\xi = \mathbf{d}^H (\mathbf{I}_K - \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}}) \mathbf{d}. \quad (6.8)$$

As pointed out before, the perturbation vector is chosen as the nearest point in the lattice that minimizes the mean square error, being the optimization problem defined as

$$\mathbf{a} = \min_{\mathbf{a}' \in \tau \mathbb{Z}^K + j\tau \mathbb{Z}^K} \xi.$$

Carrying out the same procedure described in Section 5.2 for *VP-MMSE* local optimal VP approach, the inner term of Equation (6.8) is decomposed through Cholesky factorization as

$$(\mathbf{I}_K - \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \overline{\mathbf{F}}) = \mathbf{L}^H \mathbf{L},$$

where the search of \mathbf{a} can be established as

$$\mathbf{a} = \min_{\mathbf{a}' \in \tau \mathbb{Z}^K + j\tau \mathbb{Z}^K} \|\mathbf{L}(\mathbf{s} + \mathbf{a})\|_2^2. \quad (6.9)$$

As it will be seen in provided simulation results, this suboptimal design strategy outperforms both optimal linear and THP designs. Furthermore, the cost due to the inclusion of VP is greatly reduced due to the fact that the perturbation vector is not searched in each iteration. The proposed technique is described in Algorithm 11, where firstly an iterative

process is carried out for precoding and relaying computation, being l the current iteration and ϵ the convergence error. After that, the sphere encoder algorithm is computed for the search of the perturbation vector.

Algorithm 11 Calculation of \mathbf{F} , \mathbf{W} and \mathbf{a} under the MMSE criterion for problem decomposition based approach.

- 1: Initialize the variables: $\overline{\mathbf{F}}^0 = \mathbf{I}_{M \times K}$, $\overline{\mathbf{W}}^0 = \mathbf{I}_R$ and $\epsilon = \infty$.
 - 2: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 3: Calculate $\overline{\mathbf{F}}^{l+1}$ (6.6) for a fixed $\overline{\mathbf{W}}^l$.
 - 4: Computation of $\overline{\mathbf{W}}^{l+1}$ (6.7) for the updated $\overline{\mathbf{F}}^{l+1}$.
 - 5: Error calculation: $\epsilon = \|\overline{\mathbf{F}}^{l+1} - \overline{\mathbf{F}}^l\|^2 + \|\overline{\mathbf{W}}^{l+1} - \overline{\mathbf{W}}^l\|^2$.
 - 6: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 7: $\overline{\mathbf{F}}^l = \overline{\mathbf{F}}^{l+1}$.
 - 8: $\overline{\mathbf{W}}^l = \overline{\mathbf{W}}^{l+1}$.
 - 9: **end if**
 - 10: **end while**
 - 11: Inner term of (6.8) computation $\implies \mathbf{X}$.
 - 12: Cholesky factorization of the previously calculated matrix \mathbf{X} .
 - 13: Execute SE algorithm to obtain the perturbation vector \mathbf{a} (6.9).
-

6.2.3 Simulation Results

In order to conclude with the analysis started in Section 6.2 about the suboptimal VP solutions for the design of relaying and precoding matrices, simulation results are provided in this section. $4 \times 4 \times \{4 \times 1\}$ set-up (A) has been chosen for simulations. It should be underlined that the independent hop-by-hop VP solution is named *VP-MMSE-HH*, whilst *VP-MMSE-Sub* stands for the approach based on problem decomposition.

The following proposals have been tested in the results provided in this section:

- *Lin-MMSE* (Section 3.1)
- *THP-MMSE* (Section 5.1)
- *VP-MMSE* (Section 5.2)
- *THP-MMSE-Sub* (Section 6.1)
- *VP-MMSE-HH* (Section 6.2.1)
- *VP-MMSE-Sub* (Section 6.2.2)

In order to reduce the computational cost of the search of the perturbation vector in a lattice, we decided to limit the amount of points that form the lattice, which is defined as

$\nu \triangleq \{x + yj : |x|, |y| \leq (\sqrt{|\nu|} - 1)/2\}$, to 25 values ($|\nu| = T=25$) according to the results provided in [Barrenechea12].

To conclude, it should be pointed out that just for *VP-MMSE-Sub* proposal the convergence error $\epsilon = 0.001$ is imposed. Three different measurement are presented in the following sections: BER performance, achievable sum-rate as defined in (5.18) and the mean square error.

6.2.3.1 BER Performance

Figure 6.4 shows the BER performance of the cited approaches when ρ_2 is set to 15 dB (Figure 6.4(a)) and ρ_1 is fixed to 15 dB (Figure 6.4(b)).

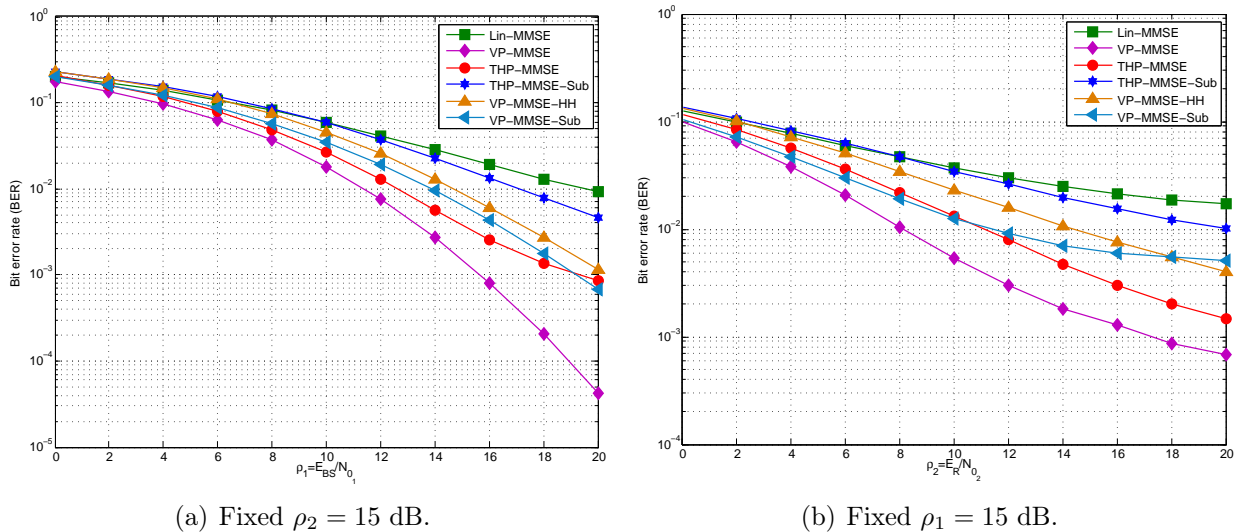


Figure 6.4: BER performance for MMSE non-linear VP suboptimal precoding techniques in $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation.

Figure 6.4(a) shows the performance of the suboptimal VP approaches compared to linear and THP proposals. The best suboptimal performance is achieved by *VP-MMSE-Sub*, which obtains a gain of 1 dB with respect to *VP-MMSE-HH*. The performance of *VP-MMSE-Sub* operates near *THP-MMSE*, considered the local optimal THP solution. This improvement is specially visible at high SNRs, from 17 dB on. Figure 6.4(a) depicts the effectiveness of VP suboptimal approaches over linear and *THP-MMSE-Sub* solution, getting a gain of 5 and 4 dB, respectively.

As expected Figure 6.4(b) depicts that the suboptimal VP approaches improve the suboptimal THP proposal in BER terms when ρ_1 is set to 15 dB. Nevertheless, the difference gap between *THP-MMSE* and *VP-MMSE-Sub* is higher, reaching 7 dB. We observe that at high SNR, *VP-MMSE-HH* outperforms *VP-MMSE-Sub*, being the former more effective at high ρ_2 . Comparing Figures 6.4(a) and 6.4(b) we can state that the performance is degraded

more by ρ_1 . Furthermore, we can firmly say that the performance obtained by the suboptimal VP solutions is clearly worse comparing with the optimal *THP-MMSE* and *VP-MMSE*. Nevertheless, as it will be presented in Section 6.3, the advantage of this solutions resides on the computational complexity instead on the performance.

6.2.3.2 Achievable Sum-Rate

The achievable sum-rate defined in (5.18) is measured by simulations in this section. Figure 6.5(a) shows the achievable sum-rate when ρ_2 is set to 15 dB. We observe that *VP-MMSE-Sub* outperforms the remaining systems with the exception of *VP-MMSE* and *THP-MMSE*. At low SNR, from 0 to 12 dB the difference gap that *THP-MMSE-Sub* and *Lin-MMSE* achieve over *VP-MMSE-HH* is considerable, in fact due to the simplest hop-by-hop optimization. Furthermore, it can be seen that the difference between optimal and suboptimal schemes is greater in SR than in BER terms.

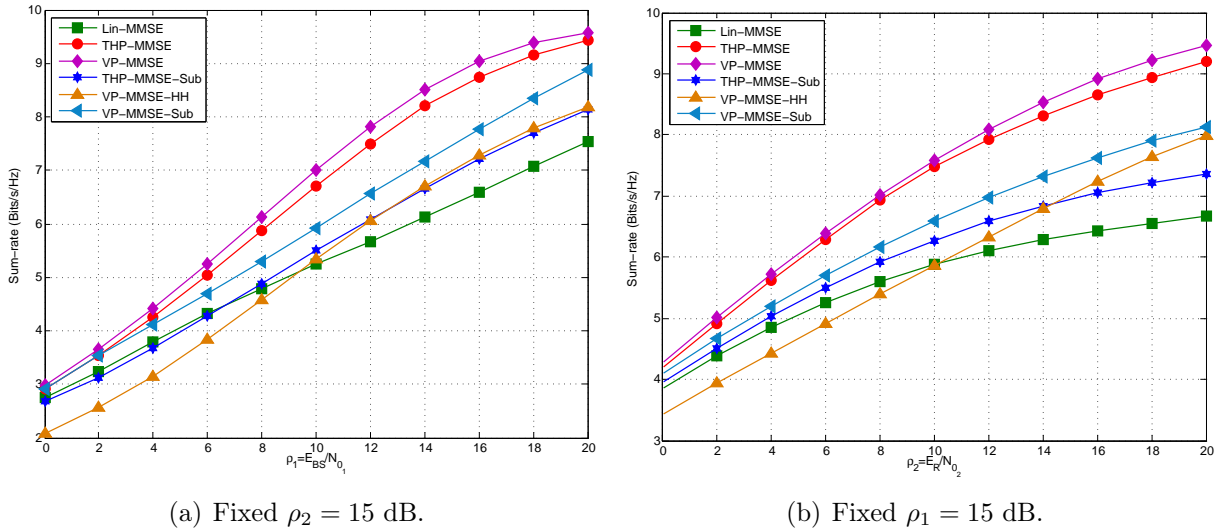


Figure 6.5: Sum-rate achieved by MMSE non-linear VP suboptimal precoding techniques in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

When ρ_1 is fixed to 15 dB as shown in Figure 6.5(b) and the non-iterative approach *VP-MMSE-HH*, we observe that *VP-MMSE-Sub* outperforms the capacity of *THP-MMSE-Sub*, *Lin-MMSE* and *VP-MMSE-HH*. The difference gap that *VP-MMSE* gets over *VP-MMSE-Sub* is greater. Until 10 dB *Lin-MMSE* obtains a better capacity compared with *VP-MMSE-HH*, being the difference considerable. Once 10 dB are reached, *VP-MMSE-HH* improves significantly linear's performance, achieving a maximum of 7 B/s/Hz, which coincides with the maximum capacity achieved by *VP-MMSE-Sub*. If more than 20 dB are simulated, *VP-MMSE-HH* might outperform *VP-MMSE-Sub*. From 14 dB on, *VP-MMSE-HH* performs better than *THP-MMSE-Sub*, getting a greater capacity at high SNR.

6.2.3.3 Mean Square Error

Mean square error has been tested for the iterative algorithm *VP-MMSE-sub*, stopping the algorithms after a certain number of iterations, as shown in Figure 6.6. The SNR has been configured in pairs. Thus $\{ \rho_1 = 10 \text{ dB} < \rho_2 = 15 \text{ dB} \}$, as well as $\{ \rho_1 = 20 \text{ dB} > \rho_2 = 15 \text{ dB} \}$ pairs have been evaluated. Figure 6.6(a) depicts the former pair where the suboptimal VP reduces significantly *Lin-MMSE*'s and *THP-MMSE-Sub*'s MSE error. Due to the non-iterative nature of *VP-MMSE-HH*, its error is plotted constant, being lower than the unit. After a few number of iterations, *VP-MMSE-Sub* approach converges. Furthermore, once the convergence is reached by *VP-MMSE-Sub* after 3 iterations, this reduces the error more than *VP-MMSE-HH*.

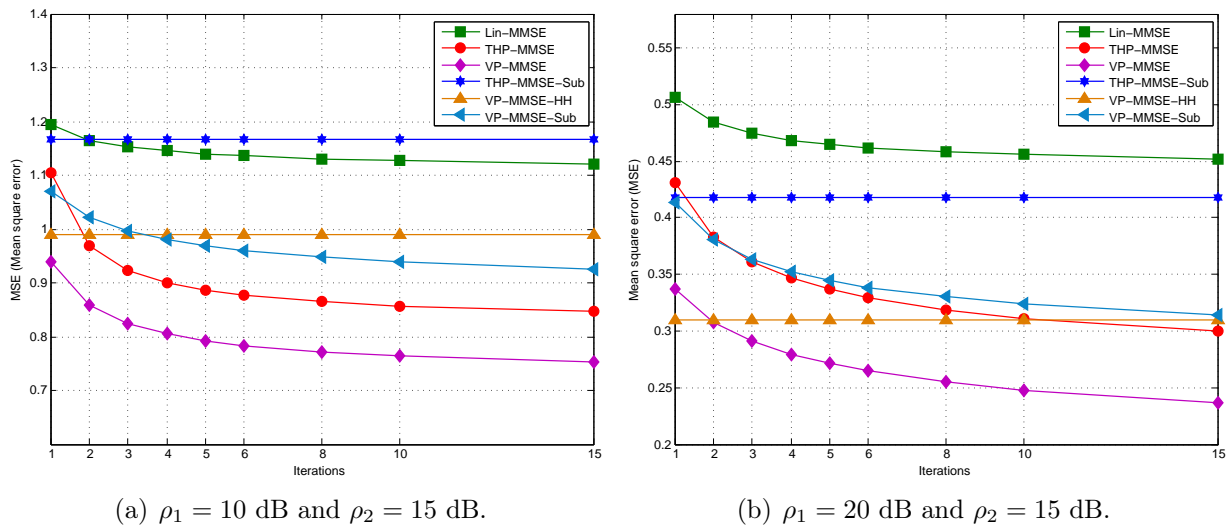


Figure 6.6: Mean square error measurement for *VP-MMSE-HH* and *VP-MMSE-Sub* approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

In addition, Figure 6.6(b) shows the convergence and the error with $\rho_1 = 20 \text{ dB}$ and $\rho_2 = 15 \text{ dB}$, where *VP-MMSE-Sub* converges faster. *VP-MMSE-HH* proposal outperforms *VP-MMSE* before its convergence. As expected, the suboptimal approaches based on VP perform better than *Lin-MMSE* and *THP-MMSE-Sub* in error terms. Furthermore, we observe that the difference between *VP-MMSE-Sub* and *VP-MMSE* is lower when ρ_1 grows

6.3 Computational Complexity

After showing the effectiveness of the suboptimal approaches for the design of precoding and relaying matrices when non-linear transmission is applied, the computational complexity of the aforementioned systems is shown in this section. Provided complexity results will show the reduction of the computational cost, which is the main design criteria. Furthermore, the computational cost is given in terms of FLOPs, complexity order and run-time.

Firstly FLOPs and complexity order are analysed based on Golub's complexity criteria [Golub96]. For example, the number of floating point operations of the product of two matrices $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\mathbf{B} \in \mathbb{C}^{n \times o}$ is $(2n + 1)mo$, while the Moore-Penrose inverse of a square matrix $\mathbf{A}^{n \times n}$ needs an amount of $(4n^3 + 6n^2 + 3n)$ FLOPs.

In addition, as accomplished previously for the earlier presented approaches, a fixed number of iterations (*it*) for *VP-MMSE-Sub* approach and $n = M = R = K$ are assumed for simplicity.

6.3.1 FLOP Operations and Complexity Order Analysis

Table 6.1 shows the number of FLOPs needed for the computation of precoding (\mathbf{F}) and relaying (\mathbf{W}) matrices as well as the computation of the feedback filter (\mathbf{B}) in THP approach. Obviously, comparing the number of floating point operations needed for the suboptimal approaches with the number of FLOPs employed for non-linear optimal approaches (see Table 5.1), the first are clearly lower meeting with their purpose. It should be pointed out that the number of operations required for SE algorithm and modulo operation have been described in Table 5.2.

| Algorithm | Operation | FLOPs |
|--------------|--------------|---|
| THP-MMSE-Sub | \mathbf{F} | $4M^3 + (3 + 2K + 8M)R^2 + (1 + 2M)K^2 + (2R + 2)MK + (3 + 5R)M$ |
| | \mathbf{B} | $4R^3K + (5K + 2K^2)R^2 + (20 + 8M)K^3 + (2M^2 + 6RM + 4M - 5)K^2 + 4M^3K + (2RK + 12K)M^2 + 8K + 3MK + 3RMK + 1$ |
| | \mathbf{W} | $4K^3 + (6 + 2R)K^2 + 2(K + 1)R^2 + (3 + 2R)K$ |
| VP-MMSE-HH | \mathbf{F} | $4M^3 + (4R + 9)M^2 + (3 + R)M$ |
| | \mathbf{W} | $10R^3 + (11 + 5K)R^2 + (3 + K)R$ |
| VP-MMSE-Sub | \mathbf{F} | $4R^3 + 4M^3(5 + 6M)R^2 + (2R + 8)M^2 + (2K + 4)RM + MK$ |
| | \mathbf{W} | $12R^3 + (4K + 2M + 19)R^2 + 2(K + 1)M^2 + (6 + 2K + 2M)R + 6MRK + 2$ |

Table 6.1: Number of FLOPs required by *THP-MMSE-Sub*, *VP-MMSE-HH* and *VP-MMSE-Sub* approaches.

Furthermore, Table 6.2 provides the complexity order of the aforementioned non-linear suboptimal approaches and the number of FLOPs needed to run the algorithm n and $n = 4$.

It should be underlined that the number of iterations (*it*) has been removed from the FLOPs for *THP-MMSE-Sub* and *VP-MMSE-HH* approaches, as a consequence of the lack of iterations. Nevertheless, the number of iterations is taken into account for \mathbf{F} and \mathbf{W}

| Algorithm | Comp. order | FLOPS | |
|---------------------|-----------------|--|--------------------|
| | | M=R=N=n | M=R=N=4 |
| THP-MMSE-Sub | $O(n^4)$ | $(28n^4 + 80n^3 + 41n^2 + 15n + 5 + N_B (6n^2 - 4n + 1))$ | 21109 |
| VP-MMSE-HH | $O(T^{n+1})$ | $(47n^3 + 42n^2 + 7n + 8 + N_B (\frac{T^{n+1}-T}{T-1} + n))$ | 40694116 |
| VP-MMSE-BD | $O(it T^{n+1})$ | $it (82n^3 + 70n^2 + 10n + 7) + N_B (\frac{T^{n+1}-T}{T-1} + n)$ | 40696815 <i>it</i> |

Table 6.2: Complexity order for *THP-MMSE-Sub*, *VP-MMSE-HH* and *VP-MMSE-Sub* approaches when $N_B = 100$ and $T = |\nu| = 25$.

derivation in *VP-MMSE-Sub*.

In addition, *THP-MMSE-Sub*'s computational order is quartic whereas the complexity of VP suboptimal approaches will be determined by the number of points that form the lattice (T), n and it for *VP-MMSE-Sub*.

Unsurprisingly, the complexity order of these suboptimal approaches is lower in comparison with the joint local optimal non-linear approaches presented in Chapter 5. Apart from reducing the FLOP operations, the suboptimal non-linear approaches are simpler due to the lack of iterations in *THP-MMSE-Sub* and *VP-MMSE-HH* approaches. For example, the complexity of *VP-MMSE-HH* will be limited by the addition of the cost of two inverse operation.

6.3.2 Run-Time Analysis

Figure 6.7 depicts the run-time required for the computation of the suboptimal algorithms, considering three terminals: BS, RS and single-antenna users. While Figure 6.7(a) compares *THP-MMSE-Sub* with linear and joint local optimal THP approaches, Figure 6.7(b) shows the run-time of suboptimal VP approaches. Obviously, the results will be lower compared to the ones given in Section 5.4.2 for the optimal frameworks.

As viewed in Figure 6.7(a) when $\rho_2 = 15$ dB, *THP-MMSE-Sub* reduces considerably the run-time of *THP-MMSE*. The maximum run-time, measured when $\rho_1 = 20$ dB, reaches 0.2 seconds for *THP-MMSE* and 0.02 seconds for *THP-MMSE-Sub*, respectively. Thus, the time needed for *THP-MMSE* algorithm computation is ten times the one required for *THP-MMSE-Sub*. Furthermore, the run-time for *THP-MMSE-Sub* system is quite constant from $\rho_1 = 0$ to $\rho_1 = 20$ dB, being lower than 0.03 seconds.

As depicted in Figure 6.7(b), at low SNR (from $\rho_1 = 0$ to $\rho_1 = 10$ dB) *VP-MMSE-HH* presents a greater run-time compared with *VP-MMSE-Sub*. The contrary happens at high

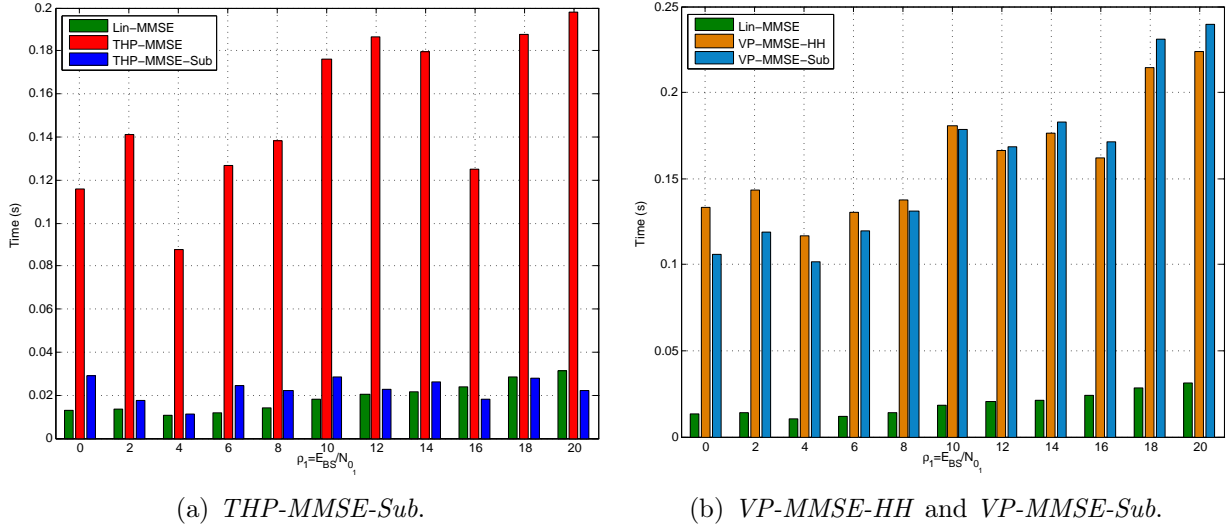


Figure 6.7: Run-time needed for *THP-MMSE-Sub*, *VP-MMSE-HH* and *VP-MMSE-Sub* algorithm for $4 \times 4 \times \{4 \times 1\}$ set-up when $\rho_2 = 15$ dB.

SNR, from 10 dB on, being the greater run-time achieved by *VP-MMSE-Sub*. Furthermore, the maximum run-time required at $\rho_1 = 20$ dB equals 0.23 seconds and 0.24 seconds for *VP-MMSE-HH* and *VP-MMSE-Sub* approaches, respectively. To conclude, it should be underlined that the run-time needed by the suboptimal algorithms is negligible in comparison to *VP-MMSE*, being the maximum run-time obtained by the latter of about 40 seconds.

6.4 Chapter Summary

This chapter has presented three novel suboptimal precoder and relaying design strategies when non-linear precoding techniques have been applied to multiuser MIMO AF relaying systems. By means of this suboptimal precoding approaches, the excessive computational cost presented by the local optimal non-linear solutions introduced in Chapter 5 has been minimized considerably.

At the beginning, a novel suboptimal approach is proposed based on THP precoding. Firstly, the second hop mean square error has been minimized to get the relaying matrix, whilst THP has been applied at the BS to cancel the interference from the transmitter to the receivers. Throughout this solution the interconnection between matrices has been broken, thus, the computational cost has been decreased greatly.

The remaining suboptimal proposals have been designed for VP transmission. The first approach has targeted the minimization of each hop independently. With the help of this simple suboptimal strategy, the precoding and relaying matrices have been found without iterations. Afterwards, VP has been employed at the BS for total error minimization and power limiting.

Finally, a second suboptimal approach has been proposed for VP transmission through the division of the non-convex optimization problem into a master problem and a sub-problem. By means of the solution of the sub-problem, the local optimal precoding and relaying matrices have been found firstly, whilst through the optimization of the master problem the perturbation vector, applied at the BS for total error minimization, has been computed. An iterative algorithm is required for this solution owing to the dependence of the precoding and relaying matrices. Nevertheless, the search of the perturbation vector in each iteration has been avoided.

Simulation results have shown the performance of the proposed novel suboptimal approaches, which has been degraded in comparison to the local optimal non-linear approaches. However, the computational complexity analysis has shown the real effectiveness of this approaches.

The contributions of this chapter turn around the proposal of suboptimal algorithms for complexity reduction. Due to the complexity presented by non-linear precoding approaches, three novel suboptimal algorithms have been proposed one for THP and the rest for VP. These are based on the independent minimization of the mean square error of each hop and the division of the optimization problem into a master problem and a sub-problem. By means of these, the search of the perturbation vector in each iteration, which adds a significant complexity, is avoided, reducing in this way the computational cost. Furthermore, the adaptation of vector precoding to multiuser MIMO relaying systems has been carried out, demonstrating the effectiveness of this precoding technique with extensive simulation results. Finally, a complete complexity analysis has been given at the end of the chapter.

Block Diagonal Geometric Mean Decomposition for Multiuser MIMO Relaying Systems

It has been shown in Section 4.3 that the wireless MIMO channel can be made diagonal by means of singular value decomposition, creating multiple non-interfering parallel subchannels. Furthermore, combining this with water-filling algorithm channel capacity can be achieved [Spencer04b]. Nevertheless, as introduced in Chapter 4, BER performance can be limited by the lowest singular values, located at the end of the diagonal matrix at the output of SVD. Basically, this entails the assignation of a larger power to the subchannels related to these singular values. As a consequence of this, a significant capacity loss happens due to the so called inverse water-filling [Mo09].

Moreover, as proved in previous chapters, when SVD is applied to multiuser MIMO relaying, not all the users will be served or not all of them will receive all the transmitted streams. However, due to its simplicity, singular value decomposition will help on transceiver design. Nevertheless, as pointed out previously, BER performance will be limited by the weakest subchannel.

In order to overcome SVD's limitation, GMD is introduced in [Jiang05b] for point-to-point MIMO framework. Mainly, GMD decomposes the channel in multiple parallel subchannels with the same SNR. This property can bring about much convenience in modulation/demodulation processes. In fact, GMD does not make tradeoffs between throughput and BER performance. Instead, it attempts to get the best of both worlds simultaneously. In [Jiang05b] transceivers are designed for sum-rate maximization under ZF criterion using GMD decomposition combined with DFE and THP for interference cancellation at the receiver and transmitter, respectively.

When multiple multi-antenna users share the wireless medium, GMD decomposition is converted to block diagonal geometric mean decomposition (BD-GMD) [Lin08], where an identical SNR is obtained in blocks, each of which corresponds to a user. Hence, all users' streams will be transmitted with the same effective signal-to-noise ratio. In [Lin08], BD-

GMD is combined with DPC for performance improvement and is also extended to uniform channel decomposition. As a consequence of assigning the same SNR for each user, equal-rate coding can be applied across the subchannels.

In [Liu08], the precoding matrix is designed for downlink multiuser MIMO systems with multiple antenna mobile terminals using BD-GMD decomposition. Furthermore, THP is applied at the BS for interference cancellation and a linear receiver is set at each receiver.

In spite of the great performance obtained for single user and multiuser MIMO systems, GMD has not been taken into consideration in multihop systems, mainly due to the design of the relaying matrix. In [Tseng09], GMD is used to derive the precoder at the transmitter and the relaying matrix considering the direct link. The single user is equipped with a DFE filter for interference cancellation. Throughout GMD, the non-convex optimization problem is converted to a convex scalar problem where a close-form solution can be obtained.

Considering the great performance obtained with the combination of GMD and THP, VP has been also considered for the transceiver design under GMD decomposition in point-to-point MIMO scenarios [Liu07] [Liu10]. In [Liu07], SVD, QR and GMD decomposition are used for transceiver design in single-user MIMO systems, where a comparative is provided among these decomposition techniques. Whilst SVD diagonalized the channel, QR and GMD decompositions obtain a triangular matrix. Since SE algorithm considers interference cancellation to find the optimal perturbation vector, it does not make sense to find the perturbation vector to cancel the interferences when the system was previously diagonalized. Due to GMD property, where the same SNR is obtained, the imbalance among subchannel gains is cancelled. This directly affects to the search of the perturbation vector. The diversity order of GMD is also derived in [Liu07], obtaining it to be $M - m + 1$, where $m = \min(M_R, M_T)$ and $M = \max(M_R, M_T)$. On the other hand, [Liu10] designs the precoder at the BS combining VP and BD-GMD, where a lattice reduction based sphere encoder are used instead of the typical SE.

In this chapter the precoding and relaying matrices are designed using BD-GMD decomposition, which will help improving the convergence and performance of the system. Moreover, BD-GMD is combined with VP in order to minimize the overall MSE. Two different approaches are here proposed: the joint optimal design of the matrices is first carried out, while a suboptimal approach is given as a second option in order to reduce the computational complexity. Provided simulation results will show the effectiveness of both approaches.

7.1 BD-GMD Algorithm Analysis

The framework employed is the one introduced in Section 3.1, where an M -antenna BS sends data through a relay equipped with R antennas to K N_k -antenna users, as depicted in Figure 3.1. As it has been stated, BD-GMD decomposes the wireless channel matrix

$\mathbf{H} \in \mathbb{C}^{N \times M}$, created from the BS to the end users, into multiple subchannels with the same SNR, according to the following formula:

$$\mathbf{H} = \mathbf{Q}\mathbf{R}\mathbf{P}^H, \quad (7.1)$$

where $\mathbf{Q} \in \mathbb{C}^{N \times z}$ and $\mathbf{P} \in \mathbb{C}^{M \times z}$ are unitary matrices while $\mathbf{R} \in \mathbb{R}^{z \times z}$ stands for an upper triangular matrix and z is the rank of matrix \mathbf{H} . The unitary matrices \mathbf{Q} and \mathbf{P} correspond to information loss-less filters applied to transmitted and received signals to minimize the maximum error rate of the network. Since the SNRs of the subchannels are identical, the diagonal elements of \mathbf{R} are equal and can given by

$$r_{i,i} = \bar{\sigma}_{\mathbf{H}} = \left(\prod_{j=1}^z \lambda_{\mathbf{H},j} \right)^{1/z} \quad 1 \leq i \leq z, \quad (7.2)$$

being $r_{i,i}$ the i^{th} value of the diagonal of matrix \mathbf{R} , whereas $\lambda_{\mathbf{H},j}$ stands for the j^{th} singular value of the channel matrix \mathbf{H} .

The implementation of GMD can be easily done though SVD decomposition, where firstly, the singular value decomposition of \mathbf{H} is carried out as The implementation of GMD can be easily done though SVD decomposition, where firstly, the singular value decomposition of \mathbf{H} is carried out as $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where matrices \mathbf{U} and \mathbf{V} denote the right and left singular vectors. In order to compute $r_{i,i}$ as in (7.2), an ensemble of $\mathbf{R}^{(o)}$ matrices are generated for $1 \leq o \leq z$, each of which has to fulfil with the following conditions:

- A. $r_{i,j}^{(o)} = 0$ when $i > j$ or $j > \max(o, i)$.
- B. $r_{i,i}^{(o)} = \bar{\sigma}_{\mathbf{H}}$ for all $i < o$, computed as (7.2).

The permutation process for the construction of the matrices is defined in Algorithm 12. Once the permutation is accomplished, the next step will be the construction of matrices \mathbf{G}_1 and \mathbf{G}_2 for the derivation of \mathbf{Q}_n and \mathbf{P}_n , by modifying the elements in the identity matrix that lie at the intersection of rows n and $n + 1$ with columns n and $n + 1$ as depicted in Figure 7.1.

The permuted $\mathbf{\Pi}\mathbf{R}^{(n)}\mathbf{\Pi}$ matrix is then left-multiplied by \mathbf{G}_2^H and right-multiplied by \mathbf{G}_1 . As shown in Figure, 7.1 by means this multiplication, the elements of a 2×2 matrix are modified at the intersection of rows n and $n + 1$ with columns n and $n + 1$ as shown in the following equation:

$$\underbrace{\begin{bmatrix} c\delta_1 & s\delta_2 \\ -s\delta_2 & c\delta_1 \end{bmatrix}}_{\mathbf{G}_2^H} \underbrace{\begin{bmatrix} \delta_1 & 0 \\ 0 & c\delta_2 \end{bmatrix}}_{\mathbf{\Pi}\mathbf{R}^{(l)}\mathbf{\Pi}} \underbrace{\begin{bmatrix} c & -s \\ s & c \end{bmatrix}}_{\mathbf{G}_1} = \underbrace{\begin{bmatrix} \bar{\sigma}_{\mathbf{H}} & a \\ 0 & b \end{bmatrix}}_{\mathbf{R}^{(l+1)}}. \quad (7.3)$$

Algorithm 12 Permutation process.

- 1: Generate \mathbf{Q}_n and $\mathbf{P}_n \implies$ Generated using a pair of Givens rotations and symmetric permutation.
 - 2: Definition of $\mathbf{R}^{(n+1)}$ as $\mathbf{R}^{(n+1)} = \mathbf{Q}_n^H \mathbf{R}^{(n)} \mathbf{P}_n$.
 - 3: **if** $\mathbf{R}^{(n)}$ fulfils with A and B **then**
 - 4: **if** $r_{n,n}^{(n)} \geq \bar{\sigma}_{\mathbf{H}}$ **then**
 - 5: Construct permutation matrix $\mathbf{\Pi}$ with the property $\implies \mathbf{\Pi} \mathbf{R}^{(n)} \mathbf{\Pi}$.
 - 6: Exchange $(n+1)$ st diagonal elements of $\mathbf{R}^{(n)}$ with any element $r_{p,p} \implies p > n$ for which $r_{p,p} \leq \bar{\sigma}_{\mathbf{H}}$.
 - 7: **else**
 - 8: Construct permutation matrix $\mathbf{\Pi}$.
 - 9: Exchange $(n+1)$ st diagonal elements of $\mathbf{R}^{(n)}$ with any element $r_{p,p} \implies p > n$ for which $r_{p,p} \geq \bar{\sigma}_{\mathbf{H}}$.
 - 10: **end if**
 - 11: Compute new diagonal elements at n and $n+1$ positions $\implies \delta_1 = r_{n,n}^{(n)}$ and $\delta_2 = r_{p,p}^{(n)}$.
 - 12: **end if**
-

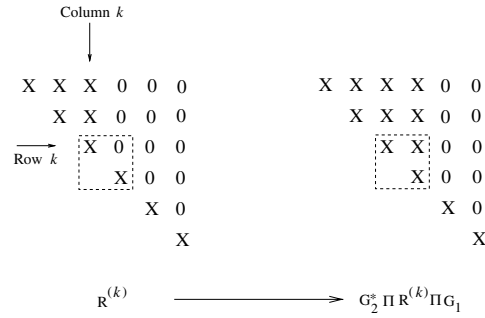


Figure 7.1: Columns and rows modification [Jiang05b]

Two options are examined. Firstly, if $\delta_1 = \delta_2 = \bar{\sigma}_{\mathbf{H}}$, the values that c and s take are: $c = 1$ and $s = 0$. Otherwise, if $\delta_1 \neq \delta_2$, $c = \sqrt{\frac{\bar{\sigma}_{\mathbf{H}}^2 - \delta_2^2}{\delta_1^2 - \delta_2^2}}$ and $s = \sqrt{1 - c^2}$. For both options, x and y are computed as:

$$x = \frac{sc(\delta_2^2 - \delta_1^2)}{\bar{\sigma}_{\mathbf{H}}} \quad \text{and} \quad y = \frac{\delta_1 \delta_2}{\bar{\sigma}_{\mathbf{H}}}.$$

As can be seen in (7.3), c and s are real scalars chosen to fulfil with $c^2 + s^2 = 1$ and $(c\delta_1^2)^2 + (s\delta_2^2)^2 = \bar{\sigma}_{\mathbf{H}}^2$. In order to satisfy (7.3), taking into account the orthogonality of left and right singular vectors, the unitary matrices \mathbf{Q} and \mathbf{P} are defined as

$$\mathbf{Q} = \mathbf{U} \left(\prod_{i=1}^{z-1} \mathbf{Q}_i \right) \quad \text{and} \quad \mathbf{P} = \mathbf{V} \left(\prod_{i=1}^{z-1} \mathbf{P}_i \right),$$

being $\mathbf{Q}_i = \mathbf{\Pi} \mathbf{G}_2$ and $\mathbf{P}_i = \mathbf{\Pi} \mathbf{G}_1$. Finally, the upper triangular matrix \mathbf{R} is equivalent to:

$$\mathbf{R} = (\mathbf{Q}_{z-1}^H \dots \mathbf{Q}_2^H \mathbf{Q}_1^H) \mathbf{\Lambda} (\mathbf{P}_1 \mathbf{P}_2 \dots \mathbf{P}_{z-1}).$$

This process is briefly summarized in Algorithm 13

Algorithm 13 Geometric mean decomposition based on singular value decomposition approach.

- 1: SVD of the channel \mathbf{H} .
 - 2: Initialize the variables: $\mathbf{Q} = \mathbf{U}$, $\mathbf{P} = \mathbf{V}$ and $\mathbf{R} = \mathbf{\Lambda}$.
 - 3: **for** $l = 1 : z - 1$ **do**
 - 4: Execute Algorithm 12.
 - 5: Construct \mathbf{G}_1 and \mathbf{G}_2 matrices (7.3).
 - 6: Replace $\mathbf{R}^{(l+1)} \iff \mathbf{G}_2^H \mathbf{R}^{(l)} \mathbf{G}_1$, $\mathbf{Q}_l \iff \mathbf{Q} \mathbf{G}_2$ and finally $\mathbf{P}_l \iff \mathbf{P} \mathbf{G}_1$.
 - 7: **end for**
-

To conclude, it should be pointed out that GMD decomposition is converted to BD-GMD for multiuser schemes, where multiple parallel subchannels are created, each one related to a user. By means of BD-GMD, the channel is diagonalized by blocks, being each block subject to the same SNR. In order to apply BD-GMD to a multiuser MIMO relaying scenario, BD is first applied, while Algorithm 13 is applied to each equivalent channel \mathbf{H}_k .

7.2 VP Transmission with Optimal BD-GMD

7.2.1 System Model

A multiuser MIMO amplify-and-forward system is here considered, where as depicted in Figure 5.2 a BS equipped with M antennas transmits data through an R -antenna relay to K N_k -antenna users. Considering (5.6), each user receives:

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k},$$

being β_1 and β_2 the scaling factors that limit the power at the BS and RS, respectively and \mathbf{D}_k user k equalization matrix.

In order to cancel the interference between users and the BS, BD [Spencer04b] is proposed. As pointed out in Chapter 3, due to the lack of convergence of block diagonalization algorithm from the transmitted to the receivers, BD was only applied to the second hop, making the second channel block-diagonal. Furthermore, in Section 5.2.3, BD is combined with VP for performance improvement. As a consequence of this, BD and VP will be used in combination with GMD in this section.

The relaying matrix \mathbf{W} is defined as $\mathbf{W} = \mathbf{W}_2 \mathbf{W}_1$, being $\mathbf{W}_2 = [\mathbf{W}_{2,1}, \dots, \mathbf{W}_{2,k}, \dots, \mathbf{W}_{2,K}] \in \mathbb{C}^{R \times N}$ used for block diagonalization and $\mathbf{W}_1 = [\mathbf{W}_{1,1}^T, \dots, \mathbf{W}_{1,k}^T, \dots, \mathbf{W}_{1,K}^T]^T \in \mathbb{C}^{N \times R}$ for processing at the relay. Taking the structure of \mathbf{W} into consideration, the signal

received by user k is

$$\begin{aligned} \hat{\mathbf{d}}_k &= \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS} + \underbrace{\sum_{\substack{j=1 \\ j \neq k}}^K \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,j} \mathbf{W}_{1,j} (\mathbf{H}_1 \mathbf{y}_{BS} + \mathbf{n}_1)}_{\text{Interference}} \\ &+ \underbrace{\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}}_{\text{Noise}}. \end{aligned}$$

In [Spencer04b] the constraint imposed for block diagonalization is $\mathbf{H}_{2,k} \mathbf{W}_{2,j} = \mathbf{0}_{N_k}$, where the symbols transmitted to the user k are mapped in the null space created by the interfering subchannels. Carrying out the procedure described in Section 5.2.3, after BD constraint application we obtain:

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k},$$

being $\mathbf{H}_{e,k} = \mathbf{H}_{2,k} \mathbf{W}_{2,k} \in \mathbb{C}^{N_k \times N_k}$ the equivalent channel created from the RS to user k after block diagonalization.

7.2.2 Joint Filter Design and BD-GMD Application

As stated in Section 7.2.1, the main objective of this approach is the joint design of the precoding (\mathbf{F}) and relaying (\mathbf{W}_1) matrices, as well as the selection of the perturbation vector \mathbf{a} under MMSE criterion. For the joint design, the MSE is defined as:

$$\begin{aligned} \xi &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{d}_k\|_2^2 \right], \\ &= \sum_{k=1}^K \beta_2^2 \beta_1^2 \mathbf{y}_{BS}^H \mathbf{H}_1^H \mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{D}_k^H \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2^2 \beta_1^2 N_{01} \text{Tr} \left(\mathbf{W}_{1,k}^H \mathbf{H}_{e,k}^H \mathbf{D}_k^H \mathbf{D}_k \right. \\ &\quad \left. \mathbf{H}_{e,k} \mathbf{W}_{1,k} \right) + \beta_2^2 N_{02} \text{Tr} \left(\mathbf{D}_k^H \mathbf{D}_k \right) - 2\beta_2 \beta_1 \Re \left(\mathbf{d}_k^H \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS_k} \right) + \mathbf{d}_k^H \mathbf{d}_k, \end{aligned}$$

being \mathbf{d}_k the perturbed symbols transmitted for user k and $\mathbf{y}_{BS_k} = \mathbf{F}_k \mathbf{d}_k$ the precoded symbols, respectively.

In order to simplify the analysis, the channel is decomposed. The proposal presented here decomposes the equivalent channel created from the BS to the users by means of BD-GMD, being the equivalent channel defined as

$$\bar{\mathbf{H}}_k = \mathbf{H}_{e,k} \bar{\mathbf{W}}_{1,k} \mathbf{H}_1 \in \mathbb{C}^{N_k \times M}. \quad (7.4)$$

Assuming $N_k = \text{rank}(\bar{\mathbf{H}}_k)$ as the rank of the equivalent channel matrix $\bar{\mathbf{H}}_k$, the channel

can be easily decomposed as in Section 7.1 by means of GMD decomposition as

$$\bar{\mathbf{H}}_k = \mathbf{Q}_k \mathbf{R}_k \mathbf{P}_k^H,$$

being $\mathbf{Q}_k \in \mathbb{C}^{N_k \times N_k}$ and $\mathbf{P}_k \in \mathbb{C}^{M \times N_k}$ unitary matrices and $\mathbf{R}_k \in \mathbb{R}^{N_k \times N_k}$ an upper triangular matrix with the same elements at the diagonal. The equal elements of the diagonal, which provide the same SNR per user, can be obtained as described in (7.2), being $r_{i,i}$ the i^{th} element of diagonal of the matrix \mathbf{R}_k , related to the user k . The elements of the diagonal are equal to $\bar{\sigma}_{\bar{\mathbf{H}}_k}$ and $\lambda_{\bar{\mathbf{H}}_k, j}$ is the singular value located at the j^{th} position. For simplicity $\mathbf{D}_k = \bar{\mathbf{Q}}_k^H \in \mathbb{C}^{N_k \times L_k}$ is assumed, taking the first L_k columns of \mathbf{Q}_k . Furthermore, it should be pointed out that due to the orthogonality of $\bar{\mathbf{Q}}_k$, we can state that $\mathbf{D}_k^H \mathbf{D}_k = \mathbf{I}_{N_k}$ and $\mathbf{D}_k \mathbf{D}_k^H = \mathbf{I}_{L_k}$.

Hence, the optimization problem is defined as

$$\begin{aligned} \{\mathbf{a}, \mathbf{F}, \mathbf{W}_1, \beta_1, \beta_2\} &= \underset{\{\mathbf{a}, \mathbf{F}, \mathbf{W}_1, \beta_1, \beta_2\}}{\operatorname{argmin}} \quad \xi \\ \text{s.t.} \quad &\|\mathbf{y}_{BS}\|_2^2 = E_{BS}, \\ &\|\mathbf{y}_R\|_2^2 = E_R. \end{aligned} \quad (7.5)$$

The optimization problem defined in (7.5) is subject to two power constraints, being E_{BS} and E_R the total transmission power at the BS and RS, respectively. Furthermore, the problem is highly non-convex. Once again, by means of the numerical method introduced in [Fang06] a local optimal solution can be effectively derived.

Carrying out the steps described in Section B.2, the expression obtained for the relaying matrix is

$$\mathbf{W}_{1,k} = \frac{1}{\beta_2} \left(\mathbf{H}_{e,k}^H \mathbf{H}_{e,k} + \frac{N_{02} N}{E_R} \mathbf{I}_{N_k} \right)^{-1} \mathbf{H}_{e,k}^H \mathbf{D}_k^H \mathbf{R}_{d_k} \bar{\mathbf{F}}_k^H \mathbf{H}_1^H (\mathbf{H}_1 \bar{\mathbf{y}}_{BS} \bar{\mathbf{y}}_{BS}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R)^{-1} \quad (7.6)$$

The input signal covariance matrix \mathbf{R}_{d_k} is related to the user k and is stacked at the matrix $\mathbf{R}_d = \text{blkdiag}(\mathbf{R}_{d_1}, \dots, \mathbf{R}_{d_k}, \dots, \mathbf{R}_{d_K})$. It should be pointed out that, opposite to THP, the statistics of the vector perturbation are still unknown for multiuser MIMO relaying systems, entailing the lack of useful approximations. Hence, it forces the search of the perturbation vector for $\mathbf{W}_{1,k}$ computation. On the other hand, matrix $\bar{\mathbf{F}}_k \in \mathbb{C}^{M \times L_k}$ stands for the unscaled precoding matrix associated to user k , while $\bar{\mathbf{y}}_{BS}$ denotes the unscaled precoded symbols.

From the definition of β_2 :

$$\beta_2 = \sqrt{\frac{\sum_{k=1}^K \text{Tr} \left(\mathbf{R}_{\mathbf{d}_k} \bar{\mathbf{F}}_k^H \mathbf{H}_1^H \mathbf{A}^{-1} \mathbf{H}_1 \bar{\mathbf{F}}_k \mathbf{R}_{\mathbf{d}_k} \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{B}^{-2} \mathbf{H}_{e,k} \mathbf{D}_k \right)}{E_R}},$$

being $\mathbf{A} = (\mathbf{H}_1 \bar{\mathbf{y}}_{BS} \bar{\mathbf{y}}_{BS}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R)$ and $\mathbf{B} = (\mathbf{H}_{e,k}^H \mathbf{H}_{e,k} + \frac{N_{02} N}{E_R} \mathbf{I}_{N_k})$.

In order to obtain the precoding matrix and the scaling factor β_1 , BD-GMD is applied to (7.4), simplifying in this way the complexity of the analysis. After some manipulations, the simplified MSE stands for

$$\begin{aligned} \xi &= \beta_1^2 \mathbf{y}_{BS}^H \mathbf{P} \bar{\mathbf{R}}^H \bar{\mathbf{R}} \mathbf{P}^H \mathbf{y}_{BS} + \beta_1^2 N_{01} \text{Tr} \left(\bar{\mathbf{W}}_1^H \mathbf{H}_e^H \mathbf{H}_e \bar{\mathbf{W}}_1 \right) + \beta_2^2 N_{02} N \\ &\quad - 2\beta_1 \Re \left(\mathbf{d}^H \bar{\mathbf{R}} \mathbf{P}^H \mathbf{y}_{BS} \right) + \mathbf{d}^H \mathbf{d}. \end{aligned} \quad (7.7)$$

Matrix \mathbf{H}_e denotes the equivalent channel matrix while $\bar{\mathbf{W}}_1 = \beta_2 \mathbf{W}_1$ is the unscaled relaying matrix. Matrix $\bar{\mathbf{R}} = \text{blkdiag} = (\bar{\mathbf{R}}_1, \dots, \bar{\mathbf{R}}_k, \dots, \bar{\mathbf{R}}_K) \in \mathbb{R}^{L \times N}$ is the matrix containing the upper triangular matrices derived from BD-GMD, being $\bar{\mathbf{R}}_k \in \mathbb{R}^{L_k \times N_k}$ the upper triangular matrix related to user k , which takes the first L_k rows of \mathbf{R}_k . Finally, $\mathbf{P} = [\mathbf{P}_1, \dots, \mathbf{P}_k, \dots, \mathbf{P}_K] \in \mathbb{C}^{M \times N}$.

Obviously, an error is introduced in the precoding matrix derivation if it is compared to the ideal since the relaying matrix is updated in each iteration and used for BD-GMD decomposition of the channel from the BS to the end users. However, as depicted in simulation results, we can firmly state that the algorithm converges and the error introduced has less impact than expected. Thus, considering fixed β_2 and \mathbf{W}_1 and after carrying out the procedure depicted in Section B.2, the following expression is obtained:

$$\mathbf{y}_{BS} = \frac{1}{\beta_1} \underbrace{\left(\mathbf{P} \mathbf{R}^H \mathbf{R} \mathbf{P}^H + \frac{N_{02} N}{E_R} \mathbf{H}_1^H \bar{\mathbf{W}}_1^H \bar{\mathbf{W}}_1 \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-1}}_{\mathbf{F}} \mathbf{P} \bar{\mathbf{R}}^H \mathbf{d}, \quad (7.8)$$

where $\alpha_1 = \frac{N_{01}}{E_{BS}} \text{Tr} \left(\bar{\mathbf{W}}_1^H \left(\mathbf{H}_e^H \mathbf{H}_e + \frac{N_{02} N}{E_R} \mathbf{I}_N \right) \bar{\mathbf{W}}_1 \right)$. Comparing (7.8) with the rest of expressions derived for the precoding matrix in the previous chapters, we can state that (7.8) is much simpler. In the same way, the scaling factor β_1 is defined as

$$\beta_1 = \sqrt{\frac{\mathbf{d}^H \bar{\mathbf{R}} \mathbf{P}^H \left(\mathbf{P} \mathbf{R}^H \mathbf{R} \mathbf{P}^H + \frac{N_{02} N}{E_R} \mathbf{H}_1^H \bar{\mathbf{W}}_1^H \bar{\mathbf{W}}_1 \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \mathbf{P} \bar{\mathbf{R}}^H \mathbf{d} \right)^{-2} \mathbf{P} \bar{\mathbf{R}}^H \mathbf{d}}{E_{BS}}}.$$

Once the precoding and relaying matrices are obtained, as well as the scaling factors β_1 and β_2 that constrain the power at the BS and RS, respectively, the optimal precoding

vector has to be chosen as the nearest point in the lattice. As described in Section 5.2, the perturbation vector is selected for MSE minimization. Applying (7.8) to (7.7), the MSE function is reduced to

$$\xi = \mathbf{d}^H (\mathbf{I}_L - \overline{\mathbf{R}}\mathbf{P}^H\overline{\mathbf{F}}) \mathbf{d}. \quad (7.9)$$

Carrying out the steps enumerated in [Schmidt05], by means of the Cholesky factorization of the inner term of (7.9), the triangular matrix \mathbf{L} can be obtained since:

$$(\mathbf{I}_L - \overline{\mathbf{R}}\mathbf{P}^H\overline{\mathbf{F}}) = \mathbf{L}^H\mathbf{L},$$

which simplifies the computation of the perturbation vector \mathbf{a} to:

$$\mathbf{a} = \underset{\mathbf{a}' \in \tau\mathbb{Z}^L + j\tau\mathbb{Z}^L}{\operatorname{argmin}} \|\mathbf{L}(\mathbf{s} + \mathbf{a})\|_2^2. \quad (7.10)$$

Due to the search of the closest point in a lattice, a SE is implemented at the BS to reduce the computation cost. Comparing BD-GMD with other decomposition methods, such as SVD, we observe that whilst for the former the matrix \mathbf{R} is upper triangular, for the latter \mathbf{R} is diagonal. Furthermore, when SVD is applied, the lowest value of the diagonal of \mathbf{R} limits the BER performance of the overall system. It should be pointed out that, SE algorithm considers interference cancellation to find the perturbation vector, so SE cannot take any benefit from the diagonal matrix. Moreover, GMD method gives a triangular matrix with equal diagonal elements, which eliminates the imbalance among subchannel gains.

The solution of the optimization problem is computationally intense and hard to solve due to the interdependence of the matrices in (7.8) and (7.6). For this reason, an iterative algorithm has to be applied increasing the computational complexity of the system, which monotonously decreases the MSE. Moreover, BD-GMD decomposition and the search of the perturbation vector have to be computed in each iteration, being the SE algorithm executed in each iteration.

The algorithm is described in Algorithm 14, where $\epsilon_{\min} = 0.001$ denotes the minimum convergence error, computed as $\epsilon = \|\mathbf{F}^{l+1} - \mathbf{F}^l\|_2^2 + \|\mathbf{W}_1^{l+1} - \mathbf{W}_1^l\|_2^2$ where l is the current iteration. Once the convergence error is reached, the algorithm is stopped.

Provided simulation results will show the effectiveness of BD-GMD applied to multiuser MIMO relaying systems, where the performance will be shortly penalized due to the use of the updated $\overline{\mathbf{W}}_1^{l+1}$ for BD-GMD decomposition. Nevertheless, the real improvement over previously introduced non-linear precoding schemes will be viewed in the computational cost.

Algorithm 14 Computation of \mathbf{F} , \mathbf{W}_1 and the perturbation vector \mathbf{a} under the MMSE criterion for BD-GMD based approach.

- 1: Block diagonalization of the channel \mathbf{H}_2 (See Section 3.3.3).
 - 2: Initialize the variables: $\bar{\mathbf{F}}^0 = \mathbf{I}_{M \times L}$, $\bar{\mathbf{W}}_1^0 = \mathbf{I}_{N \times R}$, $\epsilon_{\min} = 0.001$ and $\epsilon = \infty$.
 - 3: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 4: Computation of $\bar{\mathbf{W}}_1^{l+1}$ (7.6) for $\bar{\mathbf{F}}^l$.
 - 5: Calculate the equivalent channel matrix $\bar{\mathbf{H}} = \mathbf{H}_e \bar{\mathbf{W}}_1 \mathbf{H}_1 \implies$ BD-GMD computation.
 - 6: Computation $\bar{\mathbf{F}}^{l+1}$ (7.8).
 - 7: Search of the perturbation vector \mathbf{a} (7.10) \implies SE algorithm execution.
 - 8: Error calculation: $\epsilon = \|\bar{\mathbf{F}}^{l+1} - \bar{\mathbf{F}}^l\|^2 + \|\bar{\mathbf{W}}_1^{l+1} - \bar{\mathbf{W}}_1^l\|^2$.
 - 9: **if** $\epsilon \geq \epsilon_{\min}$ **then**
 - 10: $\bar{\mathbf{F}}^l = \bar{\mathbf{F}}^{l+1}$.
 - 11: $\bar{\mathbf{W}}_1^l = \bar{\mathbf{W}}_1^{l+1}$.
 - 12: **end if**
 - 13: **end while**
-

7.3 VP Transmission with Suboptimal BD-GMD

In order to avoid the complexity of joint iterative optimization of the precoding and relaying matrices based on BD-GMD decomposition described in Section 7.2, a novel design strategy is here proposed, where both matrices are calculated directly under MMSE criterion with no iterations. Thus, the main purpose of this suboptimal approach is the avoidance of the iterative algorithm for obtaining the precoding and relaying matrices.

The proposed algorithm can be summed up in three main steps. Firstly, GMD is carried out in the first hop. Secondly, BD-GMD is performed at the second hop, whilst finally VP is applied at the BS to minimize the overall mean square error and limit the power at both terminals.

By means of this suboptimal approach, the equivalent channel matrix between the BS and the end users becomes upper triangular, keeping the properties of BD-GMD decomposition, i.e, providing every subchannel with the same SNR. Therefore, GMD is applied to the first channel matrix \mathbf{H}_1 :

$$\mathbf{H}_1 = \mathbf{Q}_1 \mathbf{R}_1 \mathbf{P}_1^H, \quad (7.11)$$

where $\text{rank}(\mathbf{H}_1) = \min(R, M)$ is assumed. The channel is decomposed into three matrices: $\mathbf{Q}_1 \in \mathbb{C}^{R \times R}$ and $\mathbf{P}_1 \in \mathbb{C}^{M \times M}$, which are unitary, and $\mathbf{R}_1 \in \mathbb{R}^{R \times M}$ which is an upper triangular matrix with the same elements at the diagonal computed as in (7.2).

In order to apply BD-GMD at the second hop, BD [Spencer04b] technique is applied at the relay to cancel the interferences between users. In addition, the relaying matrix is

defined as:

$$\mathbf{W} = \mathbf{W}_2 \mathbf{W}_1,$$

being $\mathbf{W}_2 \in \mathbb{C}^{R \times N}$ employed for second hop channel block diagonalization whereas $\mathbf{W}_1 \in \mathbb{C}^{N \times R}$ is used for relaying processing. In the same way, the equalization matrix at the relay is defined as $\mathbf{W}_1 = \mathbf{W}_{1\text{post}} \mathbf{W}_{1\text{pre}}$, where $\mathbf{W}_{1\text{post}} \in \mathbb{C}^{N \times N}$ and $\mathbf{W}_{1\text{pre}} \in \mathbb{C}^{N \times R}$ are employed for the post-processing of the signal, adjusting it to the second hop, and for the pre-processing, respectively. Thus, $\mathbf{W}_{1\text{pre}} = \overline{\mathbf{Q}}_1^H \in \mathbb{C}^{R \times N}$ takes the first N columns of the matrix \mathbf{Q}_1 . Matrix $\overline{\mathbf{Q}}_1$ is defined as $\overline{\mathbf{Q}}_1 = [\overline{\mathbf{Q}}_{1,1}, \dots, \overline{\mathbf{Q}}_{1,k}, \dots, \overline{\mathbf{Q}}_{1,K}]$, being $\overline{\mathbf{Q}}_{1,k} \in \mathbb{C}^{R \times N_k}$ the pre-processing matrix belonging to user k .

After the application of GMD at the first hop, the signal transmitted by the relay is expressed as

$$\mathbf{y}_R = \beta_1 \mathbf{W}_2 \mathbf{W}_{1\text{post}} \overline{\mathbf{R}}_1 \mathbf{P}_1^H \mathbf{y}_{BS} + \beta_1 \mathbf{W}_2 \mathbf{W}_{1\text{post}} \overline{\mathbf{Q}}_1^H \mathbf{n}_1.$$

The matrix $\overline{\mathbf{R}}_1 \in \mathbb{R}^{N \times M}$ takes the first N rows of matrix \mathbf{R}_1 . It should be underlined that, due to the orthogonality of $\overline{\mathbf{Q}}_1$, the noise is not amplified where the covariance of the noise stands for $\mathbf{R}_{\mathbf{n}_1} = \mathbb{E}[\mathbf{n}_1 \mathbf{n}_1^H] = \mathbb{E}[\overline{\mathbf{Q}}_1^H \mathbf{n}_1 \mathbf{n}_1^H \overline{\mathbf{Q}}_1] = N_{01} \mathbf{I}_R$.

After scaling the signal by β_2 , which is the factor that limits the power to E_R , the signal vector received at the k^{th} terminal is

$$\hat{\mathbf{d}}_k = \underbrace{\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS} + \sum_{\substack{j=1 \\ j \neq k}}^K \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,j} \mathbf{W}_{1,j\text{post}} \left(\overline{\mathbf{R}}_{1,j} \mathbf{P}_1^H \mathbf{y}_{BS} + \overline{\mathbf{Q}}_{1,j}^H \mathbf{n}_1 \right)}_{\text{Interference}} + \underbrace{\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}_{2,k} \mathbf{W}_{1,k\text{post}} \overline{\mathbf{Q}}_{1,k}^H \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}}_{\text{Noise}},$$

where $\overline{\mathbf{R}}_{1,k} \in \mathbb{R}^{N_k \times M}$ are the rows related to user k and stacked into matrix $\overline{\mathbf{R}}_1 = [\overline{\mathbf{R}}_{1,1}^T, \dots, \overline{\mathbf{R}}_{1,K}^T]^T$.

In order to diagonalize the second hop, the constraint employed in [Spencer04b] is applied, i.e. $\overline{\mathbf{H}}_{2,j} \mathbf{W}_{2,k} = \mathbf{0}$ for $k = 1, \dots, j-1, j+1, \dots, K$. The matrix used for the diagonalization, $\mathbf{W}_{2,k}$, should lie in the null space of $\tilde{\mathbf{H}}_{2,j} = [\mathbf{H}_{2,1}^T, \dots, \mathbf{H}_{2,j-1}^T, \mathbf{H}_{2,j+1}^T, \dots, \mathbf{H}_{2,K}^T]^T \in \mathbb{C}^{N-N_k \times R}$. From the singular value decomposition of $\tilde{\mathbf{H}}_{2,j} = \tilde{\mathbf{U}}_j \tilde{\mathbf{\Lambda}}_j \left[\tilde{\mathbf{V}}_j^{(1)} \quad \tilde{\mathbf{V}}_j^{(0)} \right]^H$, where $\tilde{\mathbf{V}}_j^{(0)} \in \mathbb{C}^{R \times N_j}$ is the matrix containing the right singular vectors that form an orthogonal basis for the null space $\tilde{\mathbf{H}}_{2,j}$, becoming the right choice for the precoding matrix $\mathbf{W}_{2,j}$.

After BD application, the received signal at the k^{th} terminal is reduced to:

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k} \mathbf{H}_1 \mathbf{y}_{BS} + \underbrace{\beta_2 \beta_1 \mathbf{D}_k \mathbf{H}_{e,k} \mathbf{W}_{1,k_{\text{post}}} \bar{\mathbf{Q}}_{1,k}^H \mathbf{n}_1 + \beta_2 \mathbf{D}_k \mathbf{n}_{2,k}}_{\text{Noise}},$$

being the interferences cancelled at the second hop and $\mathbf{H}_{e,k} = \mathbf{H}_{2,k} \mathbf{W}_{2,k} \in \mathbb{C}^{N_k \times N_k}$ the equivalent channel created from the relaying terminal to user k . Then BD-GMD is applied at the second hop where GMD is carried out at the equivalent channel $\mathbf{H}_{e,k}$ as

$$\mathbf{H}_{e,k} = \mathbf{Q}_{e,k} \mathbf{R}_{e,k} \mathbf{P}_{e,k}^H, \quad (7.12)$$

where $\text{rank}(\mathbf{H}_{e,k}) = N_k$ is assumed. Whilst $\mathbf{Q}_{e,k} \in \mathbb{C}^{N_k \times N_k}$ and $\mathbf{P}_{e,k} \in \mathbb{C}^{N_k \times N_k}$ are unitary matrices, $\mathbf{R}_{e,k} \in \mathbb{R}^{N_k \times N_k}$ is an upper triangular matrix with the same diagonal elements. Defining $\mathbf{D}_k = \bar{\mathbf{Q}}_{e,k}^H$ and $\mathbf{W}_{1,k_{\text{post}}} = \mathbf{P}_{e,k}$, the total received signal at terminal k becomes

$$\hat{\mathbf{d}}_k = \beta_2 \beta_1 \bar{\mathbf{R}}_{e,k} \bar{\mathbf{R}}_{1,k} \mathbf{P}_1^H \mathbf{y}_{BS} + \beta_2 \beta_1 \bar{\mathbf{R}}_{e,k} \bar{\mathbf{Q}}_{1,k}^H \mathbf{n}_1 + \beta_2 \bar{\mathbf{Q}}_{e,k}^H \mathbf{n}_{2,k}, \quad (7.13)$$

being $\bar{\mathbf{Q}}_{e,k} \in \mathbb{C}^{N_k \times L_k}$ the first L_k column elements of matrix $\mathbf{Q}_{e,k}$. We can firmly state that, employing the unitary matrices derived from GMD and BD-GMD, the channel from the BS to the end users is upper diagonal and the received signal expression becomes simpler. In the same way, $\bar{\mathbf{R}}_{e,k} \in \mathbb{R}^{L_k \times N_k}$ denotes the first L_k rows of $\mathbf{R}_{e,k}$. As happens previously, the second hop channel is not amplified due to the orthogonality of $\bar{\mathbf{Q}}_{e,k}^H$, being the second hop covariance matrix $\mathbf{R}_{\mathbf{n}_2} = \mathbb{E} \left[\bar{\mathbf{Q}}_{e,k}^H \mathbf{n}_{2,k} \mathbf{n}_{2,k}^H \bar{\mathbf{Q}}_{e,k} \right] = \mathbb{E} \left[\mathbf{n}_{2,k} \mathbf{n}_{2,k}^H \right] = \mathbf{N}_{0_2}$.

Moreover, by means of this approach we demonstrate that the structure of the resultant matrix of the combination of GMD and BD-GMD keeps the structural properties of the application of BD-GMD from the BS to the end users.

In order to continue with the design of the precoder at the BS, the minimization of the MSE has been employed as optimization function, which is defined as

$$\begin{aligned} \xi &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{d}_k\|_2^2 \right] \\ &= \beta_1^2 \beta_2^2 \mathbf{y}_{BS}^H \mathbf{P}_1 \bar{\mathbf{R}}_1^H \bar{\mathbf{R}}_e^H \bar{\mathbf{R}}_e \bar{\mathbf{R}}_1 \mathbf{P}_1^H \mathbf{y}_{BS} + \beta_2^2 \beta_1^2 \mathbf{N}_{0_1} \text{Tr} \left(\bar{\mathbf{R}}_e^H \bar{\mathbf{R}}_e \right) + \beta_2^2 \mathbf{N}_{0_2} L \\ &\quad - 2\beta_2 \beta_1 \Re \left(\mathbf{d}^H \bar{\mathbf{R}}_e \bar{\mathbf{R}}_1 \mathbf{P}_1^H \mathbf{y}_{BS} \right) + \mathbf{d}^H \mathbf{d}, \end{aligned} \quad (7.14)$$

where $\bar{\mathbf{R}}_e = \text{blkdiag}(\bar{\mathbf{R}}_{e,1}, \dots, \bar{\mathbf{R}}_{e,k}, \dots, \bar{\mathbf{R}}_{e,K})$ and $\hat{\mathbf{d}}_k$ is defined in (7.13). Now the optimization problem can be formulated excluding \mathbf{W}_1 and \mathbf{D} . The power scaling factor β_2 is maintained due to the power limit established at the relay. Finally, the optimization problem

is defined as

$$\begin{aligned} \{\mathbf{a}, \mathbf{F}, \beta_1, \beta_2\} &= \underset{\{\mathbf{a}, \mathbf{F}, \beta_1, \beta_2\}}{\operatorname{argmin}} \quad \xi \\ \text{s.t.} \quad & \|\mathbf{y}_{BS}\|_2^2 = E_{BS}, \\ & \|\mathbf{y}_R\|_2^2 = E_R, \end{aligned}$$

where ξ is defined in (7.14).

The optimization is carried out in two steps. Firstly \mathbf{F} , β_1 and β_2 are obtained without considering \mathbf{a} . Secondly, the perturbation vector is obtained to minimize the overall MSE. The solution is solved by means of the Lagrangian multipliers, getting the following result:

$$\mathbf{y}_{BS} = \frac{1}{\beta_1} \underbrace{\left(\mathbf{P}_1 \bar{\mathbf{R}}_1^H \mathbf{A} \bar{\mathbf{R}}_1 \mathbf{P}_1^H + \frac{N_{01} \operatorname{Tr}(\mathbf{A}) \mathbf{I}_M}{E_{BS}} \right)^{-1} \mathbf{P}_1 \bar{\mathbf{R}}_1 \bar{\mathbf{R}}_e^H \mathbf{d}}_{\mathbf{F}}, \quad (7.15)$$

$$\beta_1 = \sqrt{\frac{\mathbf{d}^H \bar{\mathbf{R}}_e \bar{\mathbf{R}}_1 \mathbf{P}_1^H \mathbf{A}^{-2} \mathbf{P}_1 \bar{\mathbf{R}}_1^H \bar{\mathbf{R}}_e^H \mathbf{d}}{E_{BS}}}$$

and finally

$$\beta_2 = \sqrt{\frac{\mathbf{d}^H \bar{\mathbf{R}}_e \bar{\mathbf{R}}_1 \mathbf{P}_1^H \mathbf{P}_1 \bar{\mathbf{R}}_1^H \mathbf{A}^{-2} \bar{\mathbf{R}}_1 \mathbf{P}_1^H \mathbf{P}_1 \bar{\mathbf{R}}_1^H \bar{\mathbf{R}}_e^H \mathbf{d} + \beta_1^2 N_{01} N}{E_R}},$$

where $\mathbf{A} = \left(\bar{\mathbf{R}}_e^H \bar{\mathbf{R}}_e + \frac{N_{02} L}{E_R} \mathbf{I}_N \right)$. As a result of GMD and BD-GMD decomposition, the parameters necessary for the precoding and processing at the relay are not interdependent.

Finally, once the precoder is obtained, the perturbation vector is selected as the nearest point in a lattice that minimizes the MSE between the transmitted symbols at the source and the received symbols at destination. The mean square error is simplified to

$$\xi = \mathbf{d}^H \left(\mathbf{I}_L - \bar{\mathbf{R}}_e \bar{\mathbf{R}}_1 \mathbf{P}_1^H \left(\mathbf{P}_1 \bar{\mathbf{R}}_1^H \mathbf{A} \bar{\mathbf{R}}_1 \mathbf{P}_1^H + \frac{N_{01}}{E_{BS}} \operatorname{Tr}(\mathbf{A}) \mathbf{I}_M \right)^{-1} \mathbf{P}_1 \bar{\mathbf{R}}_1^H \bar{\mathbf{R}}_e^H \right) \mathbf{d}, \quad (7.16)$$

being $\mathbf{A} = \left(\bar{\mathbf{R}}_e^H \bar{\mathbf{R}}_e + \frac{N_{02} L}{E_R} \mathbf{I}_N \right)$. Carrying out the procedure accomplished in previous sections and introduced in [Schmidt05], the Cholesky factorization of the inner term of (7.16) is

$$\left(\mathbf{I}_L - \bar{\mathbf{R}}_e \bar{\mathbf{R}}_1 \mathbf{P}_1^H \left(\mathbf{P}_1 \bar{\mathbf{R}}_1^H \mathbf{A} \bar{\mathbf{R}}_1 \mathbf{P}_1^H + \frac{N_{01}}{E_{BS}} \operatorname{Tr}(\mathbf{A}) \mathbf{I}_M \right)^{-1} \mathbf{P}_1 \bar{\mathbf{R}}_1^H \bar{\mathbf{R}}_e^H \right) = \mathbf{L}^H \mathbf{L}.$$

This suboptimal approach, depicted in Algorithm 15, simplifies greatly the analysis, largely owing to the independence of the matrices.

Algorithm 15 Search of \mathbf{F} and the perturbation vector \mathbf{a} under the MMSE criterion for suboptimal GMD approach.

- 1: Block diagonalization of the channel \mathbf{H}_2 (See Section 3.3.3).
 - 2: Geometric mean decomposition of the first hop channel (7.11).
 - 3: Block diagonal geometric mean decomposition of the second equivalent channel (7.12).
 - 4: Computation of the equalizer $\mathbf{D} = (\mathbf{D}_1, \dots, \mathbf{D}_K)$ where $\mathbf{D}_k = \overline{\mathbf{Q}}_{e,k}^H$.
 - 5: Calculate \mathbf{W}_1 .
 - 6: Compute \mathbf{F} (7.15).
 - 7: Search of the perturbation vector \mathbf{a} (7.10) \implies SE algorithm execution.
-

7.4 Simulation Results

Simulation results are shown in this section, where six different proposals have been considered, being the following:

- *Lin-MMSE* (Section 3.2)
- *VP-MMSE-BD* (Section 5.2.3)
- *VP-MMSE-HH* (Section 6.2.1)
- *VP-MMSE-Sub* (Section 6.2.2)
- *VP-MMSE-BD-GMD* (Section 7.2)
- *VP-MMSE-GMD-Sub* (Section 7.3)

The system simulated is the one depicted in Figure 3.1, where a base station equipped with M antennas sends information through a relay equipped with R antennas to K N_k -antenna users. Only the set-up $4 \times 4 \times \{4 \times 1\}$ has been evaluated.

Results have been obtained by means of Monte Carlo simulations, being the parameters set in Section 5.3. Two different SNR configurations have been tested: fixed $\rho_2 = 15$ dB and fixed $\rho_1 = 15$ dB.

To conclude, it should be pointed out that the results are given for BER performance, achievable SR and MSE, showing the effectiveness of *VP-MMSE-BD-GMD* and *VP-MMSE-GMD-Sub*.

7.4.1 BER Performance

Figure 7.2 shows BER performance when $\rho_2 = 15$ dB (see Figure 7.2(a)) and $\rho_1 = 15$ dB (see Figure 7.2(b)), respectively. Provided BER simulations show that, when ρ_2 is set to 15 dB, *VP-MMSE-BD-GMD* performs near the local optimal *VP-MMSE-BD*, being the maximum difference lower than 0.5 dB. Unsurprisingly, due to the fact that \mathbf{W}_1 is estimated in each iteration and afterwards employed for *BD-GMD*, some degradation is introduced. This might be the reason why *VP-MMSE-BD-GMD* performs slightly worse than the local optimal *VP-MMSE-BD*.

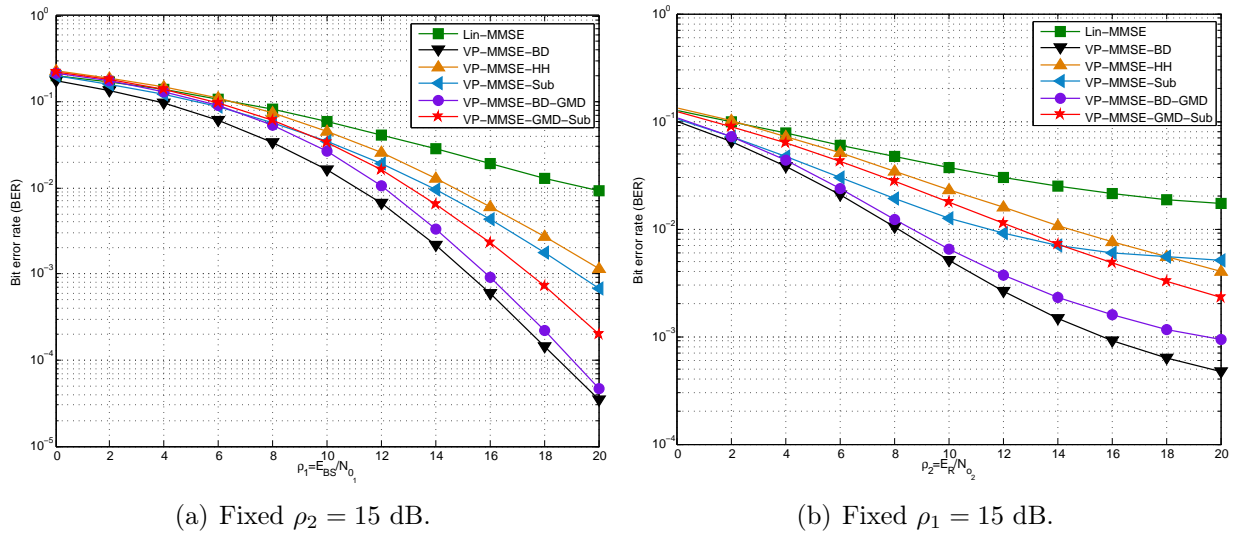


Figure 7.2: BER performance for MMSE BD-GMD systems in the $4 \times 4 \times \{4 \times 1\}$ with QPSK modulation.

Furthermore, *VP-MMSE-GMD-Sub*, designed for complexity reduction, performs near *VP-MMSE-BD-GMD* and *VP-MMSE-BD*, being the maximum difference of 2 and 2.5 dB, respectively. It should be underlined the remarkable performance improvement of GMD systems, which outperform the result of suboptimal VP approaches.

On the other hand, Figure 7.2(b) depicts the BER performance when ρ_1 is set to 15 dB. Matching up with the previous results, *VP-MMSE-BD-GMD* continues performing near *VP-MMSE-BD*, where the maximum difference is achieved at high SNR, being around 4 dB. It should be pointed out that at high SNR, from 14 on, *VP-MMSE-GMD-Sub* outperforms *VP-MMSE-Sub*. Furthermore, the main maximum difference between *VP-MMSE-BD-GMD* and *VP-MMSE-GMD-Sub* is over 6 dB, confirming the considerable improvement of GMD approaches over other suboptimal VP systems. Finally we can firmly state that obtaining the same SNR per user improves the BER performance of the systems.

7.4.2 Achievable Sum-Rate

The achievable sum-rate is shown in this section. Figure 7.3 shows the sum-rate achieved when ρ_2 and ρ_1 are set to 15 dB, respectively, as depicted in Figures 7.3(a) and 7.3(b).

When $\rho_2 = 15$ dB, *VP-MMSE-BD-GMD* achieves a lower sum-rate compared to *VP-MMSE-BD*. Furthermore, until $\rho_1 = 10$ dB, *VP-MMSE-BD-GMD* performs worse than *VP-MMSE-Sub*. *VP-MMSE-GMD-Sub* approach performs also worse than the suboptimal approaches based on VP, specially at low ρ_2 . Thus, we can state that GMD is effective for BER performance improvement but not very influential for SR maximization.

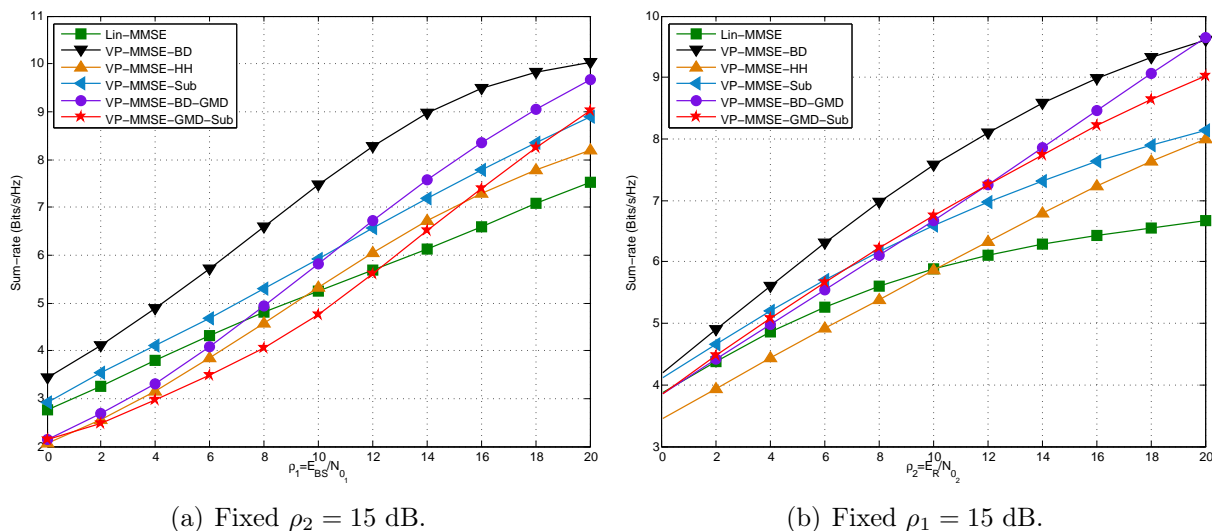


Figure 7.3: Achievable sum-rate for MMSE BD-GMD approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

The situation changes slightly when ρ_1 is set to 15 dB as depicted in Figure 7.3(b). At low SNR the sum-rate achieved by GMD approaches is the same as the one obtained by the suboptimal VP approaches. The sum-rate achieved by *VP-MMSE-BD* is not outperformed by the systems proposed in this section but the difference respect to fixed ρ_2 configuration is lower. It should be underlined that when ρ_1 is fixed to 15 dB, the sum-rate achieved by *VP-MMSE-GMD-Sub* is better than the one obtained by the suboptimal VP approaches.

7.4.3 MSE Measurement

The convergence and the MSE error are measured in this section by means of two different simulations. Taking into consideration the iterative nature of *VP-MMSE-BD-GMD* approach, BER performance of the system is measured when the number of iterations is fixed and then the MSE error is measured for a fixed number of iterations, showing for *VP-MMSE-BD-GMD* approach that increasing the number of iterations the error is reduced. Whilst

the former simulation considers $\rho_2 = 15$ dB, for the latter simulation the SNRs are set in pairs, therefore $\{ \rho_1 = 10$ dB and $\rho_2 = 15$ dB $\}$ and $\{ \rho_1 = 20$ dB and $\rho_2 = 15$ dB $\}$.

Figure 7.4 shows the BER performance when the number of iterations (it) is fixed to 1, 3, 6 and 10. Unsurprisingly, we observe that increasing the number of iterations improves the performance, reducing the difference respect to the local optimal $VP-MMSE-BD-GMD$. Figure 7.4 shows that more than 10 iterations are needed for being near to the local optimal $VP-MMSE-BD-GMD$.

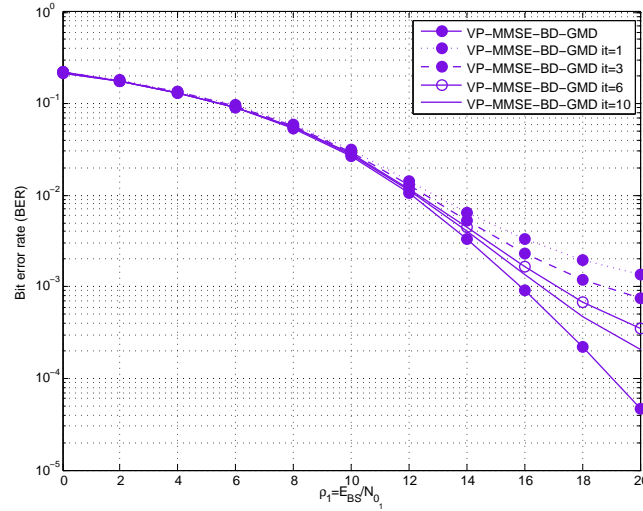


Figure 7.4: BER curves of MMSE BD-GMD in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) with QPSK modulation when $\rho_2 = 15$ dB and the number of iterations is fixed.

Figure 7.5 depicts the MSE when $\{ \rho_1 = 10$ dB and $\rho_2 = 15$ dB $\}$ (Figure 7.5(a)) and $\{ \rho_1 = 20$ dB and $\rho_2 = 15$ dB $\}$ (Figure 7.5(b)). Watching Figure 7.5(a) we observe that the error difference between $VP-MMSE-BD$ and $VP-MMSE-BD-GMD$ is negligible. Furthermore, $VP-MMSE-BD-GMD$ converges in a few number of iteration. It can be seen that GMD proposals reduce the error more than the suboptimal VP approaches. Finally it should be said that the MSE of $VP-MMSE-GMD-Sub$ is around 0.85, matching up with $VP-MMSE-BD-GMD$ when $it = 1$.

The same conclusions hold when $\rho_1 = 20$ dB and $\rho_2 = 15$ dB, as depicted in Figure 7.5(b). The main difference resides on the considerable error improvement of GMD proposals ($VP-MMSE-BD-GMD$ and $VP-MMSE-GMD-Sub$) over $VP-MMSE-BD$, being the error 0.12 and 0.22 once the convergence is reached. It is remarkable to underline that $BD-MMSE-GMD-Sub$ obtains a lower error compared with $VP-MMSE-BD$. To conclude, it should be said that GMD approaches achieve a lower MSE compared to the suboptimal VP systems.

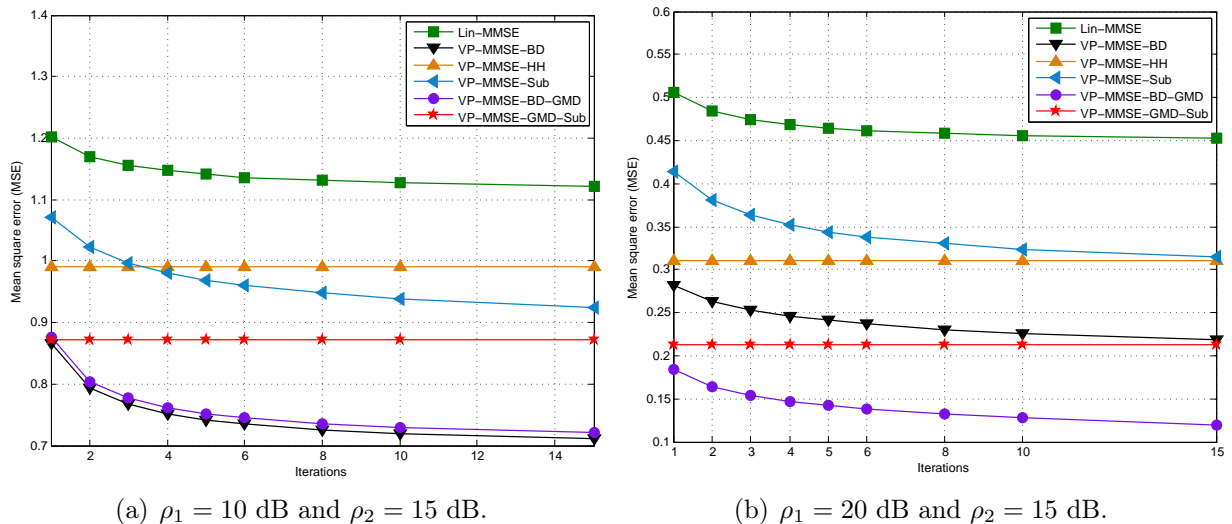


Figure 7.5: Error measurement for MMSE BD-GMD approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A).

7.5 Complexity Analysis

The computational complexity is analysed in this section by means of the complexity order of the algorithm and the number of floating point operations. In addition, the run-time required by each approach for the computation of the precoding and relaying matrices is given, computed as the addition of the run-time measured at three different terminals.

For simplicity, $M = R = N = L = n$ is assumed while N_B is the length of the transmitted data block. Moreover, $T = |\nu|$ stands for the number of points that form the lattice and z is the rank of the channel matrix. To conclude, it denotes the total number of executed iterations, which is considered fixed for simplicity.

7.5.1 Number of FLOPs and Computational Complexity Analysis

Table 7.1 shows the number of FLOPs for the computation of the precoder and the relaying matrix for *VP-MMSE-BD-GMD* and *VP-MMSE-GMD-Sub* proposals. We observe that the number of FLOPs is greater for *VP-MMSE-BD-GMD* than for *VP-MMSE-GMD-Sub*. Surprisingly, the number of FLOPs required for \mathbf{W} computation in *VP-MMSE-GMD-Sub* approach is remarkably low, reducing it to $(2N + 1)NR$.

In addition, Table 7.2 shows the number of FLOPs necessary for the computation of the perturbation vector through the SE algorithm, the modulo operation and the GMD decomposition, being m and l the dimensions of the channel matrix $\mathbf{H} \in \mathbb{C}^{m \times l}$ and z is the rank of the matrix computed as $z = \text{rank}(\mathbf{H})$.

Finally Table 7.3 depicts the complexity order and the number of FLOPs when $M = R = N = n$ and $M = R = N = 4$. The complexity order depends directly on the number of points

| Algorithm | Operation | FLOPs |
|-----------------|-----------|--|
| VP-MMSE-BD-GMD | F | $4M^3 + 2(9+N+R)M^2 + 6N^2M + (1+2N)ML + (4N+1)MR + (2N+3)M$ |
| | W | $8N^3 + 4R^3 + (10+2L)N^2 + (4N+9)R^2 + 2(M+1)NR + 2NL^2 + (3+M+L)N + (4+5M)R + 3$ |
| VP-MMSE-GMD-Sub | F | $8M^3 + (2L+4+2M)N^2 + (6N+10)M^2 + (2L+3)MN + (3+L)M + 3$ |
| | W | $(2N+1)NR$ |

 Table 7.1: Number of FLOPs required by BD-GMD approaches for the computation of **F** and **W**.

| Operation | FLOPs |
|------------------|---|
| Sphere encoder | $\left(\frac{T^{K+1}-T}{T-1} + K\right) N_B$ [Mohaisen11] |
| Modulo Operation | KN_B |
| GMD | $(m+l)z$ [Jiang05b] |

Table 7.2: Number of FLOPs for sphere encoder, modulo operation and GMD algorithm.

that compose the lattice, which is assumed to be $T = 25$, and on n for *VP-MMSE-GMD-Sub*, whilst the number of iterations should be added to this values for *VP-MMSE-BD-GMD* approach. As expected, the number of FLOPs required by the suboptimal GMD approach is lower. Furthermore, the amount of FLOPs in *VP-MMSE-BD-GMD* is limited by the number of iterations, increasing considerably the cost of the suboptimal.

| Algorithm | Comp. order | FLOPs | |
|-----------------|-----------------|--|--------------------|
| | | M=R=N=n | M=R=N=4 |
| VP-MMSE-BD-GMD | $O(it T^{n+1})$ | $it (36n^3 + 56n^2 + 16n + 9 + N_B (\frac{T^{n+1}-T}{T-1} + n))$ | $4 \times 10^7 it$ |
| VP-MMSE-GMD-Sub | $O(T^{n+1})$ | $(25n^3 + 23n^2 + 7n + 3 + N_B (\frac{T^{n+1}-T}{T-1} + n))$ | 4×10^7 |

 Table 7.3: Complexity order for *VP-MMSE-BD-GMD* and *VP-MMSE-GMD-Sub*.

7.5.2 Running-Time Analysis

In order to show the running-time of GMD systems in more detail two different figures are provided showing the running-time required for iterative approaches in one figure and the one needed by the non-iterative proposals in the other.

Figure 7.6(a) depicts the run-time needed for the iterative *VP-MMSE*, *VP-MMSE-BD* and *VP-MMSE-BD-GMD* systems. We observe that the running-time is higher when the SNR is higher. Furthermore, the time required for *VP-MMSE-BD* computation is reduced considerably by *VP-MMSE-BD-GMD*, being the maximum run-time of 23 seconds per Monte Carlo for *VP-MMSE-BD-GMD*, whilst 38 seconds per Monte Carlo are required by *VP-MMSE-BD*. However, the running-time of *VP-MMSE-BD-GMD* is higher in comparison to *VP-MMSE*, due to the block diagonalization process and GMD decomposition.

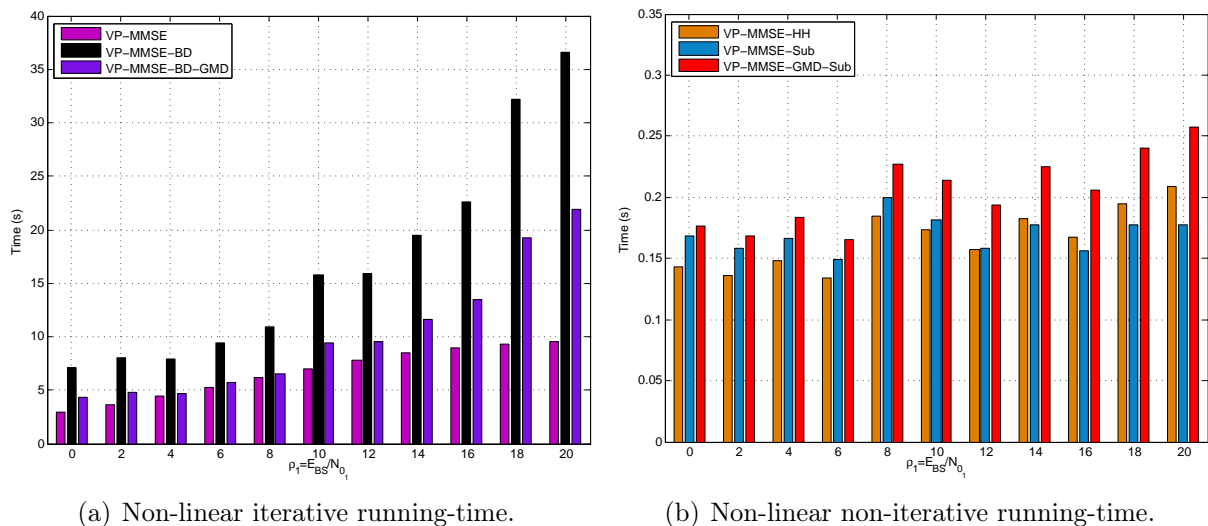


Figure 7.6: Running-time for MMSE BD-GMD approaches in the $4 \times 4 \times \{4 \times 1\}$ set-up (A) when $\rho_2=15$ dB.

Finally, Figure 7.6(b) depicts the run-time necessary for the suboptimal VP systems. The conclusions obtained for the previous iterative approaches are not the same, where the running-time does not increase with the SNR. The maximum running-time is obtained at 20 dB, needing 0.26, 0.22 and 0.18 seconds per Monte Carlo for *VP-MMSE-GMD-Sub*, *VP-MMSE-HH* and *VP-MMSE-Sub* approaches, respectively.

It should be pointed out that the main objective of *VP-MMSE-GMD-Sub* is the reduction of the computational complexity presented by *VP-MMSE-BD-GMD* as shown in Figure 7.6(a), being the maximum execution time of 0.26 seconds for the former and 20 seconds for the latter.

Surprisingly, the running-time obtained for *VP-MMSE-BD-GMD* in comparison to *VP-MMSE-BD* is lower in spite of the decomposition process. The main reason might be that the same SNR is obtained by each user cancelling the imbalance among subchannel gains, which

directly affects to the perturbation vector search. By means of GMD decomposition the search and the optimization are accomplished faster. Furthermore, the number of iterations needed for the convergence of the local optimal GMD system is lower.

7.5.3 Complexity-Performance Trade-Off

Chapter 5 and 7 analyse the use of non-linear transmitter in multiuser MIMO relaying networks, being the precoding and relaying matrices designed jointly. The former introduces firstly the optimal non-linear THP and VP designs, while the latter applies BD-GMD for matrix derivation.

The main drawback of the application of non-linear precoding techniques resides on the computational cost. For example, when THP transmitter is chosen the modulo operation executed on the iterative loop and the successive interference cancellation increase considerably the computational cost. On the other hand, when VP is employed apart from computing BD, due to lack of knowledge of the statistics of perturbation vector a SE algorithm has to be executed in each iteration for the search of the perturbation vector in a lattice formed by T values. Indeed, the complexity will be limited by this search. As happened with linear approaches, suboptimal non-linear approaches have been proposed in order to reduce the computational complexity. Thus a trade-off between performance and cost should be found for the efficient design of the matrices.

On one hand, *THP-MMSE-Sub* is proposed as an alternative to *THP-MMSE* optimal approach. By means of this, the number of FLOPs is reduced to 21109 thanks to the cancellation of the iterative algorithm employed by *THP-MMSE*. A penalty in performance is introduced due to suboptimal design. Nevertheless, this outperforms the local optimal linear approach.

The suboptimal approach *VP-MMSE-HH* obtains the precoding and the relaying matrices minimizing the MSE of each hop independently, performing near *THP-MMSE-opt*. An example of complexity reduction can be found on the execution time required by each of the approaches. While *VP-MMSE* needs 40 seconds, *VP-MMSE-HH* reduces it to 0.25 seconds.

Despite the execution of an iterative algorithm, *VP-MMSE-Sub* minimizes the computational cost considerably. This approach avoids the search of the perturbation vector in each iteration, reducing the complexity order from $O(it T^{n+1})$ to $O(T^{n+1})$. Moreover, this approach is efficient in performance, outperforming *THP-MMSE* and performing near *VP-MMSE*.

Finally, *VP-MMSE-GMD-Sub* suboptimal approach is proposed as an alternative to *VP-MMSE-BD-GMD* approach. By means of this the MSE of each hop is reduced independently, suppressing the iterative loop accomplished by the optimal. This system perform really close to the optimal and reduce in the same time the cost. For example the running-time is reduced from 22 seconds to 0.28 seconds.

We can firmly state that *VP-MMSE-Sub* and *VP-MMSE-GMD-Sub* approaches are outstanding candidates for the efficient design of precoding and relaying matrices, achieving a reasonable performance with a reduced computational cost.

7.6 Chapter summary

This chapter has dealt with the joint design of the precoding and relaying matrices under MMSE criterion and using BD-GMD decomposition for matrix design. In order to overcome the problem presented by SVD decomposition, where the performance of the systems is limited by the lowest singular value, BD-GMD decomposes the wireless channel into multiple parallel subchannels with the same SNR per user.

The first part of this chapter has introduced the GMD decomposition, which can be easily obtain by means of SVD and the the system model has been introduced. Afterwards, the joint MMSE solution has been derived for the precoding and relaying matrices based on GMD decomposition, where VP non-linear precoding technique has been applied at the BS to reduce the overall error and limit the power at both BS and RS terminals. By means of BD-GMD the equivalent channel created from the transmitter to the receiver is decomposed. The obtained solution has been shown complexity intense due to the interconnection between matrices.

In order to provide a simpler solution, a suboptimal solution has been proposed where the first hop and second channels are decomposed through GMD and BD-GMD. This non-iterative solution simplifies the processing at both terminals and the decomposition maintains the properties. Therefore, through the partial application of GMD and BD-GMD, the channel is BD-GMD decomposed.

Finally, simulation results have showed the effectiveness of the aforementioned decomposition in multiuser MIMO relaying system, where the complexity is lower compared with other non-linear iterative solutions. Furthermore, as a consequence of the same SNR the perturbation vector is optimized faster, reducing in this way the running time. Nevertheless, BD-GMD systems demonstrate its effectiveness in BER performance, being not as effective in sum-rate terms.

This chapter contributes to the first derivation of precoding and relaying matrices under BD-GMD decomposition and VP application. BD-GMD leads to an efficient search of the perturbation vector, reducing its computational cost. Furthermore, the suboptimal VP BD-GMD solution has been adapted first to multiuser multihop systems, reducing the complexity of the design significantly. Extensive simulation results have been given, demonstrating the effectiveness of the decomposition in combination with VP. Furthermore, a complete complexity analysis has been provided in order to compare the performance of the systems.

Summary and Conclusions

This chapter summarizes the work developed during this PhD dissertation, along with its main contributions and associated publications. Principally, this PhD dissertation deals with the design of precoding algorithms for multiuser MIMO-BC AF relaying frameworks.

Chapter 1 has given a brief introduction about the motivation and the main objectives of this dissertation. Furthermore, the main contributions and the structure of this thesis have been introduced.

Chapter 2 has introduced the theoretical background of the wireless MIMO technology, giving the main features and defining the capacity of this technology. This chapter has continued with the analysis of multiuser MIMO systems, where multiple users are forced to share resources such as time, bandwidth or transmission power. The main features of uplink and downlink channels have also been summarized in this chapter. Finally, the main research of this dissertation has been introduced, which is multiuser MIMO relaying. Firstly, relaying strategies have been presented, making a distinction between non-regenerative and regenerative models. As their name indicates, non-regenerative relaying or amplify-and-forward strategy only amplifies the received signal before transmission, oppositely to regenerative strategies, where the signal is decoded and regenerated before forwarding it. The general multiuser MIMO relaying model has been also presented in Chapter 2, where a base station sends data to multiple users through a relay. Assuming the full knowledge of the channels, a precoding stage is included at the BS, being the precoding techniques classified as linear and non-linear.

Chapter 3 has contributed into the joint linear design of precoding and relaying matrices, where both ZF and MMSE solutions have been given. As a consequence of the non-convexity of the optimization problem, obtaining a global solution has been shown unfeasible. A local optimal solution has been given by means of a numerical method that obtains a precoder for a known relaying matrix and vice versa. Several simulation results have shown the effectiveness of MMSE approaches over ZF systems, being the performance limited by the SNR of the first hop. This conclusion holds for the rest of the approaches presented in this PhD dissertation. In multiuser MIMO systems, BD has been demonstrated to be a good alternative to linear processing when several antennas are used per user terminal, since

multiuser interference is cancelled mapping each user signal on the null space created by the interfering users. Provided results have shown the lack of convergence in multiuser MIMO relaying systems. In order to provide a local optimal solution, linear precoding has been combined with BD, being applied the former at the base station for user separation and the latter at the relay for interference suppression. Simulation results have concluded that BD is more effective than the local optimal linear approach when users are equipped with multiple antennas.

Chapter 4 has proposed the novel suboptimal linear designs of the precoding and relaying matrices with the aim of reduce the computational complexity of the local optimal approaches presented in the previous chapter. Firstly, a naive solution has been suggested where all-pass relaying has been settled at the relay and BD has been performed at the BS for interference cancellation. Afterwards, a simple hop-by-hop independent optimization has been proposed, being the MSE of each hop minimized independently. In both approaches, due to the non-iterative nature of the matrices, the computational complexity is reduced substantially. Finally, system diagonalization based approaches have been given for MSE minimization and sum-rate maximization. By means of these approaches, multiple parallel subchannels are created from the BS to the end users, deriving the optimal power allocation matrices by means of the water-filling algorithm. The main advantage of these proposals are the performance improvement and the minimization of the complexity as a consequence of the scalarization of the problem. However, the main disadvantage resides on the impossibility to serve all the users with some channel conditions. In order to overcome this problem, two algorithms have been proposed based on the equally distributed power allocation and user selection based algorithms. The latter performs near the local optimal approaches as a consequence of the diversity introduced by user selection.

Taking into consideration the effectiveness of non-linear precoding transmitters in multiuser MIMO systems, these have been considered in Chapter 5 for the joint design of the matrices, deriving the MMSE solution for THP and VP transmitters. Apart from giving the local optimal solutions for these precoding techniques, a novel proposal has been considered, combining BD at the relay for interference suppression and VP at the BS for total error minimization. It should be underlined that VP has been considered first here for multiuser MIMO relaying, applying it at the base station for the reduction of the overall MSE. Despite its better performance, VP based designs have showed a disadvantage. Due to the interconnection of the filters, an iterative process has to be applied and the perturbation vector has to be found in an infinite lattice at each iteration. Thus, the complexity suffers a considerable increment. In order to reduce the computational cost of the search, tree search algorithms are executed and the lattice has been limited to a certain values. Provided simulation results have shown the effectiveness of non-linear techniques, outperforming linear performance considerably.

In order to reduce the computational cost of these approaches, some novel design strategies for non-linear transmitters have been given in Chapter 6. The iterative process has been avoided for THP design obtaining the relaying matrix throughout the minimization of the MSE of the second hop, whilst THP has been applied at the BS for successive interference cancellation. Another solution resides on the minimization of each hop independently obtaining a non-dependant solution and applying VP for the overall error reduction. Finally, the optimization problem has been divided into a master problem and sub-problem with the aim of avoiding the search of the perturbation vector in each iteration. By means of this, the filters have been searched iteratively and, once the filters have been optimized, the perturbation vector is searched for error reduction. A complete computational analysis has been given to demonstrate the effectiveness of this approaches, which, apart from reducing the cost, perform near the optimal solutions.

Finally, an interest decomposition approach has been considered in Chapter 7. By means of this, the same SNR is obtained per user, which apart from improving the performance, reduces the cost. BD-GMD has been used in combination with VP for the joint design of the matrices. Furthermore, it should be underlined that this decomposition cancels the imbalance among streams, reducing the time needed for the search the perturbation vector that minimizes the MSE from the BS to the end users. As happened in previous chapters, the perturbation vector, chosen as the nearest point in the lattice, has to be search in each iteration as a consequence of the interconnection between matrices, increasing the computational cost. With the aim of reducing it, a suboptimal approach based on VP BD-GMD has been given, reducing considerably the cost, while performing near the optimal VP systems.

8.1 Thesis Contributions

The main contributions of this research work are the following:

- Implementation and analysis of suboptimal linear precoding algorithms [Jimenez11] [Jimenez13a] (Submitted).
- Adaptation of block diagonalization technique to multiuser multihop scenarios. This algorithms have been proposed for both linear and vector precoding techniques [Jimenez13a] (Submitted).
- Suboptimal linear precoding analysis based on SVD decomposition, being mean square error minimization and sum-rate maximization metrics considered for the derivation of the precoding and relaying matrices [Jimenez13a] (Submitted).
- Power allocation algorithms based on equally power distribution and user selection

for the cases where not all the users can be served with the precoding techniques [Jimenez13a] (Submitted).

- Vector precoding at multiuser MIMO relaying scenarios. The local optimal solution for the design of precoding and relaying matrices for vector precoding and THP has been proposed in [Jimenez12a] [Jimenez12b].
- Design and implementation of suboptimal algorithms for the minimization of the computational complexity for non-linear precoding [Jimenez12b] [Jimenez11].
- Design and implementation of local optimal approaches based on geometric decomposition for linear and vector precoding [Jimenez11] [Jimenez13b] (Submitted).
- Computational cost analysis and extensive simulation of the algorithms [Jimenez13a] [Jimenez13b] (Submitted).

8.2 Suggestions for Further Research

Many issues described in this PhD dissertation can be addressed in the future as improvements and extensions of the current work. These are some of the suggestions for further research:

- The direct link established between the base station to the end users has been neglected for simplicity. As future research work, this can be considered for the design of the filters and the receivers algorithms.
- Both channels (first and second hop) are considered perfectly known at the terminals. Channel estimation and feedback techniques can be applied for the acquisition of the channels. In addition, imperfect CSI can be assumed, introducing a new design challenge. By means of this, an error is introduced in the channel models, requires a more robust design of the filters.
- The hop-by-hop suboptimal solutions, based on the minimization of the mean square error independently, might be beneficial when more than a relay is introduced between the base station to the end users. It would be really challenging the analysis of this scenario, as well as to provide a general solution.
- Taking into account the characteristics of BD-GMD decomposition, which triangularized the channel from the BS to the users, the combination of this decomposition technique with THP may be an interesting research challenge.

- We demonstrate that SVD decomposition limits the performance of the system due to the lowest singular value. Taking into consideration that GMD decomposition assures multiple parallel subchannels with the same SNR, this decomposition technique should be considered in combination with water-filling algorithm for the design of the filters. In this way, the negative power allocation is avoided, keeping away from the application of user selection algorithms and/or equally power allocation techniques.
- At the scenarios based on vector precoding, it is observed that the main drawback comes from the lack of knowledge of the statistics of the perturbation vector. A complete analysis of these statistics is required in order to reduce the computational complexity of the systems.
- In order to find a global optimal solution for the joint design of the precoding and relaying matrices interior point methods, also referred as barrier methods, might be considered for convex optimization, which can be efficiently solved by means of Newton's method with equality constraint.
- This PhD dissertation has focused on the theoretical analysis of the precoding and relaying algorithms. Hardware implementation of this algorithms may be an interesting future research challenge considering the high computational complexities indeed.

Appendix A

Publications

The following papers have been published or are under preparation for publication in refereed journal and conference proceedings.

Journal paper:

- I. Jimenez, M. Mendicute, M. Barrenechea and E. Arruti, “Block diagonalization for MMSE transceiver design in multiuser MIMO AF relaying systems”, submitted to *EURASIP Journal on Wireless Communications and Networking*.
- I. Jimenez, M. Mendicute and E. Arruti, “Block diagonalization with MMSE vector precoding for AF multiuser MIMO relaying systems”, submitted to *IEEE Communication Letters*.

International conference papers:

- I. Jimenez, M. Barrenechea, M. Mendicute and E. Arruti, “Iterative joint MMSE design of relaying MIMO downlink schemes with non-linearly precoded transmission”, in *IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp.424-428, Cesme, Turkey, June 2012.
- I. Jimenez, M. Barrenechea, M. Mendicute and E. Arruti, “Non-linear precoding approaches for non-regenerative multiuser MIMO relay systems”, in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 1399-1403, Bucharest, Romania, August 2012.
- I. Jimenez, S. Weiss, M. Mendicute and E. Arruti, “Multiuser MIMO amplify-and-forward relaying with vector precoding”, in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp.514-519, Bilbao, Spain, December 2011.

- M. Barrenechea, M. Mendicute, I. Jimenez and E. Arruti, “Implementation of complex enumeration for multiuser MIMO vector precoding”, in *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, pp. 739-743, Barcelona, Spain, August 2011.
- M. Barrenechea, L. Barbero, I. Jimenez, E. Arruti and M. Mendicute, “High-throughput implementation of tree-search algorithms for vector precoding”, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1689-1692, Czech Republic, Prague, May 2011.

Joint Filter Design Analysis

This appendix sums up the analysis accomplished for the joint filter designs for ZF and MMSE criteria.

B.1 Zero Forcing Based Design

This section analyses the joint filter design under ZF criterion. Before start, it should be underlined that the constraint set to fulfil with the ZF criterion is $\beta_2\beta_1\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{F}\mathbf{R}_d - \mathbf{R}_d = \mathbf{0}_L$. After setting this constraint into the mean square error equation, this is reduced to:

$$\xi_{ZF} = \beta_2^2\beta_1^2N_{01}\text{Tr}(\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{W}^H\mathbf{H}_2^H\mathbf{D}^H) + \beta_2^2N_{02}\text{Tr}(\mathbf{D}\mathbf{D}^H).$$

The optimization problem is defined as

$$\begin{aligned} \{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\} &= \underset{\{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\}}{\text{argmin}} \quad \xi_{ZF} \\ \text{s.t} \quad &\|\mathbf{y}_{BS}\|_2^2 = E_{BS}, \\ &\|\mathbf{y}_R\|_2^2 = E_R, \\ &\beta_2\beta_1\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{F}\mathbf{R}_d = \mathbf{R}_d. \end{aligned}$$

Due to the equality constraints, the solution is obtained through the Lagrangian multipliers, getting

$$\begin{aligned} \zeta(\mathbf{F}, \mathbf{W}, \beta_1, \beta_2, \gamma_1, \gamma_2, \boldsymbol{\Lambda}) &= \xi_{ZF} - \gamma_1(\text{Tr}(\mathbf{F}\mathbf{R}_d\mathbf{F}^H) - E_{BS}) - \gamma_2(\beta_1^2\text{Tr}(\mathbf{W}(\mathbf{H}_1\mathbf{y}_{BS}\mathbf{y}_{BS}^H \\ &\quad \mathbf{H}_1^H + N_{01}\mathbf{I}_R)\mathbf{W}^H) - E_R) - 2\Re(\text{Tr}(\boldsymbol{\Lambda}(\beta_2\beta_1\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{F}\mathbf{R}_d - \mathbf{R}_d))). \end{aligned}$$

The solution is get in an alternating fashion. Thus, for a fixed \mathbf{F} and β_1 , \mathbf{W} and β_2 are derived and vice versa. Firstly the derivative of ζ respect to the relaying matrix \mathbf{W} is set

and made equal to zero as

$$\begin{aligned} \frac{\partial \zeta(\dots)}{\partial \mathbf{W}} &= -\gamma_2 \beta_1^2 \mathbf{W} (\mathbf{H}_1 \mathbf{F} \mathbf{R}_d \mathbf{F}^H \mathbf{H}_1^H + \mathbf{N}_{0_1} \mathbf{I}_R) - \beta_2 \beta_1 \mathbf{H}_2^H \mathbf{D}^H \mathbf{\Lambda}^H \mathbf{F}^H \mathbf{H}_1^H \\ &+ \beta_2^2 \beta_1^2 \mathbf{N}_{0_1} \mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 \mathbf{W} = \mathbf{0}_R. \end{aligned}$$

Knowing that $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$ is the unscaled version of the precoding matrix, the previous equation is reduced to

$$-\gamma_2 \mathbf{W} (\mathbf{H}_1 \bar{\mathbf{F}} \mathbf{R}_d \bar{\mathbf{F}}^H \mathbf{H}_1^H + \mathbf{N}_{0_1} \beta_1^2 \mathbf{I}_R) - \beta_2 \mathbf{H}_2^H \mathbf{D}^H \mathbf{\Lambda}^H \bar{\mathbf{F}}^H \mathbf{H}_1^H + \beta_2^2 \beta_1^2 \mathbf{N}_{0_1} \mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 \mathbf{W} = \mathbf{0}_R,$$

and knowing that the ZF constraint $\beta_2 \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1^H \bar{\mathbf{F}} = \mathbf{I}_L$, after some manipulations we get

$$\mathbf{W} = \frac{1}{\beta_2} (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2)^{-1} \mathbf{H}_2^H \mathbf{D}^H \mathbf{R}_d \bar{\mathbf{F}}^H \mathbf{H}_1^H (\mathbf{H}_1 \bar{\mathbf{F}} \mathbf{R}_d \bar{\mathbf{F}}^H \mathbf{H}_1^H)^{-1}.$$

For getting \mathbf{F} , \mathbf{W} and β_2 are considered known. After setting the derivatives of the Lagrangian respect to \mathbf{F} equal to zero and assuming the unscaled version of the relaying matrix $\bar{\mathbf{W}} = \beta_2 \mathbf{W}$, we get

$$\frac{\partial \zeta(\dots)}{\partial \mathbf{F}} = -\gamma_1 \mathbf{F} + \alpha_2 \beta_1^2 \mathbf{H}_1^H \bar{\mathbf{W}}^H \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{F} - \beta_1 \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{\Lambda}^H = \mathbf{0}_{M \times L},$$

being $\alpha_2 = -\gamma_2 / \beta_2^2$. After some manipulations, we obtain

$$\mathbf{F} = \frac{1}{\beta_1} \left(-\gamma_1 \mathbf{I}_M + \alpha_2 \beta_1^2 \mathbf{H}_1^H \bar{\mathbf{W}}^H \bar{\mathbf{W}} \mathbf{H}_1 \right)^{-1} \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{\Lambda}^H. \quad (\text{B.1})$$

From the ZF constraint, replacing (B.1) into $\beta_1 \mathbf{D} \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{F} = \mathbf{I}_L$, we get

$$\mathbf{\Lambda}^H = \beta_1^{-2} \left(\mathbf{D} \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \left(-\gamma_1 \mathbf{I}_M + \alpha_2 \beta_1^2 \mathbf{H}_1^H \bar{\mathbf{W}}^H \bar{\mathbf{W}} \mathbf{H}_1 \right)^{-1} \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \right). \quad (\text{B.2})$$

Replacing (B.2) into (B.1) and after some manipulations, the precoding matrix designed under ZF criterion stands for

$$\mathbf{F} = \frac{1}{\beta_1} \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \left(\mathbf{D} \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \right)^{-1}.$$

B.2 Minimum Mean Square Error Based Design

In the same way, the filters under MMSE criterion are designed in this section. Therefore, the main aim of this design is the minimization of the mean square error, defined as:

$$\begin{aligned}
 \xi_{MMSE} &= \sum_{k=1}^K \mathbb{E} \left[\|\hat{\mathbf{d}}_k - \mathbf{d}_k\|_2^2 \right], \\
 &= \sum_{k=1}^K \beta_1^2 \beta_2^2 \mathbf{y}_{BS}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{H}_{2,k}^H \mathbf{D}_k^H \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2^2 \beta_1^2 N_{01} \text{Tr} (\mathbf{W}^H \mathbf{H}_{2,k}^H \mathbf{D}_k^H \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W}) \\
 &\quad + \beta_2^2 N_{02} \text{Tr} (\mathbf{D}_k^H \mathbf{D}_k) - 2\beta_2 \beta_1 \Re (\mathbf{d}_k^H \mathbf{D}_k \mathbf{H}_{2,k} \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS,k}) + \mathbf{d}_k^H \mathbf{d}_k, \\
 &= \beta_2^2 \beta_1^2 \mathbf{y}_{BS}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS} + \beta_2^2 \beta_1^2 N_{01} \text{Tr} (\mathbf{W}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 \mathbf{W}) \\
 &\quad + \beta_2^2 N_{02} \text{Tr} (\mathbf{D}^H \mathbf{D}) - 2\beta_2 \beta_1 \Re (\mathbf{d}^H \mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS}) + \mathbf{d}^H \mathbf{d}. \tag{B.3}
 \end{aligned}$$

The optimization problem is

$$\{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\} = \underset{\{\mathbf{F}, \mathbf{W}, \beta_1, \beta_2\}}{\text{argmin}} \quad \xi_{MMSE}$$

$$\text{s.t} \quad \|\mathbf{y}_{BS}\|_2^2 = E_{BS}, \tag{B.4a}$$

$$\|\mathbf{y}_R\|_2^2 = E_R, \tag{B.4b}$$

where ξ_{MMSE} denotes the MSE error defined in (B.3). The solution, subject to two power constraints at the BS (B.4a) and at the relay (B.4b), respectively, is get under Lagrangian multipliers in an alternating fashion. For a fixed \mathbf{W} and β_2 , \mathbf{F} and β_1 are optimized and vice versa, getting a local optimal solution.

$$\begin{aligned}
 \zeta (\mathbf{F}, \mathbf{W}, \beta_1, \beta_2, \gamma_1, \gamma_2) &= \xi_{MMSE} - \gamma_1 (\mathbf{y}_{BS}^H \mathbf{y}_{BS} - E_{BS}) - \gamma_2 (\beta_1^2 \mathbf{y}_{BS}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{W} \mathbf{H}_1 \mathbf{y}_{BS} \\
 &\quad + \beta_1^2 \text{Tr} (\mathbf{W}^H \mathbf{W}) - E_R). \tag{B.5}
 \end{aligned}$$

Assuming known \mathbf{F} and β_1 , after setting the derivatives of (B.5) respect to \mathbf{W} and making this equal to zero, we get

$$\begin{aligned}
 \frac{\partial \zeta (\dots)}{\partial \mathbf{W}} &= \beta_2^2 (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \mathbf{W} (\mathbf{H}_1 \bar{\mathbf{y}}_{BS} \bar{\mathbf{y}}_{BS}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R) - \beta_2 \mathbf{H}_2^H \mathbf{D}^H \mathbf{d} \bar{\mathbf{y}}_{BS}^H \mathbf{H}_1^H \\
 &= \mathbf{0}_R, \tag{B.6}
 \end{aligned}$$

where $\bar{\mathbf{y}}_{BS}$ are the unscaled precoded symbols defined as $\bar{\mathbf{y}}_{BS} = \beta_1 \mathbf{y}_{BS}$ and $\alpha_2 = -\gamma_2 / \beta_2^2$.

Multiplying (B.6) by \mathbf{W}^H , applying the trace operation and taking into consideration that $\text{Tr} (\mathbf{A}\mathbf{B}) = \text{Tr} (\mathbf{B}\mathbf{A})$, we obtain

$$\beta_2 \text{Tr} (\mathbf{W}^H (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \mathbf{W} (\mathbf{H}_1 \bar{\mathbf{y}}_{BS} \bar{\mathbf{y}}_{BS}^H \mathbf{H}_1^H + \beta_1^2 N_{01} \mathbf{I}_R)) = \text{Tr} (\mathbf{D} \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \bar{\mathbf{y}}_{BS} \mathbf{d}^H),$$

where a Hermitian operation has been applied. From the power constraint at the relay considered in (B.4b), the equation is reduced to

$$\beta_2 E_R \text{Tr} (\mathbf{D}\mathbf{H}_2\mathbf{H}_2^H\mathbf{D}^H) + \beta_2\alpha_2 E_R = \text{Tr} (\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\bar{\mathbf{y}}_{BS}\mathbf{d}^H). \quad (\text{B.7})$$

In the same way, the derivative of ζ respect to β_2 is set as

$$\begin{aligned} \frac{\partial \zeta(\dots)}{\partial \beta_2} &= 2\beta_2 \text{Tr} (\mathbf{D}\mathbf{H}_2\mathbf{W} (\mathbf{H}_1\bar{\mathbf{y}}_{BS}\bar{\mathbf{y}}_{BS}^H\mathbf{H}_1^H + \beta_1^2 N_{01}\mathbf{I}_R) \mathbf{W}^H\mathbf{H}_2^H\mathbf{D}^H) + 2\beta_2 N_{02} \text{Tr} (\mathbf{D}\mathbf{D}^H) \\ &- 2\beta_2 \text{Tr} (\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\bar{\mathbf{y}}_{BS}\mathbf{d}^H) = 0, \end{aligned}$$

where after the application of the constraint (B.4b) we get

$$\beta_2 E_R \text{Tr} (\mathbf{D}\mathbf{H}_2\mathbf{H}_2^H\mathbf{D}^H) + \beta_2 N_{02} \text{Tr} (\mathbf{D}\mathbf{D}^H) = \text{Tr} (\mathbf{D}\mathbf{H}_2\mathbf{W}\mathbf{H}_1\bar{\mathbf{y}}_{BS}\mathbf{d}^H). \quad (\text{B.8})$$

In order to get the Lagrangian multiplier α_2 , (B.8) is made equal to (B.7), getting

$$\alpha_2 = \frac{N_{02}}{E_R} \text{Tr} (\mathbf{D}\mathbf{D}^H).$$

Hence, the relaying matrix is defined as

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H\mathbf{D}^H\mathbf{D}\mathbf{H}_2 + \frac{N_{02} \text{Tr} (\mathbf{D}\mathbf{D}^H)}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H\mathbf{D}^H\mathbf{R}_d\bar{\mathbf{F}}^H\mathbf{H}_1^H\bar{\boldsymbol{\Xi}}^{-1},$$

where $\bar{\boldsymbol{\Xi}} = \mathbf{H}_1\bar{\mathbf{F}}\mathbf{R}_d\bar{\mathbf{F}}^H\mathbf{H}_1^H + \beta_1^2 N_{01}\mathbf{I}_R$ and $\mathbf{R}_d = \text{E} [\mathbf{d}\mathbf{d}^H]$. In the expression, β_1^2 can be replaced by $\beta_1^2 = \frac{\text{Tr} (\bar{\mathbf{F}}\mathbf{R}_d\bar{\mathbf{F}}^H)}{E_{BS}}$. The precoding factor β_2 is defined as

$$\beta_2 = \sqrt{\frac{\text{Tr} \left(\mathbf{R}_d\bar{\mathbf{F}}^H\mathbf{H}_1^H\bar{\boldsymbol{\Xi}}^{-1}\mathbf{H}_1\bar{\mathbf{F}}\mathbf{R}_d\mathbf{D}\mathbf{H}_2 \left(\mathbf{H}_2^H\mathbf{D}^H\mathbf{D}\mathbf{H}_2 + \frac{N_{02} \text{Tr} (\mathbf{D}\mathbf{D}^H)}{E_R} \mathbf{I}_R \right)^{-2} \mathbf{H}_2^H\mathbf{D}^H \right)}{E_R}}.$$

In the same fashion, in order to obtain the expressions for \mathbf{F} and β_1 , the derivative of the Lagrangian ζ are set respect to this variables, considering known \mathbf{W} and β_2 .

$$\begin{aligned} \frac{\partial \zeta(\dots)}{\partial \mathbf{y}_{BS}} &= \beta_1^2 \mathbf{H}_1^H\bar{\mathbf{W}}^H\mathbf{H}_2^H\mathbf{D}^H\mathbf{D}\mathbf{H}_2\bar{\mathbf{W}}\mathbf{H}_1\mathbf{y}_{BS} - \beta_1 \mathbf{H}_1^H\bar{\mathbf{W}}^H\mathbf{H}_2^H\mathbf{D}^H\mathbf{d} - \gamma_1 \mathbf{y}_{BS} \\ &+ \alpha_2 \beta_1^2 \mathbf{H}_1^H\bar{\mathbf{W}}^H\bar{\mathbf{W}}\mathbf{H}_1\mathbf{y}_{BS} = \mathbf{0}_{M \times L}, \end{aligned}$$

being $\bar{\mathbf{W}} = \beta_2 \mathbf{W}$ the unscaled version of the precoding matrix. If $\alpha_1 = -\gamma_1/\beta_1^2$,

$$\beta_1 \left(\mathbf{H}_1^H\bar{\mathbf{W}}^H (\mathbf{H}_2^H\mathbf{D}^H\mathbf{D}\mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \bar{\mathbf{W}}\mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right) \mathbf{y}_{BS} = \mathbf{H}_1^H\bar{\mathbf{W}}^H\mathbf{H}_2^H\mathbf{D}^H\mathbf{d}. \quad (\text{B.9})$$

Multiplying (B.9) by \mathbf{y}_{BS}^H and applying the trace operation we get

$$\begin{aligned} & \beta_1 \left(\text{Tr} \left((\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{y}_{BS} \mathbf{y}_{BS}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \right) + \alpha_1 \text{Tr} (\mathbf{y}_{BS} \mathbf{y}_{BS}^H) \right) \\ & = \text{Tr} (\mathbf{d}^H \mathbf{D} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{y}_{BS}), \end{aligned} \quad (\text{B.10})$$

where the trace property $\text{Tr}(\mathbf{A}\mathbf{B}) = \text{Tr}(\mathbf{B}\mathbf{A})$ has been used and the Hermitian application has been employed. From the power constraint at the BS defined in (B.4a), (B.10) stands for

$$\beta_1 \left(E_{BS} \text{Tr} \left((\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \right) + \alpha_1 E_{BS} \right) = \text{Tr} (\mathbf{D} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{y}_{BS} \mathbf{d}^H). \quad (\text{B.11})$$

Similarly β_1 is derived from the Lagrangian function ζ getting

$$\frac{\partial \zeta(\dots)}{\partial \beta_1} = 2\beta_1 E_{BS} \text{Tr} (\mathbf{H}_1 \mathbf{A} \mathbf{H}_1^H) + 2\beta_1 N_{01} \text{Tr} (\mathbf{A}) - 2\text{Tr} (\mathbf{D} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{y}_{BS} \mathbf{d}^H) = 0, \quad (\text{B.12})$$

where $\mathbf{A} = \overline{\mathbf{W}}^H (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \overline{\mathbf{W}}$. The constraint (B.4a) has been introduced into the equation. Finally after some manipulations, making equal the expressions (B.11) and (B.12), the Lagrangian multiplier α_1 denotes

$$\alpha_1 = \frac{N_{01}}{E_{BS}} \text{Tr} \left(\overline{\mathbf{W}}^H (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \overline{\mathbf{W}} \right),$$

getting the following expressions for the precoding matrix

$$\mathbf{y}_{BS} = \frac{1}{\beta_1} \underbrace{\left(\mathbf{H}_1^H \overline{\mathbf{W}}^H (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \overline{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{d}}_{\mathbf{F}},$$

and the scaling factor at the BS

$$\beta_1 = \sqrt{\frac{\mathbf{d}^H \mathbf{D} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H (\mathbf{H}_2^H \mathbf{D}^H \mathbf{D} \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \overline{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{D}^H \mathbf{d}}{E_{BS}}}$$

Derivation of Power Allocation Matrices for Sum-MSE Minimization

The Lagrangian function with respect to $x_{k,i}$ can be written as:

$$\zeta_1 = \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} x_{k,i} + N_{01})}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})} + \mu_1 \left(\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} - E_{BS} \right) - \sum_{k=1}^K \sum_{i=1}^{L_k} \nu_{k,i} x_{k,i},$$

where μ_1 is chosen to satisfy the power constraint in (4.10). It should be pointed out that $\mu_{k,i} \geq 0$, $\nu_{k,i} \geq 0$ for all $i = 1, \dots, L_k$ and $k = 1, \dots, K$. As a result of the optimization problem, the solution can be obtained by means of KKT conditions defined as

$$\frac{\partial \zeta_1}{\partial x_{k,i}} = - \frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})^2} + \mu_1 - \nu_{k,i} = 0, \quad (\text{C.1a})$$

$$\mu_1, \nu_{k,i}, x_{k,i} \geq 0,$$

$$\nu_{k,i} x_{k,i} = 0, \quad (\text{C.1b})$$

$$\mu_1 \left(\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} - E_{BS} \right) = 0.$$

From (C.1a) we know that

$$- \frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})^2} + \mu_1 = \nu_{k,i}. \quad (\text{C.2})$$

Substituting (C.2) in (C.1b)

$$\left(- \frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})^2} + \mu_1 \right) x_{k,i} = 0. \quad (\text{C.3})$$

Taking into consideration that $x_{k,i} \geq 0$, from (C.3) we derive that

$$\sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(b_{k,i} y_{k,i} + N_{02}) (a_{k,i} x_{k,i} + N_{01})^2} = \mu_1. \quad (\text{C.4})$$

From (C.4)

$$x_{k,i} = \frac{1}{a_{k,i}} \left[\sqrt{\frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{\mu_1 (b_{k,i} y_{k,i} + N_{02})}} - N_{01} \right]^+,$$

where $[p]^+ = \max(p, 0)$ or in other words, if the result obtained is negative the power allocation will be equal to zero. Due to the structure of $x_{k,i}$ and taking into account that

$\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} - E_{BS} = 0$, after some manipulations the water level is given by

$$\mu_1 = \left[\frac{\sum_{k=1}^K \sum_{i=1}^{L_k} \frac{1}{a_{k,i}} \sqrt{\frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(b_{k,i} y_{k,i} + N_{02})}}}{E_{BS} + \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{01}}{a_{k,i}}} \right]^2.$$

Derivation of Power Allocation Matrices for Sum-Rate Maximization

The Lagrangian function for $x_{k,i}$ optimization is defined as

$$\begin{aligned} \zeta_1 &= -\frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{L_k} \log_2 \left(1 + \frac{a_{k,i} b_{k,i} x_{k,i} y_{k,i}}{N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} x_{k,i} + N_{01})} \right) + \mu_1 \left(\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} - E_{BS} \right) \\ &\quad - \sum_{k=1}^K \sum_{i=1}^{L_k} \nu_{k,i} x_{k,i}, \end{aligned}$$

where μ_1 is the water-level to satisfy the power constraint. The conditions for obtaining the solution are

$$-\frac{1}{2\ln(2)} \frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} b_{k,i} + N_{01})) (a_{k,i} x_{k,i} + N_{01})} + \mu_1 - \nu_{k,i} = 0, \quad (\text{D.1a})$$

$$\mu_1, \nu_{k,i}, x_{k,i} \geq 0,$$

$$\nu_{k,i} x_{k,i} = 0, \quad (\text{D.1b})$$

$$\mu_1 \left(\sum_{k=1}^K \sum_{i=1}^{L_k} x_{k,i} - E_{BS} \right) = 0, \quad (\text{D.1c})$$

From (D.1a) we know that

$$-\frac{1}{2\ln(2)} \frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} b_{k,i} + N_{01})) (a_{k,i} x_{k,i} + N_{01})} + \mu_1 = \nu_{k,i}. \quad (\text{D.2})$$

Substituting (D.2) in (D.1b) we get

$$\left(-\frac{1}{2\ln(2)} \frac{N_{01} a_{k,i} b_{k,i} y_{k,i}}{(N_{01} b_{k,i} y_{k,i} + N_{02} (a_{k,i} b_{k,i} + N_{01})) (a_{k,i} x_{k,i} + N_{01})} + \mu_1 \right) x_{k,i} = 0.$$

Knowing that $x_{k,i} \geq 0$

$$\left(-\frac{1}{2\ln(2)} \frac{N_{0_1} a_{k,i} b_{k,i} y_{k,i}}{(N_{0_1} b_{k,i} y_{k,i} + N_{0_2} (a_{k,i} b_{k,i} + N_{0_1})) (a_{k,i} x_{k,i} + N_{0_1})} + \mu_1 \right) = 0.$$

Hence,

$$x_{k,i} = \frac{N_{0_1}}{2a_{k,i}} \left[\sqrt{\frac{b_{k,i}^2 y_{k,i}^2}{N_{0_2}^2} + \frac{2b_{k,i} a_{k,i} y_{k,i} \mu_1^*}{\ln(2) N_{0_2} N_{0_1}}} - \left(\frac{b_{k,i} y_{k,i}}{N_{0_2}} + 2 \right) \right]^+,$$

where $\mu_1^* = 1/\mu_1$ is the water-level and $[p]^+ = \max(p, 0)$ which means that if p is negative the value that will take p is zero. After some manipulations, $x_{k,i}$ can be rewritten as:

$$x_{k,i} = \left[\sqrt{m_{k,i} \left(\mu_1^* + \frac{n_{k,i}}{m_{k,i}} \right)} - z_{k,i} \right], \quad (\text{D.3})$$

being (D.3) the definition of the classical water-filling solution with the definition of the next scalars:

$$\begin{aligned} m_{k,i} &= \frac{N_{0_1} b_{k,i} y_{k,i}}{2\ln(2) N_{0_2} a_{k,i}}, \\ z_{k,i} &= \frac{N_{0_1} b_{k,i} y_{k,i}}{2N_{0_2}} + \frac{N_{0_1}}{a_{k,i}}, \\ n_{k,i} &= \frac{N_{0_1}^2 b_{k,i}^2 y_{k,i}^2}{4N_{0_2} a_{k,i}^2}. \end{aligned}$$

The term μ_1^* , the water-level of the first hop, can be derived from (D.1c), taking into consideration that $\mu_1 \geq 0$ it is known that

$$\begin{aligned} \sum_{k=1}^K \sum_{i=1}^{L_k} (x_{k,i} - E_{BS}) &= 0 \\ \sum_{k=1}^K \sum_{i=1}^{L_k} \frac{N_{0_1}}{2a_{k,i}} \left[\sqrt{\frac{b_{k,i}^2 y_{k,i}^2}{N_{0_2}^2} + \frac{2b_{k,i} a_{k,i} y_{k,i} \mu_1^*}{\ln(2) N_{0_2} N_{0_1}}} - \left(\frac{b_{k,i} y_{k,i}}{N_{0_2}} + 2 \right) \right] - E_{BS} &= 0. \end{aligned} \quad (\text{D.4})$$

The non-linear equation (D.4) is solved by means of the bisection method described in Algorithm 16. This finds a root in an interval where the maximum and minimum starting values have to be set previously. In order to get the initial maximum value of μ_M denoted $\mu_{M,0}$ we define

$$D_{k,i} = \min \left(\frac{n_{k,i}}{m_{k,i}} \right), \quad (\text{D.5})$$

for all $k = 1, \dots, K$ and $i = 1, \dots, L_k$. Substituting (D.5) into (D.3) we get the upper bound of $\mu_{M,0}$

$$\mu_{M,0}^* = \left[\frac{E_{BS} + \sum_{k=1}^K \sum_{i=1}^{L_k} z_{k,i}}{\sum_{k=1}^K \sum_{i=1}^{L_k} \sqrt{m_{k,i}}} \right]^2 - D_{k,i}.$$

The minimum value of $\mu_{L,0}^*$ will be the one that fulfils with

$$\sqrt{m_{k,i} \left(m_{k,i} \mu_{L,0}^* + \frac{n_{k,i}}{m_{k,i}} \right)} \geq z_{k,i},$$

where after some manipulations the lowest value of the water-level is defined as

$$\mu_{L,0}^* = \min \left(z_{k,i}^2 - \frac{n_{k,i}}{m_{k,i}} \right) \quad \text{for all } k = 1, \dots, K \text{ and } i = 1, \dots, L_k.$$

The next lines define the bisection method for the search of the water-levels, where $\mu_{M,0}$ and $\mu_{L,0}$ are the initial maximum and minimum water-levels, respectively.

Algorithm 16 Bisection method for the search of the water-level factor.

- 1: Initialization: $\mu_M = \mu_{M,0}$, $\mu_L = \mu_{L,0}$, $\epsilon_{\min} = 0.000001$ and $\epsilon = \infty$.
 - 2: **while** $\epsilon \geq \epsilon_{\min}$ **do**
 - 3: $\mu = \frac{\mu_M + \mu_L}{2}$.
 - 4: Computation of $f(a) = \sum_{k=1}^K \sum_{i=1}^{L_k} \left[\sqrt{m_{k,i} \left(\mu^* + \frac{n_{k,i}}{m_{k,i}} \right)} - z_{k,i} \right]$, where $\mu^* = 1/\mu$.
 - 5: Calculate $f(b) = \sum_{k=1}^K \sum_{i=1}^{L_k} \left[\sqrt{m_{k,i} \left(\mu_M^* + \frac{n_{k,i}}{m_{k,i}} \right)} - z_{k,i} \right]$, where $\mu_M^* = 1/\mu_M$.
 - 6: **if** $f(a)f(b) < 0$ **then**
 - 7: $\mu_L = \mu$.
 - 8: **else**
 - 9: $\mu_M = \mu$.
 - 10: **end if**
 - 11: $\epsilon = |\mu_M - \mu_L|$.
 - 12: **end while**
-

Suboptimal THP Approach Analysis

Throughout this appendix the suboptimal Wiener THP solution analysed in Section 6.1 is obtained. Firstly, to obtain the relaying matrix and the scaling factor β_2 the Lagrangian function is defined as:

$$\zeta_2 = \xi_2 + \lambda_2 (\mathbf{W}\mathbf{R}_{\mathbf{y}_1}\mathbf{W}^H - E_R),$$

being $\xi_2 = \mathbb{E} \left[\|\hat{\mathbf{d}} - \mathbf{y}_1\|_2^2 \right]$ the mean square error of the second hop depicted as

$$\xi_2 = \text{Tr} \left(\beta_2^2 \mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H \mathbf{H}_2^H + \beta_2^2 \mathbf{I}_K - 2\beta_2 \Re(\mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1}) + \mathbf{R}_{\mathbf{y}_1} \right).$$

Setting Lagrangian's derivative respect to \mathbf{W} , we obtain

$$\frac{\partial \zeta}{\partial \mathbf{W}} = \beta_2^2 \mathbf{H}_2^H \mathbf{H}_2 \mathbf{W} - \beta_2 \mathbf{H}_2^H + \alpha_2 \mathbf{W} = \mathbf{0}_K.$$

After some manipulations, defining $\alpha_2 = \lambda_2 / \beta_2^2$

$$\beta_2 (\mathbf{H}_2^H \mathbf{H}_2 + \alpha_2 \mathbf{I}_R) \mathbf{W} = \mathbf{H}_2^H. \quad (\text{E.1})$$

In the same fashion, setting the derivative of the Lagrangian respect to β_2 , we get

$$\frac{\partial \zeta}{\partial \beta_2} = 2\beta_2 \text{Tr}(\mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H \mathbf{H}_2^H) + \frac{2\beta_2 N_{02} K}{E_R} - 2\text{Tr}(\mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1}) = 0. \quad (\text{E.2})$$

From (E.1), multiplying it by $\mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H$ and applying the trace, we get

$$\beta_2 \text{Tr}(\mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H \mathbf{H}_2) + \beta_2 \alpha_2 \text{Tr}(\mathbf{W} \mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H) = \text{Tr}(\mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1}). \quad (\text{E.3})$$

Since the power constraint stands for $\text{Tr}(\mathbf{W} \mathbf{R}_{\mathbf{y}_1} \mathbf{W}^H) = E_R$, (E.3) is reduced to

$$\beta_2 E_R \text{Tr}(\mathbf{H}_2 \mathbf{H}_2^H) + \beta_2 E_R \alpha_2 = \text{Tr}(\mathbf{H}_2 \mathbf{W} \mathbf{R}_{\mathbf{y}_1}). \quad (\text{E.4})$$

Making (E.4) equal to (E.2), the Lagrangian multiplier takes the value

$$\alpha_2 = \frac{N_{0_2}K}{E_R}.$$

Finally, after some manipulations the relaying matrix and the scaling factor of the second hop are

$$\mathbf{W} = \frac{1}{\beta_2} \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2}K}{E_R} \mathbf{I}_R \right)^{-1} \mathbf{H}_2^H$$

$$\beta_2 = \sqrt{\frac{\text{Tr} \left(\left(\mathbf{H}_1 \bar{\mathbf{F}} \mathbf{R}_{\mathbf{v},\mathbf{v}} \bar{\mathbf{F}}^H \mathbf{H}_1^H + \beta_1^2 N_{0_1} \mathbf{I}_R \right) \mathbf{H}_2 \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{0_2}K}{E_R} \mathbf{I}_R \right)^{-2} \mathbf{H}_2^H \right)}{E_R}},$$

where $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$ is the unscaled version of the precoding matrix \mathbf{F} .

After obtaining the relaying matrix and the scaling factor β_2 , the feedback filter (\mathbf{B}), the feedforward filter (\mathbf{F}) and the scaling factor β_1 are obtained jointly for the minimization of the error from the base station to the end users and to limit the power at the output of the BS.

From the optimization problem defined in (6.1), the Lagrangian function for matrix derivation is defined as

$$\begin{aligned} \zeta_1 &= \text{Tr} \left(\beta_1^2 \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{R}_{\mathbf{v}} \mathbf{F}^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H + \beta_1^2 N_{0_1} \mathbf{H}_2 \bar{\mathbf{W}} \bar{\mathbf{W}}^H \mathbf{H}_2^H + \beta_2^2 N_{0_2} \mathbf{I}_K \right. \\ &\quad \left. - 2\beta_1 \Re \left(\mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{R}_{\mathbf{v}} (\mathbf{I}_K - \mathbf{B}) \right) + (\mathbf{I}_K - \mathbf{B}) \mathbf{R}_{\mathbf{v}} (\mathbf{I}_K - \mathbf{B})^H \right) + \lambda_1 (\text{Tr} (\mathbf{F} \mathbf{R}_{\mathbf{v}} \mathbf{F}) \\ &\quad - E_{BS}) + \sum_{k=1}^K 2\Re (\mu_k^* \mathbf{S}_k \mathbf{B} \mathbf{e}_k), \end{aligned} \quad (\text{E.5})$$

being λ_1 and μ_k the Lagrangian multipliers. It should be pointed out that $\bar{\mathbf{W}} = \beta_2 \mathbf{W}$ has been introduced in (E.5). After setting the derivative of the Lagrangian respect to \mathbf{F} we obtain

$$\frac{\partial \zeta_1}{\partial \mathbf{F}} = \beta_1^2 \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \mathbf{F} - \beta_1 \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H (\mathbf{I}_K \mathbf{B}) + \lambda_1 \mathbf{F} = \mathbf{0}_{M \times K}.$$

After some manipulations we get

$$\beta_1 \left(\mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right) \mathbf{F} = \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H (\mathbf{I}_K - \mathbf{B}), \quad (\text{E.6})$$

where $\alpha_1 = \lambda_1 / \beta_1^2$. In the same way, in order to obtain β_1 's value the derivative of the

Lagrangian is set respect to β_1 such as

$$\begin{aligned} \frac{\partial \zeta_1}{\partial \beta_1} &= 2\beta_1 \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{R}_v \mathbf{F}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \right) + 2\beta_1 N_{01} \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} \overline{\mathbf{W}}^H \mathbf{H}_2^H \right) \\ &- 2\text{Tr} \left((\mathbf{I}_K - \mathbf{B}) \mathbf{R}_v \mathbf{F}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \right) = 0. \end{aligned}$$

After manipulating the last equation and knowing that the power limitation is $\text{Tr}(\mathbf{F} \mathbf{R}_v \mathbf{F}^H)$, we obtain

$$\beta_1 \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} (E_{BS} \mathbf{H}_1 \mathbf{H}_1^H + N_{01} \mathbf{I}_R) \overline{\mathbf{W}}^H \mathbf{H}_2^H \right) = \text{Tr} \left((\mathbf{I}_K - \mathbf{B}) \mathbf{R}_v \mathbf{F}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \right). \quad (\text{E.7})$$

From (E.7), multiplying it by $\mathbf{R}_v \mathbf{F}^H$ and applying the trace we obtain

$$\beta_1 \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{R}_v \mathbf{F}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \right) + \beta_1 \alpha_1 \text{Tr}(\mathbf{F} \mathbf{R}_v \mathbf{F}^H). \quad (\text{E.8})$$

Knowing that the power limit at the BS is $\text{Tr}(\mathbf{F} \mathbf{R}_v) = E_{BS}$, (E.8) can be reduced to

$$\beta_1 E_{BS} \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \right) + \beta_1 E_{BS} \alpha_1 = \text{Tr} \left((\mathbf{I}_K - \mathbf{B}) \mathbf{R}_v \mathbf{F}^H \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \right). \quad (\text{E.9})$$

Making (E.9) equal to (E.7), the value of α_1 is obtained such as

$$\alpha_1 = \frac{N_{01} \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} \overline{\mathbf{W}}^H \mathbf{H}_2^H \right)}{E_{BS}} \quad (\text{E.10})$$

Finally, the feedback filter expression is derived setting the derivative of ζ_1 respect to \mathbf{B} such as

$$\frac{\partial \zeta_1}{\partial \mathbf{B}} = \beta_1 \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{R}_v - (\mathbf{I}_K - \mathbf{B}) \mathbf{R}_{v,v} - \sum_{k=1}^K \mathbf{S}_k^T \mu_k^* \mathbf{e}_k^T = \mathbf{0}_K, \quad (\text{E.11})$$

where $\mathbf{S}_k = \mathbf{S}_{(0,k,K-k)} = [\mathbf{I}_k, \mathbf{0}_{k \times K-k}] \in \{0, 1\}^{k \times K}$ denotes a selection matrix which cuts out the first k elements of a K -dimensional vector and $\mathbf{e}_k \in \{0, 1\}^K$ stands for the k^{th} column of the identity matrix \mathbf{e} .

Including into (E.11) feedforward filter's expression defined as

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 + \alpha_1 \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H,$$

and due to the matrix inversion lemma, after some manipulations we get

$$\mathbf{B} = \mathbf{I}_K + \alpha_1^{-1} \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H + \alpha_1 \mathbf{I}_K \right) \sum_{k=1}^K \sigma_{v_k}^{-2} \mathbf{S}_k^T \mu_k^* \mathbf{e}_k^T. \quad (\text{E.12})$$

From the triangularization constraint $\mathbf{S}_k \mathbf{B} \mathbf{e}_k = \mathbf{0}$ we obtain

$$\mu_k^* = -\alpha_1 \sigma_{\mathbf{v}_k}^2 \left(\mathbf{S}_k \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H + \alpha_1 \mathbf{I}_K \right) \mathbf{S}_k^T \right)^{-1} \mathbf{S}_k \mathbf{e}_k. \quad (\text{E.13})$$

Moreover, we know that

$$\begin{aligned} \mathbf{I}_K &= \sum_{k=1}^K \mathbf{S}_k^T \mathbf{S}_k \mathbf{e}_k \mathbf{e}_k^T \\ &= \sum_{k=1}^K -\alpha_1^{-1} \sigma_{\mathbf{v}_k}^{-2} \mathbf{S}_k^T \mathbf{S}_k \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H + \alpha_1 \mathbf{I}_k \right) \mathbf{S}_k^T \mu_k^* \mathbf{e}_k^T. \end{aligned} \quad (\text{E.14})$$

Including (E.14) into (E.12) and considering $\mathbf{S}_k^T \mathbf{S}_k \mathbf{S}_k^T - \mathbf{S}_k^T = \mathbf{0}_{K \times k}$ we get

$$\mathbf{B} = \sum_{k=1}^K \alpha_1^{-1} \sigma_{\mathbf{v}_k}^{-2} \left(\mathbf{I}_K - \mathbf{S}_k^T \mathbf{S}_k \right) \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{S}_k^T \mu_k^* \mathbf{e}_k^T.$$

After some manipulations

$$\mu_k^* = -\alpha_1 \sigma_{\mathbf{v}_k}^2 \mathbf{S}_k \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H + \alpha_1 \mathbf{I}_K \right)^{-1} \mathbf{e}_k$$

and knowing that

$$\mathbf{S}_k^T \mathbf{S}_k = \mathbf{I}_K - \sum_{j=k+1}^K \mathbf{e}_j \mathbf{e}_j^T \in \{0, 1\}^{K \times k},$$

the final feedforward filter, feedback filter and scaling factor expression stand for

$$\mathbf{F} = \frac{1}{\beta_1} \left(\mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 + \frac{N_{01}}{E_{BS}} \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{W}^H \mathbf{H}_2^H \right) \mathbf{I}_M \right)^{-1} \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{e}_k \mathbf{e}_k^T,$$

$$\mathbf{B} = \beta_1 \sum_{k=1}^K \left(\mathbf{S}_k^T \mathbf{S}_k - \mathbf{I}_K \right) \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{F} \mathbf{e}_k \mathbf{e}_k^T,$$

$$\beta_1 = \sqrt{\frac{\sum_{k=1}^K \sigma_{\mathbf{v}_k}^2 \mathbf{e}_k^T \mathbf{A}^{-2} \mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{e}_k}{E_{BS}}},$$

being $\mathbf{A} = \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{H}_1 \mathbf{H}_1^H \overline{\mathbf{W}}^H \mathbf{H}_2^H + \frac{\sigma_1^2}{E_{BS}} \text{Tr} \left(\mathbf{H}_2 \overline{\mathbf{W}} \mathbf{W}^H \mathbf{H}_2^H \right) \mathbf{I}_M \right)$.

Decomposition Based Approach Analysis

Since the mean square error is defined as

$$\begin{aligned}\xi &= \text{E} \left[\|\hat{\mathbf{d}} - \mathbf{d}\|_2^2 \right] \\ &= \mathbf{d}^H \mathbf{d} - 2\Re \left(\beta_2 \beta_1 \mathbf{d}^H \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{d} \right) \\ &\quad + \beta_2^2 \beta_1^2 \mathbf{d}^H \mathbf{F}^H \mathbf{H}_1^H \mathbf{W}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{W} \mathbf{H}_1 \mathbf{F} \mathbf{d} + \beta_2^2 \beta_1^2 N_{01} \text{Tr} \left(\mathbf{W}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{W} \right) + \beta_2^2 K N_{02}.\end{aligned}$$

Assuming that the unscaled matrix versions are $\bar{\mathbf{F}} = \beta_1 \mathbf{F}$ and $\bar{\mathbf{W}} = \beta_2 \mathbf{W}$, the MSE is reduced to

$$\begin{aligned}\xi &= \mathbf{d}^H \mathbf{d} - 2\Re \left(\mathbf{d}^H \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \bar{\mathbf{F}} \mathbf{d} \right) + \mathbf{d}^H \bar{\mathbf{F}}^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \bar{\mathbf{F}} \mathbf{d} + \beta_1^2 N_{01} \text{Tr} \left(\bar{\mathbf{W}}^H \mathbf{H}_2^H \mathbf{H}_2 \bar{\mathbf{W}} \right) \\ &\quad + \beta_2^2 K N_{02}.\end{aligned}\tag{F.1}$$

From the power constraint at the BS

$$\beta_1^2 = \frac{\mathbf{d}^H \bar{\mathbf{F}}^H \bar{\mathbf{F}} \mathbf{d}}{E_{BS}},$$

and the power limit at the RS terminal

$$\beta_2^2 = \frac{\mathbf{d} \bar{\mathbf{F}}^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \bar{\mathbf{W}} \mathbf{H}_1 \bar{\mathbf{F}} \mathbf{d}}{E_R} + \frac{N_{01} \text{Tr} \left(\bar{\mathbf{W}} \bar{\mathbf{W}}^H \right)}{E_R} \frac{\mathbf{d}^H \bar{\mathbf{F}}^H \bar{\mathbf{F}} \mathbf{d}}{E_{BS}}.$$

Applying the power constraints into the MSE expression defined in (F.1) and after some manipulations we get

$$\begin{aligned}\xi &= \mathbf{d}^H \left(\mathbf{I}_K - 2\Re \left(\mathbf{H}_2 \bar{\mathbf{W}} \mathbf{H}_1 \bar{\mathbf{F}} \right) + \bar{\mathbf{F}}^H \mathbf{H}_1^H \bar{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02} K}{E_R} \mathbf{I}_R \right) \bar{\mathbf{W}} \mathbf{H}_1 \bar{\mathbf{F}} \right. \\ &\quad \left. + \frac{N_{01}}{E_{BS}} \text{Tr} \left(\bar{\mathbf{W}}^H \left(\mathbf{H}_2^H \mathbf{H}_2 + \frac{N_{02} K}{E_R} \mathbf{I}_R \right) \bar{\mathbf{W}} \right) \bar{\mathbf{F}}^H \bar{\mathbf{F}} \right) \mathbf{d}\end{aligned}$$

References

- [Agrell02] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, “Closest point search in lattices”, *IEEE Transactions on Information Theory*, vol. 48, no. 8, 2201–2214, 2002.
- [Amitay84] N. Amitay and J. Salz, “Linear equalization theory in digital data transmission over dually polarized fading radio channels”, *AT T Technical Journal*, vol. 63, 2215–2259, 1984.
- [Barbero08] L. Barbero and J. Thompson, “Fixing the complexity of the sphere decoder for MIMO detection”, *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, 2131–2142, 2008.
- [Barrenechea10] M. Barrenechea, M. Joham, M. Mendicute, and W. Utschick, “Analysis of vector precoding at high SNR: rate bounds and ergodic results”, in *IEEE Global Communications Conference (GLOBECOM 2010)*, pp. 1–5, 2010.
- [Barrenechea12] M. Barrenechea, *Design and implementation of multi-user MIMO precoding algorithms*, Ph.D. thesis, 2012.
- [Berger67] T. Berger and D. Tufts, “Optimum pulse amplitude modulation: Transmitter-receiver design and bounds from information theory”, *IEEE Transactions on Information Theory*, vol. 13, no. 2, 196–208, 1967.
- [Boyd04] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [Caire03] G. Caire and S. Shamai, “On the achievable throughput of a multi-antenna Gaussian broadcast channel”, *IEEE Transactions on Information Theory*, vol. 49, no. 7, 1691–1706, 2003.

- [Catreux02] S. Catreux, V. Erceg, D. Gesbert, and R. Heath Jr, “Adaptive modulation and MIMO coding for broadband wireless data networks”, *IEEE Communications Magazine*, vol. 40, no. 6, 108–115, 2002.
- [Chae08a] C. Chae, S. Shim, and R. Heath, “Block diagonalized vector perturbation for multiuser MIMO systems”, *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, 4051–4057, 2008.
- [Chae08b] C. Chae, T. Tang, R. Heath, and S. Cho, “MIMO relaying with linear processing for multiuser transmission in fixed relay networks”, *IEEE Transactions on Signal Processing*, vol. 56, no. 2, 727–738, 2008.
- [Costa83] M. Costa, “Writing on dirty paper”, *IEEE Transactions on Information Theory*, vol. 29, no. 3, 439–441, 1983.
- [Cover79] T. Cover and A. Gamal, “Capacity theorems for the relay channel”, *IEEE Transactions on Information Theory*, vol. 25, no. 5, 572–584, 1979.
- [Cover91] T. Cover, J. Thomas, J. Wiley, et al., *Elements of information theory*, vol. 6, Wiley Online Library, 1991.
- [Damen03] M. Damen, H. El Gamal, and G. Caire, “On maximum-likelihood detection and the search for the closest lattice point”, *IEEE Transactions on Information Theory*, vol. 49, no. 10, 2389–2402, 2003.
- [Dietrich07] F. Dietrich, *Robust Signal Processing for Wireless Communications No. 2*, Springer Publishing Company, Incorporated, 2007.
- [Fan07] Y. Fan and J. Thompson, “MIMO configurations for relay channels: Theory and practice”, *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, 1774–1786, 2007.
- [Fang06] Z. Fang, Y. Hua, and J. Koshy, “Joint source and relay optimization for a non-regenerative MIMO relay”, in *Fourth IEEE Workshop on Sensor Array and Multichannel Processing*, pp. 239–243, 2006.
- [Fischer02a] R. Fischer, C. Windpassinger, A. Lampe, and J. Huber, “MIMO precoding for decentralized receivers”, in *Proceedings IEEE International Symposium on Information Theory*, p. 496, IEEE, 2002.
- [Fischer02b] R. Fischer, C. Windpassinger, A. Lampe, and J. Huber, “Space-time transmission using Tomlinson-Harashima precoding”, *ITG FACHBERICHT*, pp. 139–148, 2002.

- [Foschini98] G. Foschini and M. Gans, “On limits of wireless communications in a fading environment when using multiple antennas”, *Wireless personal communications*, vol. 6, no. 3, 311–335, 1998.
- [Goldsmith03] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, “Capacity limits of MIMO channels”, *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, 684–702, 2003.
- [Golub96] G. Golub and C. Van Loan, *Matrix computations*, vol. 3, Johns Hopkins Univ Pr, 1996.
- [Guan08] W. Guan and H. Luo, “Joint MMSE transceiver design in non-regenerative MIMO relay systems”, *IEEE Communications Letters*, vol. 12, no. 7, 517–519, 2008.
- [Harashima72] H. Harashima and H. Miyakawa, “Matched-transmission technique for channels with intersymbol interference”, *IEEE Transactions on Communications*, vol. 20, no. 4, 774–780, 1972.
- [Héliot10] F. Héliot, R. Hoshyar, and R. Tafazolli, “Power allocation for the downlink of nonregenerative cooperative multi-user MIMO communication system”, in *Proceedings IEEE PIMRC*, 2010.
- [Hochwald05] B. Hochwald, C. Peel, and A. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication-Part II: Perturbation”, *IEEE Transactions on Communications*, vol. 53, no. 3, 537–544, 2005.
- [Hwang09] J. Hwang, K. Lee, H. Sung, and I. Lee, “Block diagonalization approach for amplify-and-forward relay systems in MIMO multi-user channels”, in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 3149–3153, 2009.
- [Jiang05a] Y. Jiang, W. Hager, and J. Li, “The geometric mean decomposition”, *Linear Algebra and Its Applications*, vol. 396, 373–384, 2005.
- [Jiang05b] Y. Jiang, J. Li, and W. Hager, “Joint transceiver design for MIMO communications using geometric mean decomposition”, *IEEE Transactions on Signal Processing*, vol. 53, no. 10, 3791–3803, 2005.
- [Jimenez11] I. Jimenez, S. Weiss, M. Mendicute, and E. Arruti, “Multiuser MIMO Amplify-And-Forward Relaying with Vector Precoding”, *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT*, 2011.

- [Jimenez12a] I. Jimenez, M. Barrenechea, M. Mendicute, and E. Arruti, "Iterative joint MMSE design of relaying MIMO downlink schemes with non-linearly precoded transmission", *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2012.
- [Jimenez12b] I. Jimenez, M. Barrenechea, M. Mendicute, and E. Arruti, "Non-linear precoding approaches for non-regenerative multiuser MIMO relay systems", *EURASIP European Signal Processing Conference (EUSIPCO)*, 2012.
- [Jimenez13a] I. Jimenez, M. Barrenechea, M. Mendicute, and E. Arruti, "Block diagonalization for joint MMSE transceiver design in multiuser MIMO AF relaying systems", 2013.
- [Jimenez13b] I. Jimenez, M. Barrenechea, M. Mendicute, and E. Arruti, "Block diagonalization with MMSE vector precoding for AF multiuser MIMO relaying schemes for geometric mean decomposition approach", 2013.
- [Jindal04] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels", *IEEE Transactions on Information Theory*, vol. 50, no. 5, 768–783, 2004.
- [Joham04] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding", *ITG FACHBERICHT*, pp. 387–394, 2004.
- [Joham05] M. Joham, W. Utschick, and J. Nosssek, "Linear transmit processing in MIMO communications systems", *IEEE Transactions on Signal Processing*, vol. 53, no. 8, 2700–2712, 2005.
- [Joham06] M. Joham and W. Utschick, "Ordered spatial Tomlinson Harashima precoding", *Smart Antennas-State-of-the-Art*, 2006.
- [Ju09] H. Ju, E. Oh, and D. Hong, "Improving efficiency of resource usage in two-hop full duplex relay systems based on resource sharing and interference cancellation", *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, 3933–3938, 2009.
- [Juntti97] M. Juntti, T. Schlosser, and J. Lilleberg, "Genetic algorithms for multiuser detection in synchronous CDMA", in *Proceedings IEEE International Symposium on Information Theory*, p. 492, 1997.

- [Kennedy95] J. Kennedy and R. Eberhart, "Particle swarm optimization", in *Proceedings IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995.
- [Khirallah06] C. Khirallah, P. Coulton, P. Rashvand, and N. Zein, "Multi-user MIMO CDMA systems using complete complementary sequences.", *IEE Proceedings-Communications*, vol. 153, no. 4, 535–540, 2006.
- [Kim09] H. Kim, S. Lee, K. Kwak, H. Min, and D. Hong, "On the design of ZF and MMSE Tomlinson-Harashima precoding in multiuser MIMO amplify-and-forward relay system", in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 2509–2513, 2009.
- [Kusume05] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Efficient Tomlinson-Harashima precoding for spatial multiplexing on flat MIMO channel", in *IEEE International Conference on Communications (ICC)*, vol. 3, pp. 2021–2025, 2005.
- [Laneman04] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior", *IEEE Transactions on Information Theory*, vol. 50, no. 12, 3062–3080, 2004.
- [Li09] G. Li, Y. Wang, T. Wu, and J. Huang, "Joint linear filter design in multi-user non-regenerative MIMO-relay systems", in *IEEE International Conference on Communications (ICC)*, pp. 1–6, 2009.
- [Lin08] S. Lin, W. Ho, and Y. Liang, "Block diagonal geometric mean decomposition (BD-GMD) for MIMO broadcast channels", *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, 2778–2789, 2008.
- [Liu07] F. Liu, L. Jiang, and C. He, "MMSE vector precoding with joint transmitter and receiver design for MIMO systems", *Signal Processing*, vol. 87, no. 11, 2823–2833, 2007.
- [Liu08] J. Liu and W. Krzymien, "A novel nonlinear joint transmitter-receiver processing algorithm for the downlink of multiuser MIMO systems", *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, 2189–2204, 2008.
- [Liu10] F. Liu, L. Jiang, and C. He, "Advanced joint transceiver design for block-diagonal geometric-mean-decomposition-based mul-

- tiuser MIMO systems”, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, 692–703, 2010.
- [Meurer04] M. Meurer, T. Weber, and W. Qiu, “Transmit nonlinear zero forcing: energy efficient receiver oriented transmission in MIMO CDMA mobile radio downlinks”, in *IEEE Eighth International Symposium on Spread Spectrum Techniques and Applications*, pp. 260–269, 2004.
- [Millar11] A. Millar, S. Weiss, and R. Stewart, “Tomlinson Harashima precoding design for non-regenerative MIMO relay networks”, in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pp. 1–5, 2011.
- [Mo09] R. Mo and Y. Chew, “MMSE-based joint source and relay precoding design for amplify-and-forward MIMO relay networks”, *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, 4668–4676, 2009.
- [Mohaisen11] M. Mohaisen and K. Chang, “Fixed-complexity sphere encoder for multi-user MIMO systems”, *Journal of Communications and Networks*, vol. 13, no. 1, 63–69, 2011.
- [Moshavi96] S. Moshavi, “Multi-user detection for DS-CDMA communications”, *IEEE Communications Magazine*, vol. 34, no. 10, 124–136, 1996.
- [Paulraj03] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*, Cambridge University Press, 2003.
- [Peel05] C. Peel, B. Hochwald, and A. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication—part I: channel inversion and regularization”, *IEEE Transactions on Communications*, vol. 53, no. 1, 195–202, 2005.
- [Prasad96] R. Prasad and S. Hara, “An overview of multi-carrier CDMA”, in *IEEE 4th International Symposium on Spread Spectrum Techniques and Applications Proceedings*, vol. 1, pp. 107–114, 1996.
- [Proakis87] J. Proakis, *Digital communications*, vol. 1221, McGraw-hill, 1987.
- [Rong11] Y. Rong, “Simplified algorithms for optimizing multiuser multi-hop MIMO relay systems”, *IEEE Transactions on Communications*, , no. 99, 1–9, 2011.
- [Sankaranarayanan04] L. Sankaranarayanan, G. Kramer, and N. Mandayam, “Capacity theorems for the multiple-access relay channel”, in *Proc. 42nd Annu.*

- Allerton Conf. Communications, Control, and Computing*, pp. 1782–1791, 2004.
- [Schmidt05] D. Schmidt, M. Joham, and W. Utschick, “Minimum mean square error vector precoding”, in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. 107–111, 2005.
- [Shannon48] C. Shannon, “A mathematical theory of communications, I and II”, *Bell Syst. Tech. Journal*, vol. 27, 379–423, 1948.
- [Shao05] X. Shao, J. Yuan, and P. Rapajic, “Precoder design for MIMO broadcast channels”, in *IEEE International Conference on Communications (ICC)*, vol. 2, pp. 788–794, 2005.
- [Shen06] Z. Shen, R. Chen, J. Andrews, R. Heath, and B. Evans, “Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization”, *IEEE Transactions on Signal Processing*, vol. 54, no. 9, 3658–3663, 2006.
- [Shen07] Z. Shen, R. Chen, J. Andrews, R. Heath, and B. Evans, “Sum capacity of multiuser MIMO broadcast channels with block diagonalization”, *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, 2040–2045, 2007.
- [Shim08] S. Shim, J. Kwak, R. Heath, and J. Andrews, “Block diagonalization for multi-user MIMO with other-cell interference”, *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, 2671–2681, 2008.
- [Spencer04a] Q. Spencer, C. Peel, A. Swindlehurst, and M. Haardt, “An introduction to the multi-user MIMO downlink”, *IEEE Communications Magazine*, vol. 42, no. 10, 60–67, 2004.
- [Spencer04b] Q. Spencer, A. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels”, *IEEE Transactions on Signal Processing*, vol. 52, no. 2, 461–471, 2004.
- [Tang06] T. Tang, C. Chae, R. Heath, and S. Cho, “On achievable sum rates of a multiuser MIMO relay channel”, in *IEEE International Symposium on Information Theory*, pp. 1026–1030, 2006.
- [Tang07] X. Tang and Y. Hua, “Optimal design of non-regenerative MIMO wireless relays”, *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, 1398–1407, 2007.

- [Telatar99] E. Telatar, “Capacity of multi-antenna gaussian channels”, *European transactions on telecommunications*, vol. 10, no. 6, 585–595, 1999.
- [Tomlinson71] M. Tomlinson, “New automatic equaliser employing modulo arithmetic”, *Electronics letters*, vol. 7, no. 5, 138–139, 1971.
- [Tseng09] F. Tseng and W. Wu, “Joint source/relay precoders design in amplify-and-forward relay systems: a geometric mean decomposition approach”, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2641–2644, IEEE, 2009.
- [Tseng10] F. Tseng and W. Wu, “Linear MMSE transceiver design in amplify-and-forward MIMO relay systems”, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, 754–765, 2010.
- [Tseng11] F. Tseng, M. Chang, and W. Wu, “Joint Tomlinson–Harashima source and linear relay precoder design in amplify-and-forward MIMO relay systems via MMSE criterion”, *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, 1687–1698, 2011.
- [Van Der Meulen71] E. Van Der Meulen, “Three-terminal communication channels”, *Advances in applied Probability*, pp. 120–154, 1971.
- [Verdú89] S. Verdú, “Computational complexity of optimum multiuser detection”, *Algorithmica*, vol. 4, no. 1, 303–312, 1989.
- [Vishwanath02] S. Vishwanath, N. Jindal, and A. Goldsmith, “On the capacity of multiple input multiple output broadcast channels”, in *IEEE International Conference on Communications (ICC)*, vol. 3, pp. 1444–1450, IEEE, 2002.
- [Wang05] B. Wang, J. Zhang, and A. Host-Madsen, “On the capacity of MIMO relay channels”, *IEEE Transactions on Information Theory*, vol. 51, no. 1, 29–43, 2005.
- [Windpassinger04] C. Windpassinger, R. Fischer, and J. Huber, “Lattice-reduction-aided broadcast precoding”, *IEEE Transactions on Communications*, vol. 52, no. 12, 2057–2060, 2004.
- [Winters87] J. Winters, “On the capacity of radio communication systems with diversity in a Rayleigh fading environment”, *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, 871–878, 1987.

-
- [Xu08] C. Xu, B. Hu, L. Yang, and L. Hanzo, “Ant-colony-based multiuser detection for multifunctional-antenna-array-assisted MC DS-CDMA systems”, *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, 658–663, 2008.
- [Xu11] W. Xu, X. Dong, and W. Lu, “Joint precoding optimization for multiuser multi-antenna relaying downlinks using quadratic programming”, *IEEE Transactions on Communications*, vol. 59, no. 5, 1228–1235, 2011.
- [Yu01] W. Yu and J. Cioffi, “Trellis precoding for the broadcast channel”, in *IEEE Global Telecommunications Conference (GLOBECOM’01)*, vol. 2, pp. 1344–1348, 2001.
- [Yu10] Y. Yu and Y. Hua, “Power allocation for a MIMO relay system with multiple-antenna users”, *IEEE Transactions on Signal Processing*, vol. 58, no. 5, 2823–2835, 2010.
- [Zhang11] W. Zhang, U. Mitra, and M. Chiang, “Optimization of amplify-and-forward multicarrier two-hop transmission”, *IEEE Transactions on Communications*, vol. 59, no. 5, 1434–1445, 2011.
- [Zhou11] E. Zhou and L. Hao, “Ordered Tomlinson-Harashima precoding in multiuser MIMO relay system”, in *Fifth International Workshop on Signal Design and its Applications in Communications (IWSDA)*, pp. 181–184, 2011.
- [Zschau67] E. Zschau, *A primal decomposition algorithm for linear programming*, Stanford University, 1967.