SAPHYRE

Contract No. FP7-ICT-248001

Adaptive and Robust Signal Processing in Multi-User and Multi-Cellular Environments (final)

D3.1b

Contractual date:	M36
Actual date:	M36
Authors:	Jianshu Zhang, Martin Haardt, Miltiades Filippou, Rajeev Gangula, David Gesbert, Zuleita Ka Ming Ho, Eduard Jorswieck, Eleftherios Karipidis, Erik G. Larsson, Jianhui Li, Johannes Lindblom, Christian Scheunert, Jan Sýkora, Lei Yu, Haibin Zhang
Participants:	CTU, ECM, LiU, TNO, TUD, TUIL
Work package:	WP3
Security:	Public
Nature:	Report
Version:	1.0
Number of pages:	179

Keywords

Spectrum Sharing, Infrastructure Sharing, SAPHYRE Gain, Interference Channel, Relay.

Executive Summary

This document describes the signal processing techniques which have been developed within work package (WP) 3 of the SAPHYRE project. Many fundamental aspects of physical resource source sharing in future wireless networks, including the spectrum and/or the infrastructure sharing, are investigated. First, the information exchange mechanisms between operators, which are a prerequisite for physical resource sharing, are discussed in Chapter 2. This part of the deliverable contains an overview of the information exchange requirements as well as an individual analysis for three sharing scenarios. Afterwards, advanced signal processing algorithms are proposed to accomplish different forms of spectrum and infrastructure sharing majorly by exploiting MIMO techniques. To be specific,

- In Chapter 3 interference mitigation techniques in spectrum sharing multi-cell wireless networks are studied.
  - Section 3.1 considers the joint precoding across the transmitters of cooperating operators (sharing the same spectrum) with varying backhaul data symbol routing overhead. With the assumption of perfect CSI, a sum rate maximization problem with limited backhaul capacity is formulated. Since the conventional precoders cannot be applied directly in this setting, precoder for a given data symbol allocation pattern is designed. Using the developed precoding technique, an algorithm aiming to find the optimal data symbol allocation subject to a constraint on the backhaul capacity is proposed.
  - Section 3.2 deals with the problem of channel estimation in the presence of multi-operator interference generated by spectrum sharing, and more specifically pilot contamination. First, a Bayesian channel estimation method making explicit use of covariance information in the inter-operator interference scenario with pilot contamination is developed. It is shown that the channel estimation performance is a function of the degree to which dominant signal subspaces pertaining to the desired and interference channel covariance overlap with each other. Therefore, the fact that the desired user signals and interfering user signals are received at the operator side with (at least approximately) finite-rank covariance matrices is exploited. Finally, a pilot sequence assignment strategy based on carefully assigning selected groups of users to identical pilot sequences is proposed.
  - Section 3.3 considers the downlink of a multicell network where neigh-

boring multi-antenna base stations share the spectrum and coordinate their frequency and spatial resource allocation strategies to improve the overall network performance. The coordinated scheduling and multiuser transmit beamforming problem is combinatorial. Thus, a mixed-integer second-order cone program is formulated and a low-complexity branch & bound algorithm that yields the optimal solution is proposed.

- In Chapter 4 distributed MIMO signal processing and resource allocations algorithms for multi-operator spectrum sharing networks are developed and discussed.
  - In Section 4.1 a distributed beamforming algorithm is proposed for the two-user multiple-input single-output (MISO) interference channel (IC). The algorithm is iterative and uses as bargaining value the interference that each transmitter generates towards the receiver of the other user. It enables cooperation among the transmitters in order to increase both users' rates by lowering the overall interference. The algorithm is also equally applicable when the transmitters have either instantaneous or statistical channel state information (CSI). Moreover, the outcome of the proposed algorithm is approximately Pareto-optimal.
  - In Section 4.2 the problem of adaptive allocation of spectrum, power, and rate is considered for the downlink of multicell orthogonal frequency-division multiple-access (OFDMA) networks. The considered networks are assumed to have frequency reuse one and discrete-level rates. The the joint allocation problem as a nonlinear mixed integer program (MIP) is formulated, which is computationally intractable to solve optimally for practical problem sizes. Thus, using the fact that the receivers can measure the perceived interference-plus-noise on every subcarrier, the original problem is decomposed into subproblems that are solved by individual base stations. Each subproblem is a linear MIP and its optimal solution can be obtained by means of standard branch-and-cut solvers, at least for small to medium problem sizes. In order to further reduce the complexity, a distributed iterative algorithm that capitalizes on the subgradient method is proposed.
- In Chapter 5 the relay-aided spectrum and infrastructure sharing scenarios are presented. Problems of optimizing the system performance with respect to the sum rate, the secrecy rate, and the energy efficiency are formulated and analyzed.
  - Section 5.1 focuses on a multi-user wireless network which consists of multiple one-way smart instantaneous relays (i.e., the signals of the source-destination link arrive at the same time as the source-relay-destination link) and one-way dumb relays, where the smart relays are able to gather channel state information, perform linear processing and forward sig-

nals whereas the dumb relays are only able to serve as amplifiers. The achievable rate region is studied in the scenario of a) uninformed non-cooperative source-destination nodes (source and destination nodes are not aware of the existence of the relay and are non-cooperative) and (b) informed and cooperative source-destination nodes. Furthermore, an algorithm which is based on interference neutralization is developed and analyzed.

- Section 5.2 studies the users' privacy in heterogeneous dense networks, where the spectrum is shared. On a multi-antenna relay-assisted multi-carrier interference channel, each user shares the frequency and spatial resources with all other users. When the receivers are not only interested in their own signals but also in eavesdropping other users' signals, the cross talk on the frequency and spatial channels becomes information leakage. To avoid such a leakage, a novel secrecy rate enhancing relay strategy is proposed, which utilizes both frequency and spatial resources, termed as *information leakage neutralization*.
- In Section 5.3 a two-operator relay sharing scenario with a one-way amplify-and-forward MIMO relay is presented. The problem of minimizing the transmit power at the relay subject to SINR constraints at the destination nodes is studied. Both optimal and suboptimal precoding solutions are derived.
- Section 5.4 investigates a sharing scenario where both the BS and a two-way amplify-and-forward MIMO relay are shared among multiple operators. The sum rate maximization problem for this system is non-convex and in general NP-hard. Thus, heuristic methods which are based on the channel inversion, the concept of projection based separation of multiple operators (ProBaSeMO), and zero-forcing dirty paper coding are developed and compared.
- The sharing scenario in Section 5.5 consists of multiple two-way amplify-and-forward relays where each relay has its own transmit power constraint. The sum rate maximization problem of such a network is NP-hard. First, a global optimal solution based on the polyblock algorithm is derived. Due to its high computational complexity, this algorithm can only be used as a benchmark. Afterwards, inspired by the polynomial time difference of convex functions (POTDC) method, a sub-optimal solution which has lower complexity but comparable performance is developed. To further reduce the computational complexity, the total SINR eigen-beamformer and an interference neutralization based design are proposed. Detailed comparison and analysis of the proposed algorithms are performed.
- Section 5.6 considers a relay sharing scenario with asymmetric relay roles,

where the sources use network coded modulation (NCM) and the relay applies the hierarchical decode and forward (HDF) strategy. The NCM/HDF technique is vulnerable to the mutual phase rotations of the signals from the sources. Taking the advantage of additional degrees of freedom in SIMO channels, a specific beamforming is tailored for the given NCM/HDF strategy. The resulting beamforming strategy is equalizing the signals at the relay to optimize the processing of hierarchical information to take the full advantage of HDF.

Contents

Notation

Abbreviations

AF	amplify and forward
ANOMAX	algebraic norm maximization
AOA	Angle of Arrival
AQM	Active Queue Management
BD	block diagonalization
BS	base station
CB	covariance based
CBS	Commit Burst Size
CIR	Commit Information Rate
CoZF	coordinated zero-forcing
CPA	Coordinated Pilot Assignment
CSI	channel state information
DCM	dual channel matching
DFT	discrete Fourier transform
EBS	Excessive Burst Size
EIR	Excessive Information Rate
EVC	Ethernet Virtual Circuit
FDD	Frequency Division Duplexing
FlexCoBF	flexible coordinated beamforming
IC	interference channel
ICIC	Inter-Cell Interference Coordination
IETF	Internet Engineering Task Force
IRC	interference relay channel
LTE	Long Term Evolution
LS	Least Squares
MAP	Maximum a Posteriori
MIMO	multiple input multiple output
MISO	multiple input single output
MMSE	minimum mean square error
MR	maximum ratio
MRC	maximum ratio combing
NBS	Nash bargaining solution
NE	Nash equilibrium
P2P	point to point
PDF	Probability Density Function

PIR	Peak Information Rate
PO	Pareto optimal
QoS	Quality of Service
RBD	regularized block diagonalization
RC	relay channel
RED	Random Early Detection
SIMO	Single Input Multiple Output
SINR	signal-to-interference-plus-noise ratio
SLA	Service Level Agreement
SNR	signal-to-noise ratio
SP	Service Provider
SR	sum rate
srTCM	single rate Three Color Marker
SVD	singular value decomposition
TCP	Transmission Control Protocol
TDMA	time division multiple access
trTCM	dual rate Three Color Marker
TSW2CM	Time Sliding Window Two Color Marker
TSW3CM	Time Sliding Window Three Color Marker
TWR	two-way relaying
ULA	Uniform Linear Array
UT	user terminal
VLAN	Virtual LAN
ZF	zero forcing

Mathematical Notation

$\mathbb{C}$	set of complex numbers
$\mathbb{R}$	set of real numbers
$\mathbb{Z}$	set of integers
$\mathcal{CN}$	complex normal distribution
$\mathrm{tr}\{\cdot\}$	trace of a matrix
$\mathrm{rank}\{\cdot\}$	rank of a matrix
$\mathcal{R}\{\cdot\}$	range (span) of a matrix
$\mathcal{N}\{\cdot\}$	nullspace of a matrix
$\mathbf{\Pi_Z}$	orthogonal projection onto the range of $\mathbf{Z}$
$\mathbf{\Pi_Z^\perp}$	orthogonal projection onto the orthogonal complement of the range of $\mathbf{Z}$
$\mathbf{I}$	identity matrix
$\mathbb{E}\{\cdot\}$	expectation operator
$\mathbf{X}^T$	transpose of a matrix
$\mathbf{X}^*$	conjugate of a matrix
$\mathbf{X}^H$	conjugate transpose of a matrix

$\|\cdot\|_F$	Frobenius norm
$\mathbf{X} \otimes \mathbf{Y}$	Kronecker product of two matrices $\mathbf{X}$ and $\mathbf{Y}$

1 Introduction

In current wireless communication systems, the radio spectrum and the infrastructure are typically used such that interference is avoided by exclusive allocation of frequency bands and employment of base stations. SAPHYRE will demonstrate how equal-priority resource sharing in wireless networks improves the spectral efficiency, enhances coverage, increases user satisfaction, leads to increased revenue for operators, and decreases capital and operating expenditures.

The physical resources which are shared can be divided into two classes, namely spectrum and infrastructure. These are shared with respect to a set of 'players', consisting of operators and users. Each player has a set of private information, e.g., operators have their business models and their revenue strategies, users have their private interests and their partly private state information including traffic, mobility, channel parameters. These goals and parameters are usually not revealed to others.

The spectrum sharing is performed with respect to a set of constraints. These constraints are divided into two areas, namely regulatory and environmental constraints. They can partly overlap as in the case of spectrum masks and power constraints which are both regulatory and environmental. The main difference between these two areas is that regulatory constraints contain fairness and social welfare or legal issues whereas environmental constraints contain fundamental limitations imposed by physics.

The resource sharing problems are interdisciplinary and require regulatory and political bodies, business and market experts, and technical input from communication and network engineers. The ongoing discussion about spectrum commons is led mainly from a regulatory and market point of view. However, advances in communication systems (e.g., multi-antenna systems, multi-carrier transmission techniques, adaptive receivers, software defined radio, interference cancellation) are recognized already to have a very strong impact since they enable the efficient and concurrent use of spectrum.

From a communications engineering point of view, different types of orthogonality in frequency, time, space of coding domain have been used for resource allocation depending on the type of interference: For users in one cell operated by one operator (intracell interference) TDMA combined with FDMA (used in GSM systems) or CDMA (combined with TDMA/FDMA in 3G systems) is applied to separate their signals at the receivers. For different sectors or cells, the intercell interference is controlled by applying different frequency reuse factors [1]. Fractional and adaptive

frequency reuse is discussed in LTE and WiMAX [2]. Very recently, techniques for separating transmissions from different operators (inter-operator interference) without orthogonal resource allocation have been developed: First flexible resource sharing approaches have been developed and results indicate that the overall efficiency of the system can be improved by sharing different resources in the network between several operators [3, 4]. Sharing of spectrum or infrastructure ends up in creating interference on the physical layer. Therefore, interest in physical and MAC layer optimization for resource sharing has increased recently.

This report is a summary of the design and performance analysis of adaptive and robust signal processing algorithms to exploit the additional degrees of freedom in multi-user and multi-cellular environments, which are achieved in work package WP3. As already pointed out in the SAPHYRE deliverable D3.1a, it is important to:

- show the gain (loss) with respect to the chosen performance metric as compared to a non-sharing scenario,
- point out conditions when a significant gain can be achieved for the chosen scenario (topology), and
- illustrate the order of magnitude of this gain.

This sharing gain, namely the SAPHYRE gain, can be defined as the performance comparison in terms of various performance metrics (e.g., the system sum-rate, the achievable rate region, etc.). Now we briefly review the two types of SAPHYRE gains which are defined in the SAPHYRE deliverable D3.1a. These definitions will be also applied in the following this report. The SAPHYRE gain is in general defined in terms of a particular performance metric (e.g., the sum rate) of the simultaneous sharing scenario compared to the time-shared use of the spectrum and infrastructure by the operators (time division case, in this case, the operators and users are multiplexed in the time domain). If the sum rate is chosen as the performance criterion, the absolute SAPHYRE gain is defined as

$$\Xi_{\mathrm{A}} = \sum_{k=1}^{K} U_k - \frac{1}{K} \sum_{k=1}^{K} U_k^{\mathrm{SU}}, \tag{1.1}$$

and the fractional SAPHYRE gain is defined as

$$\Xi_{\mathrm{F}} = \frac{\sum_{k=1}^{K} U_k}{\frac{1}{K} \sum_{k=1}^{K} U_k^{\mathrm{SU}}}, \tag{1.2}$$

where $k \in \{1, 2, \cdots, K\}$ is the index of the users. The utility function of the kth user in the sharing scenario and the time division case are denoted by U_k and U_k^{SU}, respectively.

This report is categorized into three parts. In the first part (Chapter 2) the information exchange mechanism, which is required for the algorithms developed in this work package, is discussed. In the second part (Chapters 3 and 4), transmit strategies (Sections 3.1, 3.3, and 4.1), resource allocation schemes (Section 4.2) and channel estimation algorithms (Section 3.2) are studied for multi-operator multi-cell networks. The major difference between Chapter 3 and Chapter 4 is that centralized design/coordination is required for applying the algorithms developed in Chapter 3 while a distributed implementation of the algorithms developed in Chapter 3 is possible. In the third part (Chapter 5), transmission techniques for maximizing the system sum rate (Sections 5.1, 5.4, 5.5, and 5.6), maximizing the secrecy rate (Section 5.2) or minimizing the transmit power (Section 5.3) in relay assisted physical resource sharing scenarios are presented.

2 Information Exchange Mechanisms Between Operators

2.1 Introduction

The advanced physical-layer transmission schemes, investigated by SAPHYRE WP3, intend to coordinate the radio transmission of cooperating operators networks on the same radio resource (time, frequency), and thus to mitigate or even minimize the interference among users of different operators sharing the same radio resource. For this purpose, some prior information is typically required at the transmitters, in order to adapt the transmitted radio signals of different operators in such ways that the interference among these signals (as long as they use the same radio resource simultaneously and in the vicinity of each other) is minimized.

The required prior information and their exchange (among network nodes) differ according to the investigated transmission schemes, duplexing mode (TDD or FDD), direction of the transmission (uplink or downlink), and optionally also to the considered deployment scenarios (e.g. relaying-related or not, co-sited or non-co-sited base stations). In the context of this section below, unless explicitly mentioned, the downlink of FDD systems is assumed in the analysis. Similar analysis can be derived for uplink transmission and the TDD mode with some necessary modification.

Generally, these information may be categorized according to the ways how they are exchanged (among different network nodes or different operators):

- Information exchanged over the air interface between a terminal and the base station. A typical type of such information is the channel state information (CSI) reported by the receiver to the transmitter. In reality, considering the required signaling overhead, CSI feedback is represented by a finite set of values. For example, in LTE each user terminal (UE) uses a combination of the so-called rank indicator (RI, up to 8 values depending on the number of antennas), precoding matrix indicator (PMI, up to 16 values depending on the number of antennas and the recommended RI) and channel quality indicator (CQI, 15 values) to inform the serving cell the downlink channel condition [5]. The RI provides recommendation on the MIMO transmission rank in the downlink, i.e. the number of independent data streams transmitted in parallel. The PMI indicates which of the specified precoding matrices should preferably be used for the downlink transmission. The CQI represents the maximum modulation and coding scheme (MCS) the UE is able to receive with a block-error probability of at most 10% (using the recommended RI

and PMI). These indicators are used by the base station, jointly with other information, for downlink scheduling and link adaptation. Such quantification of CSI using limited number of values (and thus limited number of signaling bits) introduces inaccuracy of CSI knowledge at the transmitter. In LTE, to further reduce the signaling overhead, these indicators are reported not per physical resource block (PRB, equal to 12 OFDM subcarriers or 180 KHz, which is the frequency unit for scheduling in LTE), but rather per group of adjacent PRBs or even the whole bandwidth. Another source of inaccuracy in CSI at the transmitter is the latency, due to the time used for measurement, signal processing, and radio transmission over the air. In LTE, for example, this latency is typically in the range of 3-4 ms. This is crucial for UEs with high mobility, since the CSI feedback received by the transmitter might be already "out-of-date" for the current transmission. Another type of information, to support retransmission (e.g. in the context of H-ARQ), may be sent by the receiver to the transmitter, indicating whether a new initial transmission or a re-transmission of a formerly transmitted data is expected.

- Information exchanged via the mobile backhaul between base stations. Such information could include e.g. the range of spectrum shared or not shared per operator, the information used to coordinate the use of reference signals among operators, and (if applicable) CSI information (reported by the UEs of one operator to their own serving cells, but for the links with the other operator's cells). Such information exchange also takes some time and consumes the capacity of the backhaul. For example, it is often assumed that signaling over the backhaul between LTE base stations (within the same network) has a typical average delay in the region of 10 ms, and in some extreme cases even up to the range of 20 ms [6]. Somewhat higher latency is suspected for inter-operator information exchange if considering e.g. across-operator authentication.

Another aspect to be considered is the synchronization among base stations of different operators. Traditionally synchronization requirements differ between FDD-mode networks and TDD-mode networks: only frequency synchronization has been required for proper operation of FDD-mode mobile networks, while for TDD-mode both frequency synchronization and phase/time synchronization are required. However, in order to support cooperative operation of base stations, new requirements of phase/time synchronization are emerging also for FDD-mode networks, to support certain features such as Multicast Broadcast Multimedia Services (MBSFN) and Coordinated Multi-Point transmission and reception (CoMP). Global Navigation Satellite Systems (GNSS), such as GPS, may be used to achieve synchronizations with required accuracy (e.g. refer to [7] for the required frequency and phase/time accuracy of some technologies). But, GNSS might not always be reliable (e.g. for indoor base stations) and cost-effective. An alternative way is synchronization via network protocols carried by the transport network among base stations. For example, frequency synchronization might be managed via the Ethernet Synchro-

nization Messaging Channel (ESMC) protocol as defined by ITU-T [7]. Phase/time synchronization may be managed at two different layers [8]: physical layer (for the synchronization of timing/phase offset), MAC layer (for the synchronization of data flows). The IEEE 1588 standard [9] is an example of physical-layer phase/time synchronization protocol. Synchronization at the MAC layer is desired for cooperative operation of base stations in a decentralized manner. For example, if the data flows are made synchronously available at each base station, joint beamforming can be implemented in a decentralized manner with each base station applying its beamforming matrices to its own data flow [8]. The authors of [7] proposed an extended version of the ESMC protocol, in order to transport both frequency and phase/time synchronization messages.

In this chapter, we will first give an overall and general analysis of the relation between WP3 transmission schemes and information exchange and synchronization, followed by individual analysis per transmission scheme reported in this deliverable.

2.2 General Analysis

2.2.1 Impact with regard to information exchange over the air interface

In order for UEs to estimate the CSI from base stations of different operators, the reference signals (or pilot patterns) of different operators should be orthogonal to each other, at least for the base stations in the same area. Although in legacy networks (such as LTE) reference signals are not operator-specific, operator-specific reference signals are theoretically possible e.g. by applying an extra operator-specific orthogonal coding to currently cell-specific or UE-specific reference signals. In reality, this means that an extra step of applying operator-specific orthogonal coding should be included in the physical-layer (PHY) technical specifications of a radio technology. Another option is that the involved operators make agreement in using the same group of cell-specific reference signals, as long as no collision occurs in using the same reference signals by two cells in the vicinity of each other. This, however, deserves more management burden of the operators.

In legacy networks, a UE mainly reports the estimated CSI from the serving cell. Only when triggered by certain events, e.g. for the purpose of handover, the UE starts to measure the CSI of neighboring cells belong to the same operator (often in the formats of SINR and/or received signal strength) and if further triggered also reports the measured results to the serving cell. While for the developed WP3 transmission schemes, the UE may often need to continuously measure and report the CSI of the other reference cell(s) (of the other operator) at a certain time scale (see next section). This increases the uplink signaling overhead, and perhaps also the downlink signaling overhead in case that acknowledgment messages are required for such reporting. Note also that this signaling overhead scales with the number of antennas.

The overhead in signaling such extra CSI on the air interface is also dependent of the number of cells participating in the cooperation. In the simplest case, only single cell per operator is involved in the cooperation, and the UE only needs to report one extra set of CSI. The drawback is that, since all the other intra- and inter-operator cells in the neighborhood are not taken into account in the determination of transmission waveforms, the UE may suffer significant inter-cell interference especially in the case of full frequency reuse (i.e. frequency reuse factor of 1). On the other hand, if more neighbor cells are involved in the cooperation, better performance might be expected at the cost of more extra signaling overhead over the air interface. Such theoretical and potential performance gain, however, may be degraded in practice due to the CSI inaccuracy in the sense of e.g. channel estimation error, quantification loss, and latency.

When retransmission is involved, another issue is whether new waveforms (or more precisely pre-coding matrices, in the case of beamforming) should be generated for each retransmission, using updated CSI at the transmitter. There is often tradeoff between processing complexity (for regeneration of new waveforms) and performance.

The sensitivity of performance to inaccurate CSI at the transmitter determines to large extent the applicability of the corresponding transmission scheme to real radio networks, since in reality the CSI is always inaccurate at the transmitter due to the limits addressed in the former section and possible radio transmission errors.

2.2.2 Impact with regard to information exchange over the backhaul

As indicated by [6], transmission over the mobile backhaul is quite reliable with negligible packet error rate. Thus, for CSI exchanged over the backhaul, we assume no extra inaccuracy is introduced. While for other information exchanged over the backhaul, we assume they are always correctly received.

Taking into account the typical latency (i.e. 10 ms) of transmission over the backhaul, it is expected that the relevant information only changes and will be exchanged at a relatively large time scale (e.g. seconds or higher). It is feasible for almost all information exchanged over the backhaul that we have in mind, e.g. the range of spectrum shared or not shared per operator (it is expected to change at time scale no shorter than hours), and the information used to coordinate the use of reference signals among operators (it is often static). One possible exception asking for special concern is, if applicable, the CSI exchanged over the backhaul. Since these CSI will be used for the (re-)generation of radio waveforms, this implies that the corresponding transmission scheme cannot be too dynamic in the sense of updating radio waveforms.

However, the above limitation is relaxed if the cooperating cells are co-located. In the case that co-siting is not possible, it is also beneficial if the cooperating base

stations are connected with fiber links and only with limited number of routers in between. In the latter case, lower latency (< 1 ms) is expected than the typical value of 10 ms [10]. The latency could be even further reduced, if the concept of centralized RAN (C-RAN) as termed by NGMN [11], or "cloud-based RAN", is used. C-RAN is a base station architecture that separates the processing part into a centralized cloud based pool of servers, at one side, and the RF part for up-conversion, power amplification and digital (usually optical) to analog radio transmission, on the other side, into a distributed set of remote radio heads (RRHs). This concept can be seen as extreme extension of the current trend for connecting a RRH on transmit tower with optical fiber (up to few tens meters long) to the processing unit in the BS container on the ground for macro- and micro- BSs. Using the concept of C-RAN, cooperating operators may pool their processing units into a centralized fashion, while the radio transmission parts of each operator are connected to the centralized processing units as RRHs. The transmission latency between the processing units and the RRHs is in the order of μs [10].

The required capacity for information exchange depends on the time scale of information exchange (how often) and the amount of information exchanged (how large). It is hard to estimate quantitatively how much the required capacity will be over the backhaul due to the lack of detailed specification of data formats. Some analysis in literature for cooperative networking, e.g. the analysis in [12] of the signaling traffic over the backhaul for support of LTE-A CoMP, calls for the necessity of deploying optical-fiber for mobile backhauling.

If in the case that the backhaul is not adequate in the terms of latency and/or capacity, an alternative solution could be that the UE reports (at least part of) the CSI to cell(s), other than its own serving cell, where the CSI will be used. However, this is not supported by legacy networks so far. One challenge is that the UE needs to set up multiple radio links with different cells, which, although not impossible theoretically, increases hardware complexity and reduces battery life of the UE significantly.

2.2.3 Impact with regard to synchronization among base stations

For synchronization (frequency and phase/time) among base stations of different operators, it is crucial that the same synchronization mechanism is adopted by the involved operators. It is especially important for synchronization with indoor base stations, since there is no GPS equipped with the latter. That's might be managed via internationally standardized network synchronization protocols like IEEE 1588 and ESMC.

For transmission schemes for which the transmission waveforms are generated in a decentralized manner, the data flows of cooperative base stations is desired to be synchronized, e.g. using one of the approaches proposed in [7]. On the other hand, for centralized approaches, the data flows are naturally synchronous.

It is for further study what the requirements of the WP3 transmission schemes are on the synchronization accuracy.

2.3 Individual Analysis per Transmission Scheme

In this section, we analyze the impact with regard to information exchange and synchronization for three of the transmission schemes reported in this deliverable.

2.3.1 Distributed MIMO signal processing and resource allocation

This transmission scheme targets at downlink transmission. In order for the cooperative base stations (of different operators) to jointly determine their beamforming precoding matrices, the base stations should have the knowledge of the CSI of each pair of transmitters and receivers (either desired or undesired) involved. Each UE needs to measure the reference signals of both its own serving cell and other reference cells. For this purpose, the reference signals of coordinated cells should be orthogonal to each other.

As in legacy networks, we expect that each UE reports their estimated downlink CSI only to its serving cell. Thus, the CSI should be further exchanged via the backhaul, in order to be jointly used for the determination of best precoding matrices. The way how information is exchanged over the backhaul is largely dependent of whether the scheme is implemented in a centralized manner or distributed manner:

- In a centralized manner, all the CSI collected by the cooperative base stations should be further forwarded to the network node where the processing is taken and the decision is made. That can be a separate network node or one of the base stations participating in the coordination. After making decision, the decided precoding matrices should be fed back to all the involved base stations.
- In a distributed manner, the cooperative base stations should exchange CSI they have collected separately. The processing is taken and the decision is made at each of the base stations.

Some basic analysis has been executed with regard to the performance impact of partial CSI, see Section 4.1. The analysis shows that in some cases this partial CSI may introduce up to 5 dB performance loss (in the sense of required SNR for the same throughput), if compared with full CSI. This may imply the importance of full CSI to this transmission scheme. However, we admit that the results might have been impacted by the way how we have modeled the partial CSI (as channel correlation matrices).

In the case that the number of users sharing the same radio resource in one single cell is higher than the number of antennas, another extra parameter needs to be

determined is the transmit power per user. In a centralized manner, this output should also be fed back to the base stations.

This scheme requires synchronization between different operators' base stations. The cooperative base stations should be synchronized to each other at the physical layer, and in the case of a distributed manner also at the MAC layer.

2.3.2 Sharing with instantaneous relaying

In this transmission scheme (see Section 5.1), the decision is made by the relay station (RS) in a centralized manner. For this purpose, for each pair of source nodes (base stations) and destination nodes (terminals) the relay station should have the CSI of three different links: the BS-to-RS link, the RS-to-UE link, and the direct BS-to-UE link. The relay station is able to estimate the CSI of the first link (downlink) or the second link (uplink), while for the other two links the CSI is either estimated by the UE (downlink) or the base station (uplink).

In order to estimate the CSI of the BS-to-RS link (downlink) and RS-to-UE link (uplink), the relay station needs to decode the received signals, and detect the position of reference signals/pilots in the data flow. The CSI estimated by the base stations and UEs could in principle be signaled to the relay node via the air interfaces, with the costs of signaling overhead and some inaccuracy. The relay station should also be able to read these signaled message from the received signals. The delay introduced by such data processing should also be taken into account in practice.

In the analysis we have made so far, we have assumed that two reference cells of different operators are involved in the coordination. In principle, it might be extended to the cases of more cells. The signaling overhead and processing complexity both increase linearly with the number of cooperating cells. This, however, may not always bring obvious performance gain if the additional cells are located relatively far away from the relay station.

This scheme requires synchronization between different operators' base stations and the relay station. The cooperative base stations and the relay station should be synchronized to each other. The synchronization between the relay station and base stations often has to rely on the radio link between the relay station and the base stations, if the relay station is not connected with backhaul or equipped with GPS (that is usually the case). In principle this could be done in a way similar to how terminals synchronize to their serving cell in radio networks: by listening to the synchronization signals in the downlink, and by emitting random access preambles/attempts in the uplink.

2.3.3 Sharing with two-way amplify-and-forward relaying

In this transmission scheme, by default, only two reference cells are involved in the coordination for sharing spectrum and the relay station. Other neighboring cells don't join the coordination. In this case, it is assumed that the relay station estimates the effective CSI (in the format of channel response) between each terminal and its serving cell, and accordingly applies "amplification matrices" before it forwards the received signals. No CSI needs to be exchanged over the air interface nor via the backhaul link between base stations.

In order for the relay station to estimate the channel response of the links to the cells and to the terminals, operator-specific reference signals should be introduced, both in downlink and uplink. In each direction, reference signals of different operators should be orthogonal to each other, in order for the receiver to distinguish signals from different sources and accordingly estimate the corresponding channel responses.

Some basic analysis has been executed with regard to the performance sensitivity of this scheme to the inaccuracy of channel responses estimated by the relay station, see Section 5.4. This basic analysis shows little impact of channel estimation error to the performance in the high SNR regime. However, we admit that since some ideal assumption (e.g. pilot pattern) is assumed in the analysis, more thorough analysis is required in the further study in order to have a more convincing judgment, especially taking practical limitations into account.

The information exchanged over the backhaul might include the information used to coordinate the use of reference signals among operators, the range of spectrum shared, etc. As discussed in the former section, such information is often either static or vary at large time scale.

This scheme requires synchronization between different operators' base stations and the relay station. The cooperative base stations and the relay station should be synchronized to each other. The synchronization between the relay station and base stations often has to rely on the radio link between the relay station and the base stations, if the relay station is not connected with backhaul or equipped with GPS (that is usually the case). In principle this could be done in a way similar to how terminals synchronize to their serving cell in radio networks: by listening to the synchronization signals in the downlink, and by emitting random access preambles/attempts in the uplink.

If more cells join the cooperation and share the same relay station, the signaling overhead increases linearly with the number of cooperating cells. This may not always bring obvious performance gain if the base stations which control these additional cells are located relatively far away from the relay station.

If taking neighboring cells (other than two reference cells) of the same operators into account in the cooperation, while these additional cells don't use the relay station,

or the direct link between cells and terminals is enabled, there will be significant change in the way of operation and the required information exchange:

- First, the precoding matrices at the base stations should also be dynamically adapted, for which knowledge of additional CSI is required at the cells/base station. The amplification matrix design at the relay station may still follow the way of the default case (only the two reference cells are involved in the cooperation).
- The base stations would need to have the CSI of the BS-to-RS link as well as the RS-to-UE link, instead of the above-mentioned effective CSI between the base station and the user terminal. For this purpose, the relay station needs to forward the estimated channel response on the RS-to-UE link to the base stations, consuming some radio resource.
- Base stations need to exchange CSI via the backhaul.

3 Interference Mitigation in Spectrum Sharing Networks

3.1 Partial Data Sharing in Multicell Spectrum Sharing Networks

In this section we consider the joint precoding across K distant transmitters (TXs) towards K single-antenna receivers (RXs) sharing the same spectrum. Joint precoding across distant TXs requires sharing of users data symbols and CSI across the cooperating TXs. Sharing of data symbols and CSI implies high backhaul and feedback overhead which increases as the number of cooperating terminals in the network increases. However, in practical networks, cooperation between TXs is limited by the constraints on the backhaul network. With the assumption of perfect CSI, we address the following issues: How to design a joint precoder for a given routing pattern? How to optimize the data symbol allocation subject to a constraint on the total number of symbols being routed?

3.1.1 System model

The considered network has K TXs, with TX k equipped with N_k antennas and K single-antenna RXs. The total number of antennas at the transmit side is denoted by $N_T = \sum_{k=1}^{K} N_k$. The k-th RX receives

$$y_k = \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{x} + \eta_k \tag{3.1}$$

where $\boldsymbol{h}_k^{\mathrm{H}} \in \mathbb{C}^{1\times N_T}$ represents the channel vector corresponding to the k-th RX, $\boldsymbol{x} \in \mathbb{C}^{N_T\times 1}$ represents the combined transmit signals of all users sent by all the transmit antennas and $\eta_k \sim \mathcal{CN}(0,\sigma^2)$ represents the i.i.d. complex circular-symmetric additive Gaussian noise at the k-th RX. The whole multiuser channel matrix of the system is

$$\mathbf{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K]^{\mathrm{H}}. \tag{3.2}$$

The entries of the channel matrix $\mathbf{H}$ read as $\{\mathbf{H}\}_{ij} = \gamma_{ij} G_{ij}$ where G_{ij} is i.i.d. $\mathcal{CN}(0,1)$ to model the Rayleigh fading and γ_{ij} is a positive real number modeling the long term attenuation. It is assumed that all the TXs have the knowledge of the CSI of the whole multiuser channel. Multi-transmitter cooperative processing in the form of joint linear precoding is adopted. Thus, the transmitted signal $\boldsymbol{x}$ is

obtained from

$$\boldsymbol{x} = \mathbf{T}\boldsymbol{s} = [\boldsymbol{t}_1 \cdots \boldsymbol{t}_K] \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix} \tag{3.3}$$

where $\boldsymbol{s} \in \mathbb{C}^{K\times 1}$ is the vector of transmit symbols (whose entries are independent $\mathcal{CN}(0,1)$), $\mathbf{T}$ is the precoding matrix and $\boldsymbol{t}_k \in \mathbb{C}^{N_T\times 1}$ is the beamforming vector of the symbol k.

Noting that in a statistically symmetric isotropic network, fulfilling a sum power constraint will lead to an equal average power used per TX, we consider for simplicity a sum power constraint $\|\mathbf{T}\|_F^2 = K \times P$. We also assume that all data streams are allocated with an equal amount of power so that $\forall k, \|\boldsymbol{t}_k\|^2 = P$.

Conventional zero-forcing (ZF) schemes result in a complete removal of interference at the receivers. This is optimal at high SNR but not necessarily at intermediate SNR. To further improve the performance of the precoder, regularized ZF precoder, which achieves good performance even at intermediate SNR is used and is given by

$$\mathbf{T} = \frac{\sqrt{KP}}{\left\|\mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \alpha\mathbf{I}\right)^{-1}\right\|}\mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \alpha\mathbf{I}\right)^{-1} \tag{3.4}$$

where $\alpha = K \times (\sigma^2/P)$ is the regularization constant.

The sum-rate of the system is equal to

$$R = \sum_{k=1}^{K} R_k = \sum_{k=1}^{K} \log_2\left(1 + \mathrm{SINR}_k\right). \tag{3.5}$$

where the SINR of the k-th data stream is given by

$$\mathrm{SINR}_k = \frac{|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_k|^2}{\sigma^2 + \sum\limits_{j=1, j\neq k}^{K} |\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_j|^2}. \tag{3.6}$$

Furthermore, the multiplexing gain (MG) of the system is defined as

$$\mathrm{M_G} \triangleq \sum_{k=1}^{K} \mathrm{M_G}_k \triangleq \sum_{k=1}^{K} \lim_{P\to\infty} \frac{R_k(P)}{\log_2(P)}. \tag{3.7}$$

Backhaul Data Symbol Routing

To represent the effect of the allocation of the users' data symbols to the TXs, we introduce the concept of *routing matrix* which specifies to which TXs the symbol of a given user is being routed. In realistic settings, e.g. cellular networks, we

are concerned with the sharing of the users' data symbols to the TXs and not to each antenna individually, as the different antennas at a given TX are collocated and perfectly cooperate. To model the data symbol sharing, routing matrix $\mathbf{D} \in \{0,1\}^{K\times K}$ is defined as the matrix whose element $\{\mathbf{D}\}_{i,j} \in \{0,1\}$ is 1 if symbol s_j is allocated to TX i and 0 otherwise. The number of users' data symbols shared in the backhaul network can be seen to be equal to $d = \|\mathbf{D}\|_{\mathbf{F}}^2$. Yet, the antenna configuration has not been taken into account in the routing matrix. Therefore, more notations are needed to represent it. Thus, the expansion matrix $\mathbf{E} \in \mathbb{C}^{N_T\times K}$ is defined as:

$$\mathbf{E} \triangleq \left[\boldsymbol{A}_1^{\mathrm{T}} \ \cdots \ \boldsymbol{A}_K^{\mathrm{T}}\right]^{\mathrm{T}} \tag{3.8}$$

where the matrix $\boldsymbol{A}_k \in \mathbb{C}^{N_k\times K}$ is defined as $\boldsymbol{A}_k \triangleq \mathbf{1}_{[N_k\times 1]}\mathbf{e}_k^{\mathrm{T}}$. The vector $\mathbf{e}_k \in \mathbb{C}^{K\times 1}$ is the k-th vector from the canonical basis. Now the matrix $\tilde{\mathbf{D}} \triangleq \mathbf{ED}$ can be defined and represents well the data sharing constraints, as will be shown in Section 3.1.2.

Optimization Problem

In order to optimize the backhaul routing directly, the following sum-rate maximization problem is formulated:

$$\begin{aligned} \underset{\mathbf{D}\in\{0,1\}^{K\times K}}{\text{maximize}} \quad & \sum_{k=1}^{K} R_k \\ \text{subject to} \quad & d = \|\mathbf{D}\|_{\mathbf{F}}^2 \leq d^*. \end{aligned} \tag{3.9}$$

where d^* is the constraint on the backhaul routing overhead. Above problem is a discrete combinatorial optimization problem and generally exhaustive search is required to find the optimal data symbol allocation. For exhaustive search, a total of $\binom{K^2}{d^*}$ data allocation combinations need to be searched over, i.e the complexity grows as $\binom{K^2}{d^*}$. This is prohibitive even for small number of cooperating nodes. Therefore, a greedy algorithm having low complexity is proposed.

3.1.2 Precoder optimization

In this section we consider the design of the precoder $\mathbf{T}$ given the routing matrix $\mathbf{D}$. Note that each beamforming vector can be derived independently due to the ZF constraints. If one TX does not receive one symbol, it cannot participate into the transmission of that symbol and the coefficient used for that beamformer at that TX is then 0. Therefore the precoder with constrained backhaul overhead will be of the form $\mathbf{T} \odot \tilde{\mathbf{D}}$, where $\odot$ is the element wise product. The beamforming vector $\boldsymbol{t}_k$ transmitting the symbol k is obtained from the following optimization:

$$\underset{\boldsymbol{t}_k}{\text{minimize}} \left\|\mathbf{H}\left(\boldsymbol{t}_k \odot \tilde{\mathbf{D}}_{:,k}\right) - \boldsymbol{\phi}_{:,k}\right\|_2^2 \tag{3.10}$$

where $\boldsymbol{\phi} = \mathbf{HT}$ and the precoder $\mathbf{T}$ is obtained using (3.4) with full routing.

In optimization (3.9) we consider no predefined constraint on the routing pattern's structure. Therefore, there may be some columns of the routing matrix containing only zero i.e. a user is not served. This is some kind of user selection and accounts for a positive MG with partial data sharing. Thus, a *non-active user* is defined as a user whose data symbol is not routed to any TX, i.e., $\mathbf{D}_{:,k} = \mathbf{0}_{K\times 1}$, or the power allocated to the user $p_k = 0$. If there are *non-active users* in the system, the precoding scheme needs to be modified so that interference is removed only at active users. We start by introducing some notations and we denote the set of indices such that $\boldsymbol{t}_k \odot \tilde{\mathbf{D}}_{:,k} \neq 0$ by $\mathcal{J}$ and the reduced vector $\bar{\boldsymbol{t}}_k(\mathcal{J}) \in \mathbb{C}^{n_2\times 1}$ with $n_2 = |\mathcal{J}|$ made of the elements of $\mathcal{J}$. Further more, the set of indices corresponding to the active-users is denoted by $\mathcal{A}$, with $n_1 = |\mathcal{A}|$. The matrix $\mathbf{H}(\mathcal{A},\mathcal{J}) \in \mathbb{C}^{n_1\times n_2}$ is used to represent the channel containing only the rows and columns consisting in $\mathcal{A}$ and $\mathcal{J}$ respectively. Finally, $\tilde{\boldsymbol{\phi}} = \boldsymbol{\phi}(\mathcal{A},\mathcal{A})$ represents a sub matrix of $\boldsymbol{\phi}$ formed by keeping the rows and columns of the active-users. The beamforming vector $\bar{\boldsymbol{t}}_k$ can now be obtained from the following optimization:

$$\underset{\bar{\boldsymbol{t}}_k(\mathcal{J})}{\text{minimize}} \left\| \mathbf{H}(\mathcal{A},\mathcal{J})\,\bar{\boldsymbol{t}}_k(\mathcal{J}) - \tilde{\boldsymbol{\phi}}_k \right\|_2^2 \tag{3.11}$$

which can be solved as a conventional Least Square problem. The beamforming vector of user k is then obtained by reinserting the coefficients obtained in $\bar{\boldsymbol{t}}_k$ at the positions corresponding to $\mathcal{J}$ in $\boldsymbol{t}_k$.

3.1.3 Multiplexing gain analysis

In this section we consider the fundamental limit behind the optimization problem (3.9) and look for the maximum achievable MG when there is a constraint on the number of data symbols allocated. In the following analysis we assume that there will always be enough users in the system so that the MG is not restricted by the total degrees of freedom available at the RXs. Mathematically, if γ is the maximum MG, then $K \geq \gamma$.

Theorem 3.1. *In order to achieve the maximum MG γ under the constraint $\|\mathbf{D}\|_F^2 \leq d^*$, the following conditions needs to be satisfied*

$$\begin{aligned} &\sum_{k=1}^{K_{TX}} N_k \geq \gamma \,(\ \textit{ZF Feasibility}\ ), \\ &\gamma \times K_{TX} \leq d^* \,(\ \textit{Sharing Constraint}\ ). \end{aligned} \tag{3.12}$$

Proof. Proof can be found in [13]. □

Corollary 3.1. *The maximum MG γ achieved with all TXs having N antennas under the constraint $\|\mathbf{D}\|_F^2 \leq d^*$ is given by*

$$\gamma = \max \left\{ N \left\lfloor \sqrt{d^*/N} \right\rfloor, \left\lfloor \frac{d^*}{\left\lceil \sqrt{d^*/N} \right\rceil} \right\rfloor \right\}. \tag{3.13}$$

Proof. The proof can be found in [13]. □

3.1.4 Greedy data symbol allocation

Greedy Algorithms

We consider the decreasing greedy algorithm (DEC), in which the routing matrix is initialized with $\mathbf{D} = \mathbf{1}_{K\times K}$ and at each iteration the element of $\mathbf{D}$ which causes the least degradation in the sum-rate is set to 0 (i.e. removing the routing link of one symbol to one of the TX) . This process is continued till the constraint on the number of backhaul links is reached.

Algorithm 1 Decreasing greedy algorithm

Input: $\mathbf{H}$, d^*, **Output:** $\mathbf{D}, \boldsymbol{p}$

Require: $\mathbf{D} = \mathbf{1}_{N\times K}$, $\boldsymbol{p} = P \times \mathbf{1}_{K\times 1}$

for $d = 1$ **to** $K^2 - d^*$ **do**
  $C_{init} = 0$, $\mathbf{D}_{temp} = \mathbf{D}$, $\boldsymbol{p} = \boldsymbol{p}'$
  for RX $k = 1$ **to** K **do**
    for TX $l = 1$ **to** K **do**
      if $\{\mathbf{D}_{temp}\}_{lk} \neq 0$ **then**
        $\{\mathbf{D}_{temp}\}_{lk} = 0$
        $\mathbf{T} =$ **precoding** $(\mathbf{D}_{temp}, \mathbf{H}, \boldsymbol{p})$ %(From Sec.3.1.2)
        $C_{sum} =$ **sumrate**$(\mathbf{T}, \mathbf{H})$ %using (3.5)
        if $C_{sum} \geq C_{init}$ **then**
          $m = l, n = k$, $C_{init} = C_{sum}$
        end if
        $\mathbf{D}_{temp} = \mathbf{D}$
      end if
    end for
  end for
  $\{\mathbf{D}\}_{mn} = 0$, $\mathbf{T} =$ **precoding** $(\mathbf{D}, \mathbf{H}, \boldsymbol{p})$
  $C_{sum} =$ **sumrate**$(\mathbf{T}, \mathbf{H})$ %using (3.5)
  $\boldsymbol{p}' =$**Power allocation**$(\boldsymbol{p}, \mathbf{T}, C_{sum})$ %(C.f 3.1.4)
end for

The greedy algorithm at any given step try to maximize the performance at each step and it does not necessarily achieve the maximum MG. By not being MG

optimal does not hurt the performance at low SNR but in intermediate and high SNR it has a considerable effect on the performance of the algorithm. Therefore, some improvements to the greedy algorithm is proposed based on the analysis done in Section 3.1.3.

Binary Power Control In Binary Power Control (BPC) power allocated to k-th user p_k takes only two values 0 or P [14]. In this work we use the idea of BPC in Algorithm 1 in step 18 to make it MG optimal. After each step of removing a data symbol from $\mathbf{D}$ we check whether by turning off a user completely, results in an increase in the sum-rate. If the sum-rate increases by turning off k-th user completely, then the power allocated to that user $p_k = 0$.

Proposition 3.1. *With single antenna TXs, the DEC with BPC as described in Algorithm 1 achieves the optimal MG.*

Proof. The proof can be found in [13]. □

3.1.5 Simulations

We investigate the performance of a network consisting of $K = 7$ cells. Each cell consists of one TX located at the center of the cell. In the simulations, an average cell edge SNR of 20 dB is maintained by selecting the proper transmit power and the cell radius. The simulation results are averaged over uniform randomly generated RX positions (such that each cell has exactly one RX) and Rayleigh fading realizations. In Figure 3.1 the performance of the greedy algorithm with multiple antennas at the TXs with $[N_1, N_2, ..., N_7]$ =[2 1 1 3 1 2 1] is considered.

The greedy algorithm outperforms the conventional time-sharing scheme when the data symbol sharing overhead is more than 10% of the full cooperation scenario. We define the SAPHYRE gain to be the ratio of sum rate performances between the proposed algorithm and a time-sharing scheme without any cooperation among transmitters. For illustration purposes, we can see from Figure 3.1 that the SAPHYRE gain is 2.4 when the backhaul overhead is 50% of the full cooperation scenario. As the allowed backhaul overhead increases the SAPHYRE gain also increases.

3.1.6 Conclusion

The performance of joint precoding across cooperating TXs under varying backhaul data symbol routing overhead is studied. Simulation results shows that the proposed routing algorithm out performs the the conventional time-sharing schemes for practical backhaul overhead values.

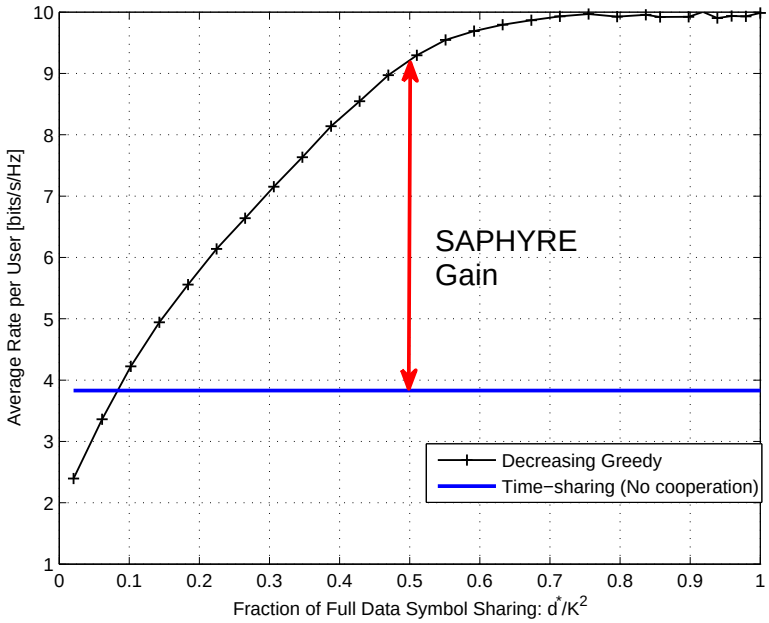


Figure 3.1: Sum-rate in terms of fraction of full data symbol sharing for a average cell edge SNR of 20dB and the antennas corresponding to the TXs are distributed as [2 1 1 3 1 2 1].

3.2 A Coordinated Approach to Channel Estimation

In this section, we address the problem of channel estimation in the presence of multi-operator interference generated by spectrum sharing, and more specifically pilot contamination. We propose an estimation method which provides a substantial improvement in performance. It relies on two key ideas. The first is the exploitation of dormant side-information lying in the second-order statistics of the user channels, both for desired and interfering users. In particular, we demonstrate a powerful result indicating that the exploitation of covariance information under certain subspace conditions on the covariance matrices can lead to a complete removal of pilot contamination effects in the large number of antennas limit. We then turn to a practical algorithm design where this concept is exploited. The key idea behind the new algorithm is the use of a covariance-aware pilot assignment strategy among spectrum sharing operators within the channel estimation phase itself.

More specifically, our contributions are the following: We first develop a Bayesian channel estimation method making explicit use of covariance information in the inter-operator interference scenario with pilot contamination. We show that the channel estimation performance is a function of the degree to which dominant signal subspaces pertaining to the desired and interference channel covariance overlap with each other. Therefore we exploit the fact that the desired user signals and interfering

user signals are received at the operator side with (at least approximately) finite-rank covariance matrices. This is typically the case in realistic scenarios due to the limited angle spread followed by incoming paths originating from street-level users [15]. Finally, we propose a pilot sequence assignment strategy based on carefully assigning selected groups of users to identical pilot sequences.

The gains are shown to depend on system parameters such as the typical angle spread measured at the service provider (SP) and the number of SP antennas. Performance close to the interference-free channel estimation scenario is obtained for moderate numbers of antennas and users.

3.2.1 Signal and channel models

We consider a network of L time-synchronized[1] spectrum sharing operators (for ease of exposition one can set L=2), with full spectrum reuse. Estimation of (block-fading) channels in the uplink is considered,[2] and all the SPs are equipped with M antennas each. To simplify the notations, we assume the 1st SP is the target SP, unless otherwise notified. We assume the pilots, of length τ, used by single-antenna users in the same SP area are mutually orthogonal. As a result, intra-cell interference is negligible in the channel estimation phase. However, non-orthogonal (possibly identical) pilots are reused from SP area to SP area, resulting in pilot contamination from $L-1$ interfering SPs. For ease of exposition, we consider the case where a single user per operator transmits its pilot sequence to its serving SP. The pilot sequence used in the l-th SP is denoted by:

$$\mathbf{s}_l = [\, s_{l1} \;\; s_{l2} \;\; \cdots \;\; s_{l\tau} \,]^T. \tag{3.14}$$

The powers of pilot sequences are assumed equal such that $|s_{l1}|^2 + \cdots + |s_{l\tau}|^2 = \tau, l = 1, 2, \ldots, L$.

The channel vector between the l-th SP area's user and the target SP is $\mathbf{h}_l$. Thus, $\mathbf{h}_1$ is the desired channel while $\mathbf{h}_l$, $l > 1$ are interference channels. All channel vectors are assumed to be $M \times 1$ complex Gaussian, undergoing correlation due to the finite multipath angle spread at the SP side [16]:

$$\mathbf{h}_l = \mathbf{R}_l^{1/2} \mathbf{h}_{\mathrm{W}l}, l = 1, 2, \ldots, L, \tag{3.15}$$

where $\mathbf{h}_{\mathrm{W}l} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the spatially white $M \times 1$ SIMO channel, and $\mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ denotes zero-mean complex Gaussian distribution with covariance matrix $\mathbf{I}_M$. In

[1] Note that assuming synchronization between uplink pilots provides a worst case scenario from a pilot contamination point of view, since any lack of synchronization will tend to statistically decorrelate the pilots.

[2] Similar ideas would be applicable for downlink channel estimation, provided the UT is equipped with multiple antennas as well, in which case the estimation would help resolve interferences originating from neighboring SPs.

this scenario, we make the assumption that the covariance matrix $\mathbf{R}_l \triangleq \mathbb{E}\{\mathbf{h}_l \mathbf{h}_l^H\}$ can be obtained separately from the desired and interference channels.

During the pilot phase, the $M \times \tau$ signal received at the target SP is

$$\mathbf{Y} = \sum_{l=1}^{L} \mathbf{h}_l \mathbf{s}_l^T + \mathbf{N}, \tag{3.16}$$

where $\mathbf{N} \in \mathbb{C}^{M \times \tau}$ is the spatially and temporally white additive Gaussian noise (AWGN) with zero-mean and element-wise variance σ_n^2.

3.2.2 Covariance-based channel estimation

Pilot Contamination due to Spectrum Sharing

Conventional channel estimation relies on correlating the received signal with the known pilot sequence (referred here as Least Squares (LS) estimate for example). Hence, using the model in (3.16), a LS estimator for the desired channel $\mathbf{h}_1$ is

$$\widehat{\mathbf{h}}_1^{\mathrm{LS}} = \mathbf{Y}\mathbf{s_1}^*(\mathbf{s_1}^T\mathbf{s_1}^*)^{-1}. \tag{3.17}$$

The conventional estimator suffers from a lack of orthogonality between the desired and interfering pilots, an effect known as pilot contamination [17], [18], [19]. In particular, when the *same* pilot sequence is reused in all L SPs, i.e., $\mathbf{s}_1 = \cdots = \mathbf{s}_L = \mathbf{s}$, the estimator can be written as

$$\widehat{\mathbf{h}}_1^{\mathrm{LS}} = \mathbf{h}_1 + \sum_{l \neq 1}^{L} \mathbf{h}_l + \mathbf{N}\mathbf{s}^*/\tau\,. \tag{3.18}$$

As it appears in (3.18), the interfering channels leak directly into the desired channel estimate. The estimation performance is then limited by the signal to interference ratio at the SP side, which in turns limits the ability to design an effective interference-avoiding beamforming solution.

Bayesian Estimation

We hereby propose an improved channel estimator with the aim of reducing the pilot contamination effect, and taking advantage of the multiple antenna dimensions. We suggest to do so by exploiting side information lying in the second order statistics of the channel vectors. The role of covariance matrices is to capture structural information related to the distribution (mainly mean and spread) of the multipath angles of arrival at the SP. Due to the typically elevated position of the operator, rays impinge on the antennas with a finite angle-of-arrival (AOA) spread and a user

location-dependent mean angle. Note that covariance-aided channel estimation itself is not a novel idea, e.g., in [20]. In [21], the authors focus on optimal design of pilot sequences and they exploit the covariance matrices of desired channels and colored interference. The optimal training sequences were developed with adaptation to the statistics of disturbance. In our scenario, however, the pilot design is shown not having an impact on interference reduction, since fully aligned pilots are transmitted. Instead, we focus on i) studying the limiting behavior of covariance-based estimates in the presence of interference and large-scale antenna arrays, and ii) how to *shape* covariance information for the full benefit of channel estimation quality.

Two Bayesian channel estimators can be formed. In the first, all channels are estimated at the target SP (including interfering ones). In the second, only $\mathbf{h_1}$ is estimated.

By vectorizing the received signal and noise, our model (3.16) can be represented as

$$\mathbf{y} = \tilde{\mathbf{S}}\mathbf{h} + \mathbf{n}, \tag{3.19}$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{n} = \text{vec}(\mathbf{N})$, and $\mathbf{h} \in \mathbb{C}^{LM\times 1}$ is obtained by stacking all L channels into a vector.

The pilot matrix $\tilde{\mathbf{S}}$ is defined as

$$\tilde{\mathbf{S}} \triangleq \left[\, \mathbf{s}_1 \otimes \mathbf{I}_M \; \cdots \; \mathbf{s}_L \otimes \mathbf{I}_M \,\right]. \tag{3.20}$$

By applying Bayes' rule and by using the maximum a posteriori (MAP) decision rule, one can obtain the following expression for the Bayesian estimator [22]:

$$\widehat{\mathbf{h}} = (\sigma_n^2 \mathbf{I}_{LM} + \mathbf{R}\tilde{\mathbf{S}}^H\tilde{\mathbf{S}})^{-1}\mathbf{R}\tilde{\mathbf{S}}^H\mathbf{y}. \tag{3.21}$$

Interestingly, the Bayesian estimate as shown in (3.21) coincides with the minimum mean square error (MMSE) estimate, which has the form

$$\widehat{\mathbf{h}}^{\text{MMSE}} = \mathbf{R}\tilde{\mathbf{S}}^H(\tilde{\mathbf{S}}\mathbf{R}\tilde{\mathbf{S}}^H + \sigma_n^2\mathbf{I}_{\tau M})^{-1}\mathbf{y}. \tag{3.22}$$

(3.21) and (3.22) are equivalent thanks to the matrix inversion identity $(\mathbf{I}+\mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I}+\mathbf{BA})^{-1}$.

Channel Estimation with Full Pilot Reuse

Matched filters require the knowledge of the desired channel only, so that interference channels can be considered as nuisance parameters. For this case, the single user channel estimation shown below can be used. For ease of exposition, the worst case situation with a unique pilot sequence reused in all L cells is considered:

$$\mathbf{s} = [\, s_1 \; s_2 \; \cdots \; s_\tau \,]^T. \tag{3.23}$$

Similar to (3.20), we define a training matrix $\bar{\mathbf{S}} \triangleq \mathbf{s} \otimes \mathbf{I}_M$. Note that $\bar{\mathbf{S}}^H\bar{\mathbf{S}} = \tau\mathbf{I}_M$. Then the vectorized received training signal at the target operator can be expressed as

$$\mathbf{y} = \bar{\mathbf{S}} \sum_{l=1}^{L} \mathbf{h}_l + \mathbf{n}. \tag{3.24}$$

Since the Bayesian estimator and the MMSE estimator are identical, we omit the derivation and simply give the expression of this estimator for the desired channel $\mathbf{h}_1$ only:

$$\widehat{\mathbf{h}}_1 = \mathbf{R}_1\bar{\mathbf{S}}^H \left(\bar{\mathbf{S}} \left(\sum_{l=1}^{L} \mathbf{R}_l \right) \bar{\mathbf{S}}^H + \sigma_n^2 \mathbf{I}_{\tau M} \right)^{-1} \mathbf{y} \tag{3.25}$$

$$= \mathbf{R}_1 \left(\sigma_n^2 \mathbf{I}_M + \tau \sum_{l=1}^{L} \mathbf{R}_l \right)^{-1} \bar{\mathbf{S}}^H \mathbf{y}. \tag{3.26}$$

Note that the MMSE channel estimation in the presence of identical pilots is also undertaken in other works such as [23].

Focusing on the mean squared error (MSE) of the proposed estimators, this can be defined as: $\mathcal{M} \triangleq \mathbb{E}\{\|\widehat{\mathbf{h}} - \mathbf{h}\|_F^2\}$, or for the single user channel estimate $\mathcal{M}_1 \triangleq \mathbb{E}\{\|\widehat{\mathbf{h}}_1 - \mathbf{h}_1\|_F^2\}$.

The estimation MSE of (3.21) is

$$\mathcal{M} = \mathrm{tr}\left\{ \mathbf{R} \left(\mathbf{I}_{LM} + \frac{\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}}{\sigma_n^2}\mathbf{R} \right)^{-1} \right\}. \tag{3.27}$$

Specifically, when identical pilots are used within all SP areas, the MSEs are

$$\mathcal{M} = \mathrm{tr}\left\{ \mathbf{R} \left(\mathbf{I}_{LM} + \frac{\tau\mathbf{J}_{\mathrm{LL}} \otimes \mathbf{I}_M}{\sigma_n^2}\mathbf{R} \right)^{-1} \right\}, \tag{3.28}$$

$$\mathcal{M}_1 = \mathrm{tr}\left\{ \mathbf{R}_1 - \mathbf{R}_1^2 \left(\frac{\sigma_n^2}{\tau}\mathbf{I}_M + \sum_{l=1}^{L} \mathbf{R}_l \right)^{-1} \right\}, \tag{3.29}$$

where $\mathbf{J}_{\mathrm{LL}}$ is an $L \times L$ unit matrix consisting of all 1s.

We can readily obtain the channel estimate of (3.26) in an interference-free scenario, by setting interference terms to zero:

$$\widehat{\mathbf{h}}_1^{\text{no int}} = \mathbf{R}_1 \big(\sigma_n^2 \mathbf{I}_M + \tau\mathbf{R}_1 \big)^{-1} \bar{\mathbf{S}}^H (\bar{\mathbf{S}}\mathbf{h}_1 + \mathbf{n}), \tag{3.30}$$

where the superscript *no int* refers to the "no interference case", and the corresponding MSE:

$$\mathcal{M}_1^{\text{no int}} = \text{tr}\left\{\mathbf{R}_1\left(\mathbf{I}_M + \frac{\tau}{\sigma_n^2}\mathbf{R}_1\right)^{-1}\right\}. \tag{3.31}$$

Large Scale Analysis

We seek to analyze the performance for the above estimators in the regime of large antenna number M. For tractability, our analysis is based on the assumption of uniform linear array (ULA) with supercritical antenna spacing (i.e., less than or equal to half wavelength).

Hence we have the following multipath model[3]

$$\mathbf{h}_i = \frac{1}{\sqrt{P}}\sum_{p=1}^{P}\mathbf{a}(\theta_{ip})\alpha_{ip}, \tag{3.32}$$

where P is the arbitrary number of i.i.d. paths, $\alpha_{ip} \sim \mathcal{CN}(0, \delta_i^2)$ is independent over channel index i and path index p, where δ_i is the i-th channel's average attenuation. $\mathbf{a}(\theta)$ is the steering vector, as shown in [24]

$$\mathbf{a}(\theta) \triangleq \begin{bmatrix} 1 \\ e^{-j2\pi\frac{D}{\lambda}\cos(\theta)} \\ \vdots \\ e^{-j2\pi\frac{(M-1)D}{\lambda}\cos(\theta)} \end{bmatrix}, \tag{3.33}$$

where D is the antenna spacing at the SP and λ is the signal wavelength, such that $D \leq \lambda/2$. $\theta_{ip} \in [0, \pi]$ is a random AOA. Note that we can limit angles to $[0, \pi]$ because any $\theta \in [-\pi, 0]$ can be replaced by $-\theta$ giving the same steering vector.

Below, we momentarily assume that the selected users exhibit multipath AOAs that do not overlap with the AOAs for the desired user, i.e., the AOA spread and user locations are such that multipath for the desired user are confined to a region of space where interfering paths are very unlikely to exist. Our main result is as follows:

Theorem 3.2. *Assume the multipath angle of arrival θ yielding channel $\mathbf{h}_j, j = 1, \ldots, L$, in (3.32), is distributed according to an arbitrary density $p_j(\theta)$ with bounded support, i.e., $p_j(\theta) = 0$ for $\theta \notin [\theta_j^{\min}, \theta_j^{\max}]$ for some fixed $\theta_j^{\min} \leqslant \theta_j^{\max} \in$*

[3]Note that the Gaussian model (3.15) can well approximate the multipath model (3.32) as long as there are enough paths. Since the number of elementary paths is typically very large, we have $P \gg 1$ this assumption is valid in practice.

$[0, \pi]$. *If the* $L-1$ *intervals* $[\theta_i^{\min}, \theta_i^{\max}]$, $i = 2, \ldots, L$ *are strictly non-overlapping with the desired channel's AOA interval*[4] $[\theta_1^{\min}, \theta_1^{\max}]$, *we have*

$$\lim_{M\to\infty} \widehat{\mathbf{h}}_1 = \widehat{\mathbf{h}}_1^{no\ int}. \tag{3.34}$$

Proof. From the channel model (3.32), we get

$$\mathbf{R}_i = \frac{\delta_i^2}{P} \sum_{p=1}^{P} \mathbb{E}\{\mathbf{a}(\theta_{ip})\mathbf{a}(\theta_{ip})^H\} = \delta_i^2 \mathbb{E}\{\mathbf{a}(\theta_i)\mathbf{a}(\theta_i)^H\},$$

where θ_i has the PDF $p_i(\theta)$ for all $i = 1, \ldots, L$. The proof of Theorem 3.2 relies on three intermediate lemmas which exploit the eigenstructures of the covariance matrices. The proofs of the lemmas are provided in [22]. The essential ingredient is to exhibit an asymptotic (at large M) orthonormal vector basis for $\mathbf{R}_i$ constructed from steering vectors at regularly sampled spatial frequencies. □

Lemma 3.1. *Define* $\boldsymbol{\alpha}(x) \triangleq [\,1\ e^{-j\pi x} \cdots e^{-j\pi(M-1)x}\,]^T$ *and* $\mathcal{A} \triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in [-1, 1]\}$. *Given* $b_1, b_2 \in [-1, 1]$ *and* $b_1 < b_2$, *define* $\mathcal{B} \triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in [b_1, b_2]\}$, *then*

- $\dim\{\mathcal{A}\} = M$
- $\dim\{\mathcal{B}\} \sim (b_2 - b_1)M/2$ *when* M *grows large.*

Proof. The proof is given in [22]. □

Lemma 3.1 characterizes the number of dimensions a linear space has, which is spanned by $\boldsymbol{\alpha}$(x), in which x plays the role of spatial frequency.

Lemma 3.2. *With a bounded support of AOAs, the rank of channel covariance matrix* $\mathbf{R}_i$ *satisfies:*

$$\frac{\text{rank}(\mathbf{R}_i)}{M} \leqslant d_i, \ \text{as } M \to \infty,$$

where d_i *is defined as*

$$d_i \triangleq \left(\cos(\theta_i^{min}) - \cos(\theta_i^{max})\right) \frac{D}{\lambda}.$$

Proof. The proof is given in [22]. □

Lemma 3.2 indicates that for large M, there exists a null space null($\mathbf{R}_i$) of dimension $(1-d_i)M$. Interestingly, related eigenstructure properties of the covariance matrices were independently derived in [25] for the purpose of reducing the overhead of downlink channel estimation and CSI feedback in massive MIMO for FDD systems.

[4] This condition is just one example of practical scenario leading to non-overlapping signal subspaces between the desired and the interference covariances, however, more general multipath scenarios could be used.

Lemma 3.3. *The null space* $\text{null}(\mathbf{R}_i)$ *includes a certain set of unit-norm vectors:*

$$\text{null}(\mathbf{R}_i) \supset \text{span}\left\{\frac{\mathbf{a}(\Phi)}{\sqrt{M}}, \forall \Phi \notin [\theta_i^{min}, \theta_i^{max}]\right\}, \text{ as } M \to \infty.$$

Proof. The proof is given in [22]. □

This lemma indicates that multipath components with AOA outside the AOA region for a given user will tend to fall in the null space of its covariance matrix in the large-number-of-antennas case.

3.2.3 Coordinated pilot assignment

In this section, we exploit the aforementioned results in order to design a suitable coordination protocol for assigning pilot sequences to users in the L SPs. The role of the coordination is to optimize the use of covariance matrices in an effort to try and satisfy the non-overlapping AOA constraint of Theorem 3.2. We assume that in all L SPs, the considered pilot sequence will be assigned to one (out of K) user in each of the L SPs. Let $\mathcal{G} \triangleq \{1, \ldots, K\}$, then $\mathcal{K}_l \in \mathcal{G}$ denotes the index of the user in the l-th SP area who is assigned the pilot sequence $\mathbf{s}$. The set of selected users is denoted by $\mathcal{U}$ in what follows.

We use the estimation MSE in (3.29) as a performance metric to be minimized in order to find the best user set. (3.28) is an alternative if we take the estimates of interfering channels into consideration. For a given user set $\mathcal{U}$, we define a network utility function

$$\text{F}(\mathcal{U}) \triangleq \sum_{j=1}^{|\mathcal{U}|} \frac{\mathcal{M}_j(\mathcal{U})}{\text{tr}\{\mathbf{R}_{jj}(\mathcal{U})\}}, \tag{3.35}$$

where $|\mathcal{U}|$ is the cardinal number of the set $\mathcal{U}$. $\mathcal{M}_j(\mathcal{U})$ is the estimation MSE for the desired channel at the j-th SP, with a notation readily extended from $\mathcal{M}_1$ in (3.29), where this time spectrum sharing operator j is the target one when computing $\mathcal{M}_j$. $\mathbf{R}_{jj}(\mathcal{U})$ is the covariance matrix of the desired channel at the j-th spectrum sharing operator.

The principle of the coordinated pilot assignment consists in exploiting covariance information at all SP areas (a total of KL^2 covariance matrices) in order to minimize the sum MSE metric. In view of minimizing the search complexity, a classical greedy approach is proposed:

1) Initialize $\mathcal{U} = \emptyset$

2) For $l = 1, \ldots, L$ do:

$\mathcal{K}_l = \arg\min_{k \in \mathcal{G}} \text{F}(\mathcal{U} \cup \{k\})$

$\mathcal{U} \leftarrow \mathcal{U} \cup \{\mathcal{K}_l\}$

3) End

The coordination can be interpreted as follows: To minimize the estimation error, a SP tends to assign a given pilot to the user whose spatial feature has most differences with the interfering users assigned the same pilot. Clearly, the performance will improve with the number of users, as it becomes more likely to find users with discriminable second-order statistics.

3.2.4 Numerical results and conclusions

In order to preserve fairness between users and avoid having high-SNR users being systematically assigned the considered pilot, we consider a symmetric multi-operator network where the users are all distributed on the SP area edge and have the same distance from their SPs. In practice, users with greater average SNR levels (but equal across operator areas) can be assigned together on a separate pilot pool. Some basic simulation parameters are given in Table 3.1. We keep these parameters in the following simulations, unless otherwise stated.

Table 3.1: Basic simulation parameters

Operator area radius	1 km
Operator area edge SNR	20 dB
Number of users per operator area	10
Distance from a user to its operator	800 m
Path loss exponent	3
Carrier frequency	2 GHz
Antenna spacing	$\lambda/2$
Number of paths	50
Pilot length	10

The channel vector between the u-th user in the l-th SP area and the target SP is

$$\mathbf{h}_{lu} = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} \mathbf{a}(\theta_{lup}) \alpha_{lup}, \tag{3.36}$$

where θ_{lup} and α_{lup} are the AOA and the attenuation of the p-th path between the u-th user in the l-th SP area and the target SP respectively. Note that the variance of $\alpha_{lup}, \forall p$ is δ_{lu}^2, which includes the distance-based path loss β_{lu} between the user and the target operator (which can be anyone of the L operators):

$$\beta_{lu} = \frac{\alpha}{d_{lu}^{\gamma}}, \tag{3.37}$$

where α is a constant dependent on the prescribed average SNR at SP area edge. d_{lu} is the geographical distance. γ is the path-loss exponent.

A bounded (uniform) AOA distribution is considered in this case. Moreover, two performance metrics are used to evaluate the proposed channel estimation scheme. The first one is a normalized channel estimation error whereas the second performance metric is the per-operator rate of the downlink obtained assuming standard MRC beamformer based on the channel estimates. The beamforming weight vector of the j-th SP is $\mathbf{w}_j^{\mathrm{MRC}} = \widehat{\mathbf{h}}_{jj}$.

Numerical results of the proposed channel estimation scheme are now shown. In the figures, "LS" stands for conventional LS channel estimation. "CB" denotes the Covariance-aided Bayesian estimation (without coordinated pilot assignment), and "CPA" is the proposed Coordinated Pilot Assignment-based Bayesian estimation.

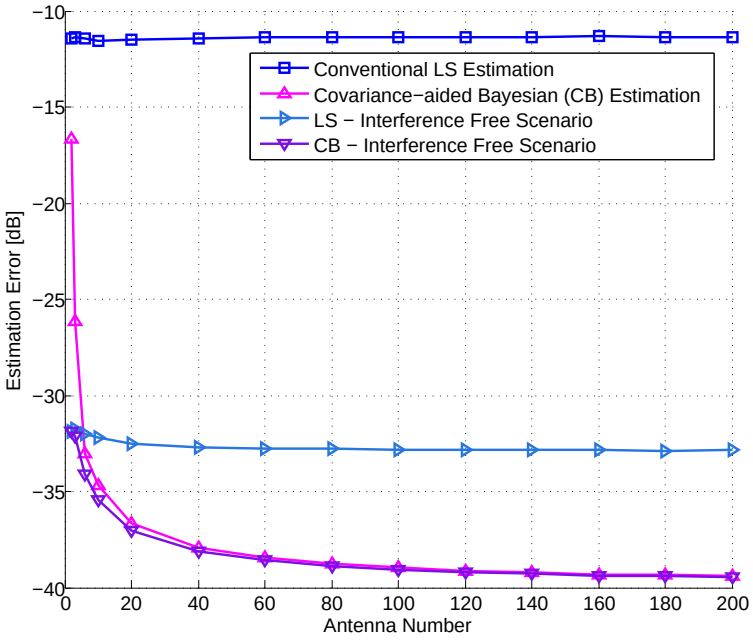


Figure 3.2: Estimation MSE vs. SP antenna number, 2-SP network, fixed positions of two users, uniformly distributed AOAs with $\theta_\Delta = 20$ degrees, non-overlapping multipath.

In Figure 3.2 we validate Theorem 3.2 with a 2-cell network, where the two users' positions are fixed. AOAs of desired channels are uniformly distributed with a mean of 90 degrees, and the angle spreads of all channels are 20 degrees, yielding no overlap between desired and interfering multipaths. The pilot contamination is quickly eliminated as the number of antennas at the SP increases.

In Figure 3.3 the per-SP sum rate is depicted as a function of the number of antennas at the SP, regarding a two-operator network and for two different cases, CPA (sharing case) and the Bayesian estimator (non-sharing case). Here, the pilot sequences are statistically correlated and the AOAs are uniformly distributed. It is evident that the sharing gain (SAPHYRE gain) increases when the number of the

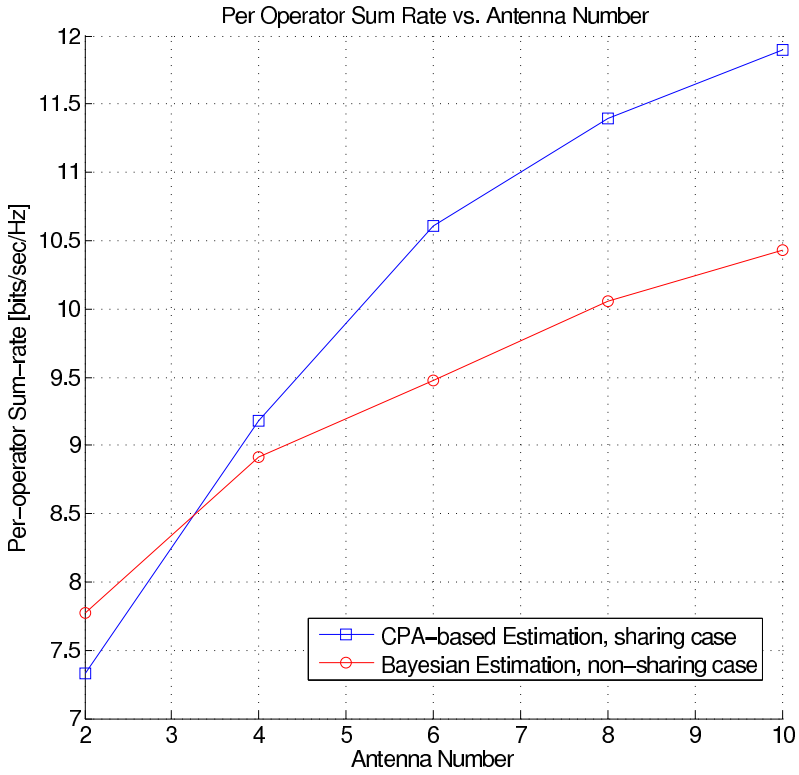


Figure 3.3: Per-operator sum-rate vs. antenna number, 2-operator network, statistically correlated pilot sequences, uniformly distributed AOAs with $\theta_\Delta = 10$ degrees - Bayesian estimator.

antennas at the SP gets larger.

As a conclusion, this work proposes a covariance-aided channel estimation framework in the context of interference-limited, multiple spectrum sharing operator, multiple antenna systems. In our scenario, we assumed the individual covariance matrices can be estimated separately. This could be done in practice by exploiting resource blocks where the desired user and interference users are known to be assigned at different times. In future networks, one may imagine a specific training design for learning second-order statistics.

3.3 Coordinated Scheduling and Beamforming for Spectrum Sharing Networks

In this section, we consider the downlink of a multicell network where neighboring multi-antenna base stations share the spectrum and coordinate their frequency and spatial resource allocation strategies to improve the overall network performance. The objective of the coordination is to maximize the number of users that can be scheduled, meeting their quality-of-service requirements with the minimum total

transmit power. The coordinated scheduling and multiuser transmit beamforming problem is combinatorial; we formulate it as a mixed-integer second-order cone program and propose a branch & bound algorithm that yields the optimal solution with relatively low-complexity. The algorithm can be used to motivate or benchmark approximation methods and to numerically evaluate the gains due to spectrum sharing and coordination.

3.3.1 System model

We consider a wireless network with L cells, where in each cell there is one base station (BS) and K mobile stations (MSs). The BS of the lth ($l \in \mathcal{L} \triangleq \{1, \cdots, L\}$) cell is denoted as BS_l and the kth ($k \in \mathcal{K} \triangleq \{1, \cdots, K\}$) MS in the lth cell is denoted as $\text{MS}_{l,k}$. Each BS has N_t antennas and each MS has a single antenna. Multiuser downlink beamforming is employed at each BS. The number of available subchannels is N (indexed by $n \in \mathcal{N} \triangleq \{1, \cdots, N\}$), which is assumed to be smaller than the total number of MSs in the network, i.e., $N < LK$. We assume that each MS can be scheduled in at most one subchannel. All the channel state information, i.e., from each BS to every MS in the network, is assumed to be known, and the channels are flat in each transmission interval.

We denote the beamforming vector for $\text{MS}_{l,k}$ in the nth subchannel as $\mathbf{w}_{l,k}^n \in \mathbb{C}^{N_t}$, and the channel from BS_j to $\text{MS}_{l,k}$ in the nth subchannel as $\mathbf{h}_{j,l,k}^n \in \mathbb{C}^{N_t}$. We have $\mathbf{w}_{l,k}^n \neq \mathbf{0}$ if $\text{MS}_{l,k}$ is scheduled in the nth subchannel, and $\mathbf{w}_{l,k}^n = \mathbf{0}$ otherwise. The signal-to-interference-plus-noise ratio (SINR) for $\text{MS}_{l,k}$ in the nth subchannel can then be expressed as

$$\Gamma_{l,k}^n = \frac{|(\mathbf{w}_{l,k}^n)^H \mathbf{h}_{l,l,k}^n|^2}{\sum\limits_{b \neq k} |(\mathbf{w}_{l,b}^n)^H \mathbf{h}_{l,l,k}^n|^2 + \sum\limits_{j \neq l} \sum\limits_{b} |(\mathbf{w}_{j,b}^n)^H \mathbf{h}_{j,l,k}^n|^2 + \sigma_{l,k}^2}, \tag{3.38}$$

where $\sigma_{l,k}^2$ is the noise power.

In order to satisfy the quality-of-service (QoS) requirement, the condition to schedule $\text{MS}_{l,k}$ in the nth subchannel is that the SINR $\Gamma_{l,k}^n$ should be no less than a target $\gamma_{l,k}$, i.e., $\Gamma_{l,k}^n \geq \gamma_{l,k}$. Here, a single SINR target is chosen for different subchannels. Thus a MS can be scheduled in the subchannel which has the best channel and interference condition. Moreover, we assume the total transmit power at each BS can not exceed a maximum budget P_l, i.e., $\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \|\mathbf{w}_{l,k}^n\|^2 \leq P_l$. This constraint allows the power to be distributed unevenly among the subchannels.

3.3.2 Optimization objectives: scheduling and beamforming

In our problem of interest, we have two objectives, namely scheduling and beamforming. Scheduling is the primary objective, which is to maximize the number of MSs that can be scheduled in the available subchannels while satisfying both the

SINR constraints and the BS power constraints. Beamforming is the secondary objective, which is to find the optimal beamforming vectors for those scheduled MSs that minimize the total transmit power.

We can describe this multi-objective optimization problem in two stages as in [26]. Let $\mathcal{S}_l^n \subseteq \mathcal{K}$ denote the subset of MSs in the lth cell that are scheduled in the nth subchannel, and $|\mathcal{S}_l^n|$ denote the cardinality of $\mathcal{S}_l^n$. The first stage is a combinatorial optimization problem given by

$$\max_{\mathcal{S}_l^n, \mathbf{w}_{l,k}^n} \sum_{n\in\mathcal{N}} \sum_{l\in\mathcal{L}} |\mathcal{S}_l^n| \tag{3.39a}$$

$$\text{s.t.} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{S}_l^n} \|\mathbf{w}_{l,k}^n\|^2 \le P_l, \quad \forall l \in \mathcal{L}, \tag{3.39b}$$

$$\Gamma_{l,k}^n \ge \gamma_{l,k}, \quad \forall n \in \mathcal{N}, \forall k \in \mathcal{S}_l^n, \forall l \in \mathcal{L}, \tag{3.39c}$$

$$\mathcal{S}_l^n \cap \mathcal{S}_l^m = \emptyset, \quad n \ne m, \forall n, m \in \mathcal{N}, \forall l \in \mathcal{L}, \tag{3.39d}$$

where constraint (3.39d) ensures that no MS can be scheduled in more than one subchannel.

With the optimal sets $\{\mathcal{S}_l^n\}$ obtained by solving (3.39), the second stage is a downlink beamforming problem,

$$\min_{\mathbf{w}_{l,k}^n} \sum_{n\in\mathcal{N}} \sum_{l\in\mathcal{L}} \sum_{k\in\mathcal{S}_l^n} \|\mathbf{w}_{l,k}^n\|^2, \tag{3.40a}$$

$$\text{s.t. (3.39b), (3.39c).} \tag{3.40b}$$

Problem (3.40) can be transformed into a SOCP problem [27] with complexity $\mathcal{O}(((\sum_n \sum_l |\mathcal{S}_l^n|)N_t)^{3.5})$ and solved efficiently by using general-purpose convex optimization toolboxes.

As an alternative to the two-stage formulation (3.39) and (3.40), we propose a joint formulation by introducing auxiliary binary variables $s_{l,k}^n \in \{0,1\}$. Let $s_{l,k}^n = 1$ if $\text{MS}_{l,k}$ is scheduled in the nth subchannel; and $s_{l,k}^n = 0$ otherwise. The joint formulation is given by

$$\min_{\mathbf{w}_{l,k}^n, s_{l,k}^n} \epsilon \sum_{n\in\mathcal{N}} \sum_{l\in\mathcal{L}} \sum_{k\in\mathcal{K}} \|\mathbf{w}_{l,k}^n\|^2 - \sum_{n\in\mathcal{N}} \sum_{l\in\mathcal{L}} \sum_{k\in\mathcal{K}} s_{l,k}^n \tag{3.41a}$$

$$\text{s.t.} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} \|\mathbf{w}_{l,k}^n\|^2 \le P_l, \qquad \forall l \in \mathcal{L}, \tag{3.41b}$$

$$\frac{|(\mathbf{w}_{l,k}^n)^H \mathbf{h}_{l,l,k}^n|^2 + M_{l,k}^n(1 - s_{l,k}^n)}{\sum_{b\ne k} |(\mathbf{w}_{l,b}^n)^H \mathbf{h}_{l,l,k}^n|^2 + \sum_{j\ne l} \sum_b |(\mathbf{w}_{j,b}^n)^H \mathbf{h}_{j,l,k}^n|^2 + \sigma_{l,k}^2} \ge \gamma_{l,k},$$

$$\forall n \in \mathcal{N}, \forall l \in \mathcal{L}, \forall k \in \mathcal{K}, \tag{3.41c}$$

$$s_{l,k}^n \in \{0,1\}, \qquad \forall n \in \mathcal{N}, \forall l \in \mathcal{L}, \forall k \in \mathcal{K}, \tag{3.41d}$$

$$\sum_{n\in\mathcal{N}} s_{l,k}^n \le 1, \qquad \forall l \in \mathcal{L}, \forall k \in \mathcal{K}. \tag{3.41e}$$

In the cost function (3.41a), the first term is the total transmit power scaled by a positive constant ϵ and the second term counts the total number of admitted MSs. Since the total transmit power is bounded by $\sum_{l\in\mathcal{L}} P_l$ and the second term is discrete with step size -1, we choose $0 < \epsilon < 1/(\sum_{l\in\mathcal{L}} P_l)$. This choice of ϵ implies that the maximum possible number of MSs will be scheduled and no other solution that schedules the same number of MSs can operate with less power [26]. Constraint (3.41c) defines N inequalities for each $\text{MS}_{l,k}$. When $s^n_{l,k} = 1$, the inequality is a standard SINR constraint; when $s^n_{l,k} = 0$, the inequality does not impose any constraint on $\{\mathbf{w}^n_{l,k}\}$ provided that $M^n_{l,k}$ is large enough to satisfy the inequality for all possible values of $\{\mathbf{w}^n_{l,k}\}$. By considering the power constraint (3.41b) and the Cauchy-Schwarz inequality, we choose the value of $M^n_{l,k}$ as $M^n_{l,k} \geq \gamma_{l,k} \sum_{j\in\mathcal{L}} P_j \|\mathbf{h}^n_{j,l,k}\|^2 + \gamma_{l,k}\sigma^2_{l,k}$. Constraint (3.41e) makes sure that each MS is scheduled in at most one subchannel. Note that, when all the binary variables $\{s^n_{l,k}\}$ are fixed, (3.41) is equivalent to (3.40) in the two-stage formulation.

Constraint (3.41c) for each n, l, k can be formulated as a SOCP constraint,

$$\left\|[(\mathbf{u}^n_{l,k})^T,\ \sigma_{l,k}]\right\| \leq \sqrt{(1+\frac{1}{\gamma_{l,k}})}(\mathbf{w}^n_{l,k})^H\mathbf{h}^n_{l,l,k} + \sqrt{\frac{M^n_{l,k}}{\gamma_{l,k}}}(1-s^n_{l,k}), \tag{3.42}$$

where $\mathbf{u}^n_{l,k}$ is a $LK \times 1$ vector defined as $\mathbf{u}^n_{l,k} = [(\mathbf{w}^n_{1,1})^H\mathbf{h}^n_{1,l,k}, \cdots, (\mathbf{w}^n_{L,K})^H\mathbf{h}^n_{L,l,k}]^T$, and $(\mathbf{w}^n_{l,k})^H\mathbf{h}^n_{l,l,k}$ is constrained to be real valued and positive. By replacing (3.41c) with (3.42), we can transform (3.41) into a MI-SOCP problem, which can be solved, e.g., by using the B&B solver CPLEX. The procedure of using CPLEX to solve this MI-SOCP problem is based on the relaxation of the binary variables. Specifically, constraint (3.41d) is relaxed as $0 \leq s^n_{l,k} \leq 1$. Thus the MI-SOCP problem changes into a SOCP problem of complexity $\mathcal{O}((LKNN_t)^{3.5})$. However, the lower bound found from relaxation in this problem is quite loose. This is because the chosen values of constants $\{M^n_{l,k}\}$ are much larger than the sum of interference and noise in (3.41c). Therefore, every constraint in (3.41c) can be satisfied with $\{s^n_{l,k}\}$ of such values that the equality in (3.41e) becomes valid for every MS, i.e., $\sum_{n\in\mathcal{N}} s^n_{l,k} = 1$. Thus the second term in the objective function (3.41a) will be LK, which seems like all the LK MSs are scheduled in the available subchannels, but in fact it is not. In the next section, we propose a customized B&B algorithm which avoids the relaxation. Our algorithm has tighter lower bound and also finds an upper bound with low complexity.

3.3.3 Branch and bound algorithm

In our B&B algorithm, we split the problem in (3.41) (i.e., the root) into subproblems (nodes) by fixing a subset of the binary variables. We define a $N \times 1$ binary vector $\mathbf{s}_i$ ($i \in \{1, \cdots, LK\}$) for each MS, for example, $\mathbf{s}_i = [s^1_{j,b}, \cdots, s^N_{j,b}]^T$ ($\forall j \in \mathcal{L}$, $\forall b \in \mathcal{K}$). Because of the constraint (3.41e), $\mathbf{s}_i$ can be either a column of the identity matrix $\mathbf{I}_{N\times N}$ or an all-zero vector, so it has $N+1$ possible combinations.

We can generate a tree with LK levels, where each level corresponds to a specified MS. The original problem (3.41) can be split into $N+1$ nodes in the first level by fixing $\mathbf{s}_1$. Each of those nodes can be further split in the second level by further fixing $\mathbf{s}_2$, etc, thus generating $(N+1)^{LK}$ nodes in the last level. Solving the SOCP problem for each and every node at the last level corresponds to the enumeration method which has prohibitive complexity. For this reason, we would like to prune nodes in the tree early on without going all the way down to the last level and we show how to achieve this in the following.

For a node in the tree, we calculate a lower bound (LB) and an upper bound (UB) for the optimal value of (3.41a). If the LB of a node is higher than the global upper bound (GUB), i.e., the tightest UB from all nodes already examined, this node and all its children nodes can be safely pruned without loss of optimality. This is because all children nodes are further restrictions of their parent node (each child node has one more binary vector fixed relative to its parent node), implying that the LB of a child node must be greater than or equal to the LB of its parent node. This implicit elimination is the key to computational savings, and it can be very effective if substantial pruning happens early in the process.

In order to make the algorithm more efficient, we need to define a proper *order* to fix the binary vectors of the MSs, i.e., for which MS it is fixed in the first level, which is fixed in the second level, etc. Since our primary objective is to maximize the number of scheduled MSs within the limited total transmit power and satisfying the individual SINR constraints, we have an intuition that the MS requiring small transmit power to satisfy its SINR constraint is more likely to be admitted, so we fix the binary vector of such a MS in a early level. Although a SOCP has to be solved to determine the required transmit power $\|\mathbf{w}_{l,k}^n\|^2$ satisfying $\Gamma_{l,k}^n \geq \gamma_{l,k}$, we can find a LB for it by considering the interference-free case. The minimum transmit power satisfying the QoS constraint can be find through the maximum ratio transmission (MRT) as $p_{l,k}^n = \gamma_{l,k}\sigma_{l,k}^2/\|\mathbf{h}_{l,l,k}^n\|^2$ and the related beamforming vector is $\mathbf{v}_{l,k}^n = \sqrt{p_{l,k}^n}\mathbf{h}_{l,l,k}^n/\|\mathbf{h}_{l,l,k}^n\|$. We refer to $p_{l,k}^n$ and $\mathbf{v}_{l,k}^n$ as MRT power and MRT beamformer, respectively. We calculate every $p_{l,k}^n$ to get a $N \times LK$ matrix $\mathbf{P} = [\mathbf{p}_{1,1}, \cdots, \mathbf{p}_{1,K}, \cdots, \mathbf{p}_{L,1}, \cdots, \mathbf{p}_{L,K}]$, where $\mathbf{p}_{l,k} = [p_{l,k}^1, \cdots, p_{l,k}^N]^T$. Then we find the minimum element in each column of $\mathbf{P}$ and sort them into a $LK \times 1$ vector in an ascending order. This vector gives the *order* that we need, i.e., the MS corresponds to the i element of this vector will be fixed in the ith level.

Then we introduce how to calculate the UB. For a specified node in the ith level, we have a subset of i fixed binary vectors $\mathbf{s}_1 \cdots \mathbf{s}_i$. By assuming all the other $\mathbf{s}_{i+1} \cdots \mathbf{s}_{LK}$ to be all-zero vectors (i.e., not scheduled in any subchannel), we get a fixed set of binary variables $\{s_{l,k}^n\}$, and the node corresponds to a SOCP problem (3.40). The complexity of this SOCP problem is $\mathcal{O}((iN_t)^{3.5})$, since there are only i SINR constraints (3.39c). However, if using the relaxation method as in CPLEX, the same node corresponds to a SOCP problem with complexity $\mathcal{O}(((i+N(LK-i))N_t)^{3.5})$, because of the additional $N(LK-i)$ constraints (3.41c) and variables. If the SOCP

problem of the node is infeasible, this node can be pruned directly. Otherwise, we get a solution which can be used as a *UB* for this node. However, this *UB* is a *loose* one because we have assumed all the other unfixed MSs are not scheduled. We can tighten this *UB* by scheduling more MSs while keeping those already scheduled ones. We implement this by extending the low-complexity admission-control method in [28] into our multi-subchannel model. In each subchannel, we keep the spatial signatures of beamformers for the already admitted MSs, but adjust their power (under the power limit) to schedule a new MS, and the beamforming vector for the new MS is calculated while satisfying all the constraints. This process is repeated until no more MS can be scheduled. In this way, we find a relatively tight *UB* avoiding to solve additional SOCPs.

For the *LB*, we keep the solution from the *loose UB*. Since the beamforming vectors for those MSs with the fixed binary vectors $\mathbf{s}_1 \cdots \mathbf{s}_i$ were calculated, we know how much power was spent in each cell. With the remaining power budget, we try to schedule as many other MSs as possible by considering their MRT power under the power constraint, i.e., assuming the interference-free case for the other MSs. Since the MRT power is the least power required to admit a MS, we get a *LB* for the node. This *LB* is tighter than that of the relaxation method.

We also find an initial *LB* and an initial *UB* before splitting the original problem into subproblems. We calculate the initial *LB* by minimizing the objective function (3.41a) under the constraints (3.41b), (3.41d) and (3.41e) while assuming $\|\mathbf{w}_{l,k}^n\|^2 = p_{l,k}^n$. For the initial *UB*, we first schedule one MS in each subchannel, which corresponds to the smallest element in each row of the MRT power matrix $\mathbf{P}$. Then we schedule more MSs with the same method as we do in tightening the *UB*.

Our proposed optimal algorithm using B&B is summarized in Algorithm 2. We denote the node that selected to be split as *s-node* and we use a *stack* to keep track of nodes that require further examination. In step 1, if the calculated initial *LB* and initial *UB* are equal, we terminate the algorithm, and the optimal solution to the problem is obtained from the initial *UB*. Otherwise, we initialize *s-node* to be the root and *stack* to be empty, set *GUB* to be equal to the initial *UB* and go to step 2. In step 2, we implement the depth-first search. Specifically, a node is split into $(N + 1)$ children nodes, but only the one with the lowest *LB* is further split into the next level, while any other unpruned nodes are inserted in the stack. We repeat step 2 and 3 until the *stack* becomes empty. The final optimal solution is obtained from the *GUB*.

For large size problems, the complexity of Algorithm 1 can be very high because a large number of nodes might be generated in step 2 and 3. Therefore, we can find a sub-optimal solution by implementing a fixed number of depth-first searches in Algorithm 1. If Q searches are implemented, the maximum number of nodes generated is $(N + 1)LKQ$.

Algorithm 2 Proposed optimal algorithm using B&B

1. Calculate an initial LB and an initial UB.
   - If they are equal, terminate; else, initialize *s-node* $\leftarrow$ root, *stack* $\leftarrow \emptyset$, $GUB \leftarrow$ initial UB, and go to 2.
2. Implement the depth-first search one time.
   (a) Split *s-node* into $N+1$ new nodes, and for every new node, solve a SOCP problem (3.40).
      - If it is infeasible, prune the node; else, calculate LB.
      - If $LB > GUB$, prune the node; else, tighten UB.
      - If $UB < GUB$, $GUB \leftarrow UB$.
   (b) *s-node* $\leftarrow$ the one has the lowest LB in the unpruned new nodes, insert the other ones in *stack*, go to (a).
   (c) Repeat (a) (b) until all new nodes are pruned or the last level is reached.
      - If *stack* $= \emptyset$, terminate; else, go to 3.
3. Delete the nodes whose $LB > GUB$ in *stack*.
   - If *stack* $= \emptyset$, terminate; else, *s-node* $\leftarrow$ the node with the lowest LB in *stack*, and go to 2.

We give an example to illustrate Algorithm 1 in Figure 3.4, where 2 cells with 2 MSs in each cell and 2 subchannels are considered. The branching *order* is that, the binary vector for $\mathrm{MS}_{1,1}$ is fixed in level 1, followed by $\mathrm{MS}_{1,2}$ in level 2, $\mathrm{MS}_{2,2}$ in level 3 and $\mathrm{MS}_{2,1}$ in level 4. The nodes are numbered in the same order as they are generated. Since the initial UB and the initial LB are not equal to each other, we split the root into three nodes in level 1 by fixing $\mathbf{s}_1$. Then node 1 and 3 are pruned because their LB are higher than GUB and we do not need to tighten their UB. Since node 2 is the only node left, it is split in level 2 where $\mathbf{s}_2$ is further fixed. In level 2, node 5 and 6 are pruned and node 4 is split into level 3. In level 3, GUB is updated to be equal to the UB of node 7, node 9 is pruned, node 8 is split into level 4 and node 7 is inserted in *stack*. In level 4, nodes 10, 11, and 12 are all pruned since their LB are higher than the updated GUB, and therefore node 8 is also pruned. Next, take node 7 out of *stack* and split it. Then nodes 13 and 15 are pruned. Now all the nodes in the tree, except node 14 and its parents (nodes 7, 4, and 2), have been pruned and *stack* becomes empty. Therefore, the optimal binary vectors in this example are $\mathbf{s}_1 = [1\ 0]^T$, $\mathbf{s}_2 = [0\ 1]^T$, $\mathbf{s}_3 = [0\ 1]^T$ and $\mathbf{s}_4 = [1\ 0]^T$, and the optimal beamforming vectors are also obtained when calculating the GUB. In this example, we find the optimal solution by generating 15 nodes, while by using the relaxation method with CPLEX there are 26 nodes generated, and by brute-force searching we have to solve all 3^4 nodes. For larger size problems, we can even save more complexity.

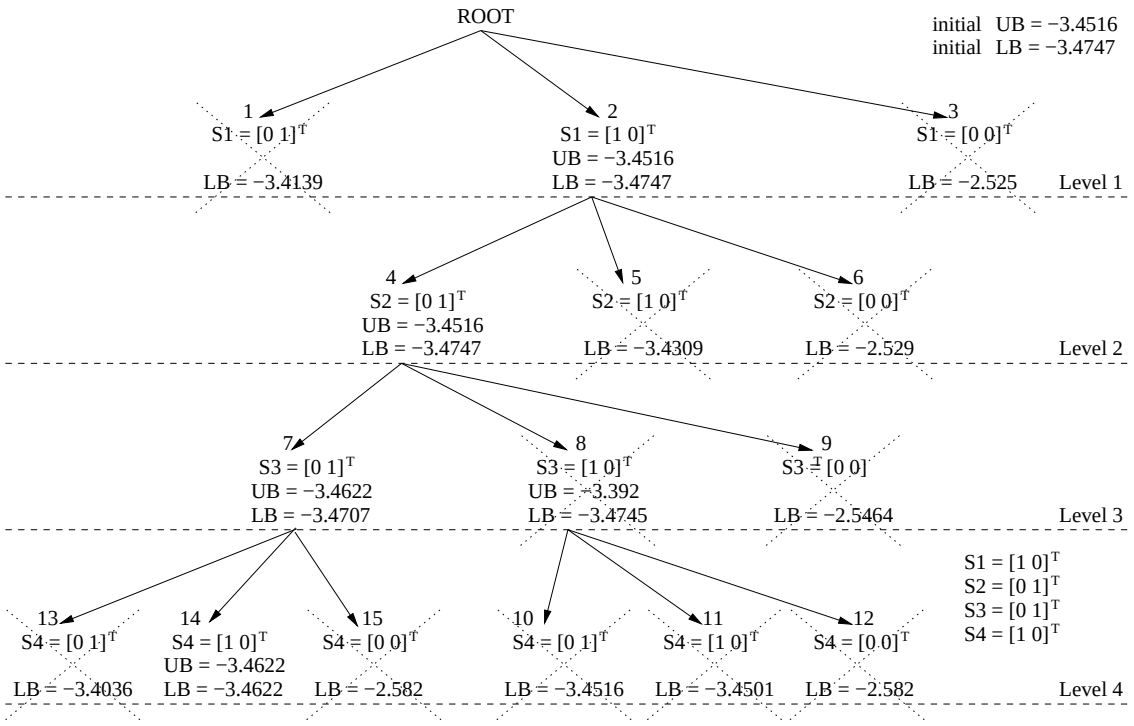


Figure 3.4: An example for the B&B algorithm

3.3.4 Simulations

In our simulations, we consider the overlapping multicell networks from two operators. As shown in Figure 3.5, two strategies for the BSs deployment are involved, namely cositing and non-cositing. In the cositing strategy, the BSs of the two operators share the same mast or located very closely; whereas in the non-cositing strategy, the BSs of one operator are located apart from those of the other. The network of each operator has 7 hexagonal sectors and each sector has one BS. An inter-site distance of 500 meters is considered. Each operator has K MSs randomly located within the center sector. The number of available subchannels is N. The radio channel traces are generated using the quasi-deterministic radio channel generator (QuaDRiGa) [29] recently released by Fraunhofer Heinrich Hertz Institute. The WINNER+ model [30] is used with the major parameters summarized in TABLE 3.2.

We consider two spectrum allocation scenarios, i.e., sharing and non-sharing. In the sharing scenario, the N subchannels are shared by both operators; while in the non-sharing scenario, each operator exclusively uses $N/2$ subchannels. Moreover, we assume that only the two center sectors (each from one operator) are coordinated, whereas all the surrounding sectors are merely included to establish a more realistic interference environment. Thus the interference from the surrounding sectors will

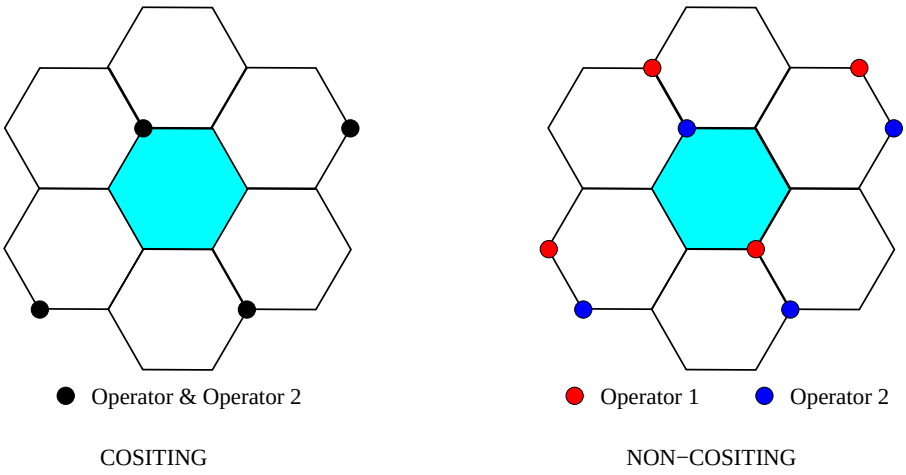


Figure 3.5: Overlapping multicell networks from two operators

Parameter name	**Parameter value**
According to WINNER+ scenario	C2 NLOS
Carrier frequency (GHz)	2.6
Path loss (dB)	$35.05\log(d) + 36.70$
Intra / inter-site correlation	1 / 0.5
Number of BSs	14
Number of antennas BSs / MSs	4 / 1
Heights of BSs / MSs (m)	32 / 2
Antenna pattern of BSs	KATHREIN 80010541 (10 degrees down tilt)
Antenna pattern of MSs	omni-directional

Table 3.2: Propagation parameters

not be mitigated during the optimal scheduling and beamforming.

We assume the two coordinated sectors have the same power budget, i.e., $P_l = P$, $(l = 1, 2)$, in both the sharing and non-sharing scenarios. The transmit power in each of the surrounding sectors is assumed to be αP, where α is an activity factor that determines the interference strength from the surrounding sectors. Moreover, the power αP is evenly distributed among the available subchannels. Therefore, the transmit power at each subchannel in each surrounding sector is $\alpha P/N$ for the sharing scenario and $2\alpha P/N$ for the non-sharing scenario. Note that the number of interfering sectors is different for the sharing and non-sharing scenario. In the sharing scenario, all 12 surrounding sectors from both operators generate interference; while in the non-sharing scenario only 6 surrounding sectors from the same operator generate interference. The interference from jth interfering sector to kth MS of ith operator in nth subchannel is $(\alpha P/N)|(\mathbf{w}_j^n)^H \mathbf{h}_{j,i,k}^n|^2$ and $(2\alpha P/N)|(\mathbf{w}_j^n)^H \mathbf{h}_{j,i,k}^n|^2$ for the sharing and non-sharing scenario, respectively. In the calculations, we use a random transmit beamforming vector $\mathbf{w}_j^n$, which has uniform norm (i.e., $\|\mathbf{w}_j^n\| = 1$), for the jth surrounding sector and nth subchannel.

We assume the SINR targets for all the MSs are the same, i.e., $\gamma_{l,k} = \gamma$. The power budget for the two coordinated sectors is 0.4 W per subchannel, i.e., $P = 0.4N$ W. By using the proposed B&B algorithm, we find the total number of scheduled MSs of the two operators in the sharing and non-sharing scenarios for both the cositing and non-cositing networks. Assume the total number of scheduled MSs is K', the throughput (i.e., sum rate of the K' MSs), can be calculated as $K' \log_2(1+\gamma)$, where γ is the SINR target. We define the SAPHYRE gain to be the ratio of throughput between the sharing scenario and the non-sharing scenario. In the following, we show a small-scale example and a large-scale example, where different number of N and K are considered. In each example, 10 realizations of the MSs positions are simulated and the results are averaged.

In the small-scale example, we assume only two subchannels are available and each operator has 10 MSs, i.e., $N = 2, K = 10$. The power budget for the two coordinated sectors is 0.8 W. In this example, for the sharing scenario, we find a suboptimal solution by implementing 5 times depth-first search in Algorithm 2; while for the non-sharing scenario, we find the optimal solution (due to the low complexity with small values of N and K). Regarding to the interference from surrounding sectors, we consider three cases: (a) no interference from surrounding sectors, i.e., $\alpha = 0$; (b) relatively weak interference with $\alpha = 0.1$; (c) relatively strong interference with $\alpha = 0.5$. The average number of scheduled MSs and average throughput for various SINR targets for the three cases are shown in Figure 3.6 - 3.11.

We first look at the case of $\alpha = 0$ in Figure 3.6 and 3.7. For the non-sharing scenario, almost the same number of MSs are scheduled in both the cositing and non-cositing network. The number of scheduled MSs in the non-sharing scenario is 8 when SINR > 8 dB. This is because each BS has 4 antenna and a maximum number of 4 MSs can be served in each subchannel with large SINR. For the sharing scenario, a larger

number of MSs can be scheduled in both networks and the numbers decrease as SINR increases. Moreover, the number of scheduled MSs in non-cositing network is larger than that in the cositing network. We find the maximum SAPHYRE gain is more than 1.5 for the non-cositing network and about 1.25 for the cositing network, both of which are achieved at SINR = 8 dB. As the SINR increases, the gain becomes smaller. At SINR = 20 dB, the SAPHYRE gain becomes about 1.25 for the non-cositing network, and it becomes 1 (i.e., no gain) for the cositing network.

The results for the case of $\alpha = 0.1$ are given in Figure 3.8 and 3.9. For the non-sharing scenario of both networks, the number of scheduled MSs does not keep at 8, but decreases when the SINR becomes larger and larger due to the extra interference from the surrounding sectors. For the sharing scenario, a larger number of MSs are scheduled in the non-cositing network than in the cositing network, and for both networks the number decreases as SINR increases. Again, in this case the SAPHYRE gain in the non-cositing network is higher than that in the cositing network.

For the case of $\alpha = 0.5$ in Figure 3.10 and 3.11, due to the relatively high interference from the surrounding sectors, the number of scheduled MSs for both networks and both scenarios drops quickly as the SINR increases. Moreover, the SAPHYRE gain becomes smaller in both networks compared with the cases of $\alpha = 0$ and $\alpha = 0.1$.

Next, we consider a large-scale example where 10 subchannels are available and each operator has 50 MSs, i.e., $N = 10, K = 50$. In this example, the power budget for the two coordinated sectors is 4 W, and we only consider the case of small interference from surrounding sectors, i.e., $\alpha = 0.1$. Moreover, due to the large value of N and K, for both the sharing and non-sharing scenario we find a suboptimal solution by implementing only one time depth-first search in Algorithm 2. The results are given in Figure 3.12 and 3.13, where we find the curves have the similar trend as those in Figure 3.8 and 3.9. In the non-sharing scenario, the numbers of scheduled MSs are almost the same in both networks. Specifically, with SINR between 6 and 16 dB, 40 MSs are scheduled, and the numbers decrease by further increasing the SINR. The maximum SAPHYRE gain is larger than 1.5 for the sharing scenario and it is larger than 1.25 for the non-sharing scenario, both of which are achieved at SINR = 6 dB. As SINR increases, the SAPHYRE gain becomes smaller in both networks.

In conclusion, we have found that a larger SAPHYRE gain can be achieved in the non-cositing network than in the cositing network. In general, the gain becomes smaller as the SINR targets increases. In our simulations, the maximum SAPHYRE gain achieved is about 1.5 in the non-cositing network and about 1.25 in the cositing network. This achievement is under the condition that the size of the user pool is 5 MSs per subchannel, and a suboptimal solution is found for the sharing scenario. We can expect a greater gain, if we have a larger size pool of MSs to choose from, or the optimal solution is found.

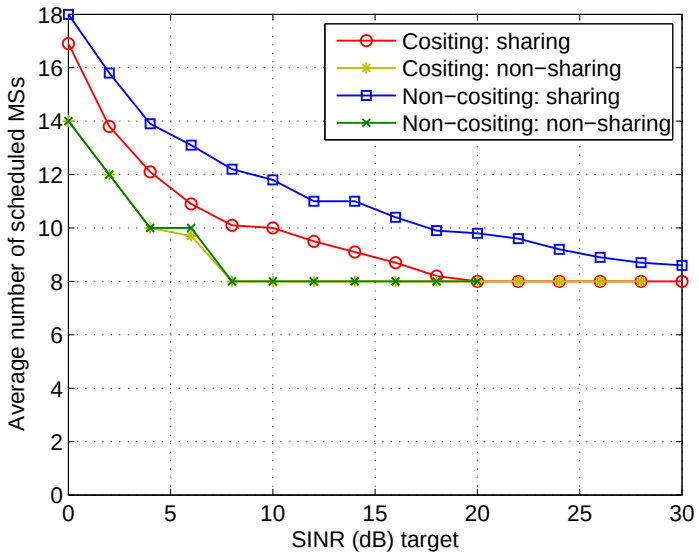


Figure 3.6: Average number of scheduled MSs with $N = 2, K = 10, \alpha = 0$

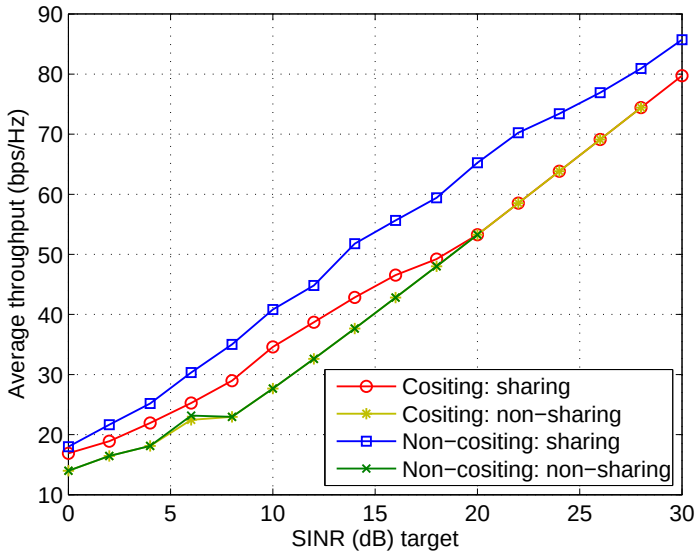


Figure 3.7: Average throughput with $N = 2, K = 10, \alpha = 0$

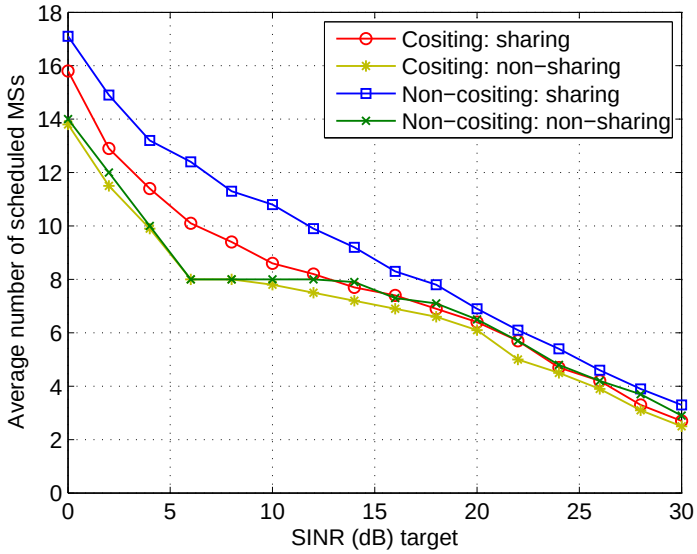


Figure 3.8: Average number of scheduled MSs with $N = 2, K = 10, \alpha = 0.1$

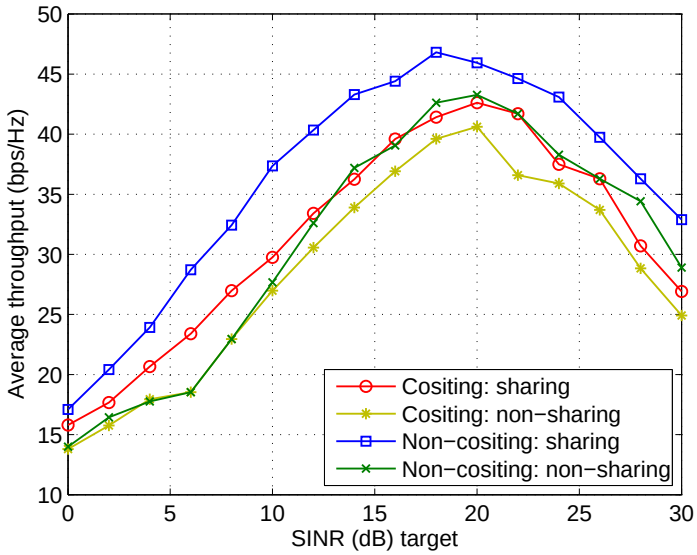


Figure 3.9: Average throughput with $N = 10, K = 50, \alpha = 0.1$

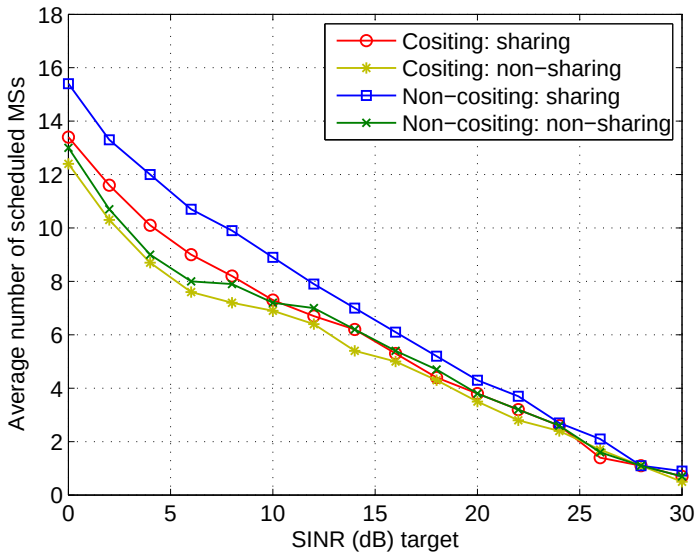


Figure 3.10: Average number of scheduled MSs with $N = 2, K = 10, \alpha = 0.5$

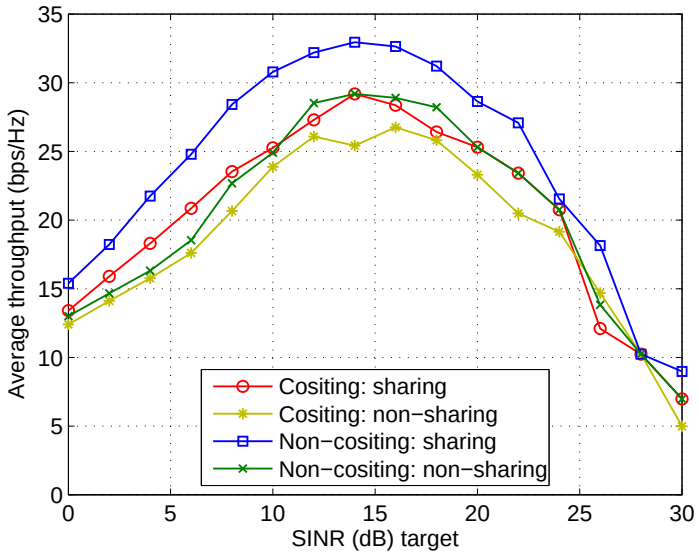


Figure 3.11: Average throughput with $N = 2, K = 10, \alpha = 0.5$

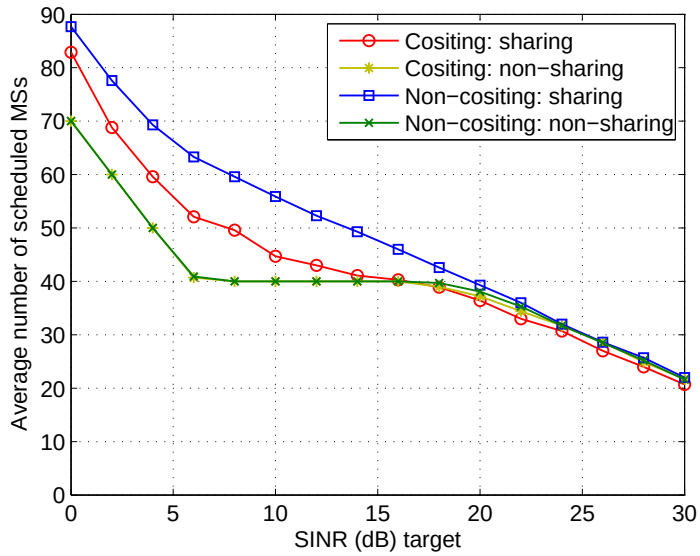


Figure 3.12: Average number of scheduled MSs with $N = 10, K = 50, \alpha = 0.1$

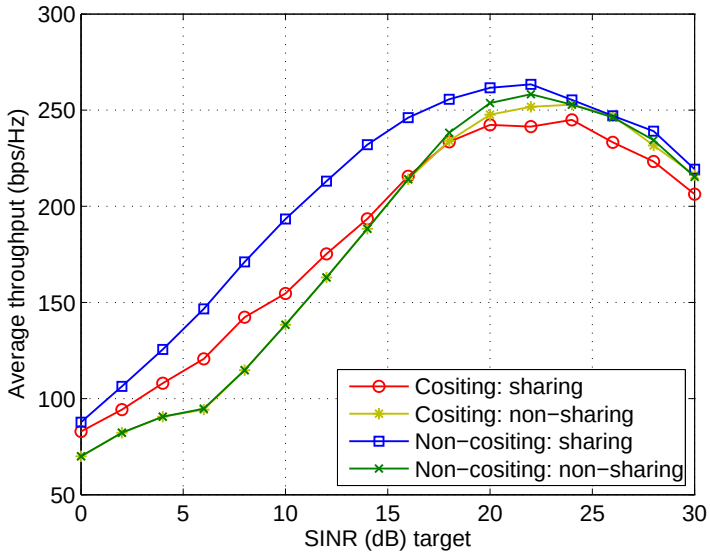


Figure 3.13: Average throughput with $N = 10, K = 50, \alpha = 0.1$

4 Distributed MIMO Signal Processing and Resource Allocation

4.1 Cooperative Beamforming

In this section, a distributed beamforming algorithm is proposed for the two-user multiple-input single-output (MISO) interference channel (IC). The algorithm is iterative and uses as bargaining value the interference that each transmitter generates towards the receiver of the other user. It enables cooperation among the transmitters in order to increase both users' rates by lowering the overall interference. In every iteration, as long as both rates keep on increasing, the transmitters mutually decrease the generated interference. They choose their beamforming vectors distributively, solving the constrained optimization problem of maximizing the useful signal power for a given level of generated interference. The algorithm is equally applicable when the transmitters have either instantaneous or statistical channel state information (CSI). The difference is that the core optimization problem is solved in closed-form for instantaneous CSI, whereas for statistical CSI an efficient solution is found numerically via semidefinite programming. The outcome of the proposed algorithm is approximately Pareto-optimal. Extensive numerical illustrations are provided, comparing the proposed solution to the Nash equilibrium, zero-forcing, Nash bargaining, and maximum sum-rate operating points.

4.1.1 System model and preliminaries

We assume that transmission consists of scalar coding followed by beamforming[1] and that all propagation channels are frequency-flat. The matched-filtered symbol-sampled complex baseband data received by RX_i is modeled as[2]

$$y_i = \mathbf{h}_{ii}^H \mathbf{w}_i s_i + \mathbf{h}_{ji}^H \mathbf{w}_j s_j + e_i, \;\; j \neq i, \;\; i,j \in \{1,2\}, \tag{4.1}$$

where $s_i \sim \mathcal{CN}(0,1)$ and $\mathbf{w}_i \in \mathbb{C}^n$ are the transmitted symbol and the beamforming vector, respectively, employed by TX_i. Also, $e_i \sim \mathcal{CN}(0,\sigma_i^2)$ models the receiver noise. The (conjugated) channel vector between TX_i and RX_j is modeled as $\mathbf{h}_{ij} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{ij})$. We denote $r_{ij} \triangleq \mathrm{rank}\{\mathbf{Q}_{ij}\}$. In the case of instantaneous CSI, TX_i accurately knows the channel realizations $\mathbf{h}_{ii}$ and $\mathbf{h}_{ij}$, whereas for statistical CSI it only knows the channel covariance matrices $\mathbf{Q}_{ii}$ and $\mathbf{Q}_{ij}$.

[1] This is optimal in the case of instantaneous CSI, but not necessarily for statistical CSI, see [31].

[2] Whenever an expression is valid for both systems, it is denoted once with respect to system i and the index $j \neq i$ refers to the other system.

The transmission power is bounded due to regulatory and hardware constraints, such as battery and amplifiers. Without loss of generality, we set this bound to 1. Hence, the set of feasible beamforming vectors is

$$\mathcal{W} \triangleq \{\mathbf{w} \in \mathbb{C}^n \mid \|\mathbf{w}\|^2 \leq 1\}. \tag{4.2}$$

Note that the set $\mathcal{W}$ is convex. In what follows, a specific choice of $\mathbf{w}_i \in \mathcal{W}$ is denoted as a *transmit strategy* of TX_i.

When the transmitters perfectly know the channel vectors and the receivers treat interference as noise, the achievable *instantaneous* rate (in bits/channel use) for link i is [32]

$$R_i(\mathbf{w}_i, \mathbf{w}_j) = \log_2\left(1 + \frac{|\mathbf{h}_{ii}^H \mathbf{w}_i|^2}{|\mathbf{h}_{ji}^H \mathbf{w}_j|^2 + \sigma_i^2}\right). \tag{4.3}$$

It is evident that the rate on each link depends on the choice of both beamforming vectors. We define the power that RX_i receives from TX_j as

$$p_{ji}(\mathbf{w}_j) \triangleq |\mathbf{h}_{ji}^H \mathbf{w}_j|^2 = \mathbf{w}_j^H \mathbf{h}_{ji} \mathbf{h}_{ji}^H \mathbf{w}_j. \tag{4.4}$$

Then, we can write (4.3) as

$$R_i(\mathbf{w}_i, \mathbf{w}_j) = \log_2\left(1 + \frac{p_{ii}(\mathbf{w}_i)}{p_{ji}(\mathbf{w}_j) + \sigma_i^2}\right), \tag{4.5}$$

which is monotonously increasing with the useful signal power $p_{ii}(\mathbf{w}_i)$ for fixed received interference power $p_{ji}(\mathbf{w}_j)$ and monotonously decreasing with $p_{ji}(\mathbf{w}_j)$ for fixed $p_{ii}(\mathbf{w}_i)$.

The main goal of the bargaining algorithm we introduce in Section 4.1.2 is to agree on a PO solution. Hence, we restrict our attention to the beamforming vectors which are candidates to achieve PO points. From [32], we know that the PO beamforming vectors use full power and that they are linear combinations of the MR and ZF strategies

$$\mathbf{w}_i^{\mathrm{PO}}(\lambda_i) = \frac{\lambda_i \mathbf{w}_i^{\mathrm{MR}} + (1-\lambda_i)\mathbf{w}_i^{\mathrm{ZF}}}{\|\lambda_i \mathbf{w}_i^{\mathrm{MR}} + (1-\lambda_i)\mathbf{w}_i^{\mathrm{ZF}}\|} \tag{4.6}$$

for $\lambda_i \in [0, 1]$, where

$$\mathbf{w}_i^{\mathrm{MR}} = \frac{\mathbf{h}_{ii}}{\|\mathbf{h}_{ii}\|} \quad \text{and} \quad \mathbf{w}_i^{\mathrm{ZF}} = \frac{\mathbf{\Pi}_{\mathbf{h}_{ij}}^{\perp} \mathbf{h}_{ii}}{\left\|\mathbf{\Pi}_{\mathbf{h}_{ij}}^{\perp} \mathbf{h}_{ii}\right\|}. \tag{4.7}$$

The outcome when both transmitters use their MR strategies is the NE. When both use their ZF we refer to the ZF point.

When the transmitters only have statistical knowledge of the channels, it is natural to design the achievable the beamforming vectors with respect to the *ergodic* rates,

which are obtained by averaging over the channel realizations. From [33], we have[3]

$$\begin{aligned} R_i(\mathbf{w}_i, \mathbf{w}_j) &\triangleq \mathbb{E}\left\{ \log_2 \left(1 + \frac{|\mathbf{h}_{ii}^H \mathbf{w}_i|^2}{|\mathbf{h}_{ji}^H \mathbf{w}_j|^2 + \sigma_i^2} \right) \right\} \\ &= \frac{p_{ii}(\mathbf{w}_i)}{\ln 2} \frac{f_i(p_{ii}(\mathbf{w}_i)) - f_i(p_{ji}(\mathbf{w}_j))}{p_{ii}(\mathbf{w}_i) - p_{ji}(\mathbf{w}_j)}, \end{aligned} \tag{4.8}$$

where

$$f_i(x) \triangleq e^{\sigma_i^2/x} \int_{\sigma_i^2/x}^{\infty} \frac{e^{-t}}{t} dt. \tag{4.9}$$

In (4.8), $p_{ji}(\mathbf{w}_j)$ denotes the *average* power that RX_i receives from TX_j

$$p_{ji}(\mathbf{w}_j) = \mathbb{E}\{ \mathbf{w}_j^H \mathbf{h}_{ji} \mathbf{h}_{ji}^H \mathbf{w}_j \} = \mathbf{w}_j^H \mathbf{Q}_{ji} \mathbf{w}_j. \tag{4.10}$$

Note that the final terms in both (4.4) and (4.10) are convex homogeneous quadratics. The difference is that the parameter (channel) matrix in (4.4) is rank-1 by definition, whereas in (4.10) it can have any rank.

The ergodic rate (4.8) has the same behavior as the instantaneous rate (4.3), i.e., it is monotonously increasing (decreasing) with $p_{ii}(\mathbf{w}_i)$ ($p_{ji}(\mathbf{w}_j)$) for fixed $p_{ii}(\mathbf{w}_i)$ ($p_{ji}(\mathbf{w}_j)$) [33]. Also, for points on the Pareto boundary we know that $\mathbf{w}_i \in \mathcal{R}\{\mathbf{Q}_{ii}, \mathbf{Q}_{ij}\}$ [31]. The MR strategy $\mathbf{w}_i^{\mathrm{MR}}$ is the dominant eigenvector of $\mathbf{Q}_{ii}$ [33]. When $\mathcal{R}\{\mathbf{Q}_{ii}\} \nsubseteq \mathcal{R}\{\mathbf{Q}_{ij}\}$, the ZF strategy $\mathbf{w}_i^{\mathrm{ZF}}$ is the dominant eigenvector of $\mathbf{\Pi}_{\mathcal{N}\{\mathbf{Q}_{ij}\}} \mathbf{Q}_{ii} \mathbf{\Pi}_{\mathcal{N}\{\mathbf{Q}_{ij}\}}$ and when $\mathcal{R}\{\mathbf{Q}_{ii}\} \subseteq \mathcal{R}\{\mathbf{Q}_{ij}\}$, e.g., when $\mathbf{Q}_{ij}$ is full-rank, then $\mathbf{w}_i^{\mathrm{ZF}} = \mathbf{0}$ [33].

In the following, we introduce some operating points, which are important in the sense that they lie on the outer boundary of the rate region; see, e.g., [34] and references therein.

Single-user (SU): The points achieved when one transmitter employs its MR strategy while the other refrains from transmission.

Maximum sum-rate (SR): The point where the sum of the rates is maximum. Graphically, it is the point where a line of slope -1 touches the Pareto boundary of the rate region.

Nash bargaining solution (NBS): The outcome of a Nash bargaining is a point $(\bar{R}_1, \bar{R}_2)$ such that $(\bar{R}_1 - R_1^*)(\bar{R}_2 - R_2^*)$ is maximized for some threat point (R_1^*, R_2^*) and $\bar{R}_i \geq R_i^*$. It is natural to use the NE as the threat point, since it is the only reasonable outcome if the systems are not able to agree on a solution. The NBS is only defined on convex utility regions, but we will call the solution to the corresponding optimization problem the NBS.

[3] We deliberately use the same symbols, as in the case of instantaneous CSI, to denote the rate and the power (R and p, respectively) in order to facilitate in the sequel a uniform treatment of both CSI scenarios.

4.1.2 Proposed algorithm

Here we elaborate the proposed bargaining algorithm that enables the transmitters to distributively design their beamforming vectors. We assume that there exists a feedback link from every receiver to every transmitter. The receivers use these links to feedback CSI. Each transmitter has CSI only on the links it is affecting. The transmitters are assumed synchronized, but no information (CSI or user data) is exchanged between them.

In the algorithm, we use as bargaining value an upper bound on the interference generated by system i to system j. This bound, denoted c_{ij}, is adjusted in every iteration. During the bargaining, the receivers feed back a one-bit message that tells the transmitters whether the iteration was successful or not, i.e. whether the rates increased or not. We denote l the iteration counter, which also acts as a quantitative measure of the overhead (total number of bits per RX-TX feedback link) and the computational complexity (total number of optimization problems that need to be solved).

A flowchart of the algorithm is illustrated in Figure 4.1. The first step of the algorithm is the decision whether the initialization point will be the NE or the ZF point. For this reason, the transmitters send two pilots using their MR and ZF beamforming vectors. The receivers measure the SINR for each transmission and feed back one-bit of information telling the transmitters which strategy yields higher SINR, hence rate. If $R_i(\mathbf{w}_i^{\mathrm{ZF}}, \mathbf{w}_j^{\mathrm{ZF}}) \geq R_i(\mathbf{w}_i^{\mathrm{MR}}, \mathbf{w}_j^{\mathrm{MR}})$ for both systems, the algorithm is initialized with the ZF point, since it is closer than the NE to the Pareto boundary. Hence, the algorithm will require fewer iterations to converge to a solution. If only one system achieves higher rate with the ZF strategies, there is no incentive for the other to accept the ZF point as initial point. The algorithm is then initialized with the NE point.

Then, the algorithm sets the stepsize for updating c_{ij}. As with any iterative algorithm, the best output is obtained for an infinitesimal stepsize. However, this is not practical, so we consider instead a fixed stepsize[4]. We assume that TX_i samples the interval $[0, p_{ij}(\mathbf{w}_i^{\mathrm{MR}})]$ uniformly in $N+1$ points, to allow up to N iterations. This gives the step $\delta_{ij} = \pm p_{ij}(\mathbf{w}_i^{\mathrm{MR}})/N$. The sign of δ_{ij} depends on the initial point. If the algorithm is initialized with the NE, δ_{ij} will be negative (decreasing interference). Otherwise, δ_{ij} will be positive (increasing interference). At iteration l, TX_i updates the interference level as $c_{ij}^l = c_{ij}^{l-1} + \delta_{ij}$ and solves the problem

$$\max_{\mathbf{w}_i \in \mathcal{W}} \; p_{ii}(\mathbf{w}_i) \tag{4.11}$$

$$\text{s.t.} \quad p_{ij}(\mathbf{w}_i) \leq c_{ij}^l. \tag{4.12}$$

The optimal solution of problem (4.11)–(4.12) is the beamforming vector which maximizes the useful power given that the generated interference is c_{ij}^l. As long

[4]Also, an adaptive stepsize can easily be incorporated to the algorithm.

as c_{ij}^l is chosen in the range $[0, p_{ij}(\mathbf{w}_i^{\mathrm{MR}})]$, there always exists a feasible solution to (4.11)–(4.12) [35]. The lower and upper end on the interference level correspond to the ZF and MR strategies, respectively. Furthermore, the bound will be tight at the optimum; hence, the inequality in (4.12) can be equivalently replaced with equality. We propose a solution to the optimization problem (4.11)–(4.12) in the following for the case of instantaneous and statistical CSI, respectively. TX_i uses $\mathbf{w}_i^l$ to transmit a pilot. RX_i measures $R_i(\mathbf{w}_i^l, \mathbf{w}_j^l)$ and if it is no smaller than $R_i(\mathbf{w}_i^{l-1}, \mathbf{w}_j^{l-1})$, it feeds back a one-bit message telling the transmitters to continue updating the interference level. As soon as the rate decreases for at least one of the receivers, the algorithm terminates and the transmitters will use the beamforming vectors from the previous iteration.

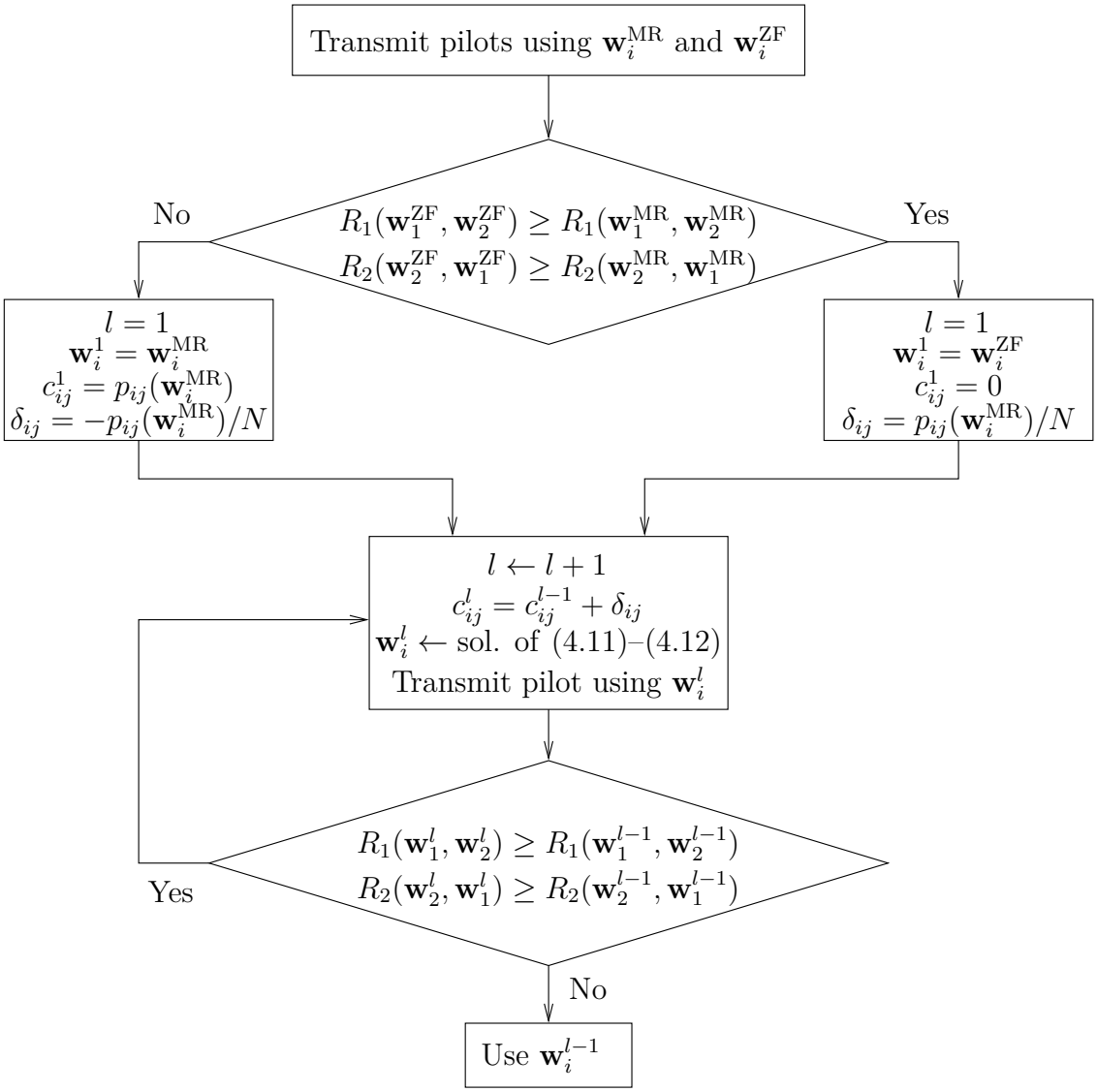


Figure 4.1: Flowchart describing the proposed cooperative beamforming algorithm

We claim that the algorithm is self-enforced. Suppose that, in one of the steps, TX_i chooses to cheat by not decreasing the interference level. Then, the rate of system j will decrease and RX_j will feedback a negative bit. According to the last step of the algorithm, the transmitters are expected to choose the beamforming vectors from the previous iteration. If TX_i does not, RX_j will notice and report it to TX_j. Then, TX_j will leave the bargaining and employ its MR beamforming vector instead. That is, if one system tries to cheat, then the cooperation is canceled and the operation falls back to the NE (the so-called threat point, in the context of Nash bargaining).

We first consider the instantaneous CSI. By inserting the expression (4.4) in (4.11)–(4.12), with inequality changed to equality, we get the problem

$$\max_{\mathbf{w}_i \in \mathcal{W}} \ |\mathbf{h}_{ii}^H \mathbf{w}_i|^2 \tag{4.13}$$

$$\text{s.t.} \quad |\mathbf{h}_{ij}^H \mathbf{w}_i|^2 = c_{ij}. \tag{4.14}$$

Since the objective of the algorithm is to find a PO point, the transmitters are only willing to use beamforming vectors that are candidates for achieving PO points. Any other beamforming vector will be a waste of power. Using (4.6) we get

$$\max_{\lambda_i \in [0,1]} \ |\mathbf{h}_{ii}^H \mathbf{w}_i^{\mathrm{PO}}(\lambda_i)|^2 \tag{4.15}$$

$$\text{s.t.} \quad |\mathbf{h}_{ij}^H \mathbf{w}_i^{\mathrm{PO}}(\lambda_i)|^2 = c_{ij}. \tag{4.16}$$

Note that the optimization (4.15)–(4.16) is now only with respect to the real scalar λ_i. Furthermore, the power constraint is obsolete, since the PO beamforming vectors use full power. That is, the inequality constraint in (4.2) is met with equality. Instead, we have a constraint on the range of the weighting factor λ_i. Finally, it is straightforward to see that the objective function (4.15) is monotonously increasing with λ_i. Thus, we can equivalently rewrite (4.15)–(4.16) as

$$\max_{\lambda_i \in [0,1]} \ \lambda_i \tag{4.17}$$

$$\text{s.t.} \quad |\mathbf{h}_{ij}^H \mathbf{w}_i^{\mathrm{PO}}(\lambda_i)|^2 = c_{ij}. \tag{4.18}$$

To simplify notation, we define

$$\alpha_i \triangleq \left(|\mathbf{h}_{ij}^H \mathbf{h}_{ii}| / \|\mathbf{h}_{ij}\|\right)^2 \text{ and } \beta_i \triangleq \left\| \mathbf{\Pi}_{\mathbf{h}_{ij}}^{\perp} \mathbf{h}_{ii} \right\| / \|\mathbf{h}_{ii}\| . \tag{4.19}$$

The values (4.19) are only calculated once per channel realization. For $c_{ij} > 0$ we write (4.18) as

$$\frac{\lambda_i^2 \alpha_i}{\lambda_i^2 + (1-\lambda_i)^2 + 2\lambda_i(1-\lambda_i)\beta_i} = c_{ij} \Leftrightarrow$$
$$\lambda_i^2(\alpha_i/c_{ij} + 2\beta_i - 2) + \lambda_i(2 - 2\beta_i) - 1 = 0.$$

When $c_{ij} = 0$, the ZF strategy is the optimal solution (i.e, $\lambda_i = 0$). Now, we write (4.17)–(4.18) as

$$\max \ \lambda_i \tag{4.20}$$
$$\text{s.t.} \quad \lambda_i^2(\alpha_i/c_{ij} + 2\beta_i - 2) + \lambda_i(2 - 2\beta_i) - 1 = 0, \tag{4.21}$$
$$0 \leq \lambda_i \leq 1. \tag{4.22}$$

The solution to (4.20)–(4.22) is the largest of the two solutions to (4.21) that satisfies (4.22).

Then we consider the statistical CSI. By inserting (4.10) in (4.11)–(4.12) we get

$$\max_{\mathbf{w}_i \in \mathbb{C}^n} \ \mathbf{w}_i^H \mathbf{Q}_{ii} \mathbf{w}_i \tag{4.23}$$
$$\text{s.t.} \quad \mathbf{w}_i^H \mathbf{Q}_{ij} \mathbf{w}_i \leq c_{ij}, \tag{4.24}$$
$$\mathbf{w}_i^H \mathbf{w}_i \leq 1. \tag{4.25}$$

Problem (4.23)–(4.25) is a quadratically constrained quadratic program (QCQP). The feasibility set determined by (4.24)–(4.25) is convex. However, the optimization is non-convex owing to the form of the objective function. However, it can still be solved optimally and efficiently using semidefinite relaxation. This is because semidefinite relaxation is tight for QCQP problems of the form in (4.23)–(4.24), as shown in [36].

We briefly elaborate the procedure, similar to the way we did in [35]. We change the optimization variables to $\mathbf{W}_i \triangleq \mathbf{w}_i \mathbf{w}_i^H$. Note that

$$\mathbf{W}_i = \mathbf{w}_i \mathbf{w}_i^H \Leftrightarrow \mathbf{W}_i \succeq \mathbf{0} \ \text{ and } \ \text{rank}\{\mathbf{W}_i\} = 1. \tag{4.26}$$

Using (4.26) and the property that $\text{tr}\{\mathbf{YZ}\} = \text{tr}\{\mathbf{ZY}\}$ for matrices $\mathbf{Y}$, $\mathbf{Z}$ of compatible dimensions, the average power term in (4.24) can be written as

$$\begin{aligned} \mathbf{w}_i^H \mathbf{Q}_{ij} \mathbf{w}_i &= \text{tr}\left\{\mathbf{w}_i^H \mathbf{Q}_{ij} \mathbf{w}_i\right\} = \text{tr}\left\{\mathbf{Q}_{ij} \mathbf{w}_i \mathbf{w}_i^H\right\} \\ &= \text{tr}\left\{\mathbf{Q}_{ij} \mathbf{W}_i\right\}. \end{aligned} \tag{4.27}$$

Due to (4.26) and (4.27), we equivalently recast (4.23)–(4.24) as

$$\max_{\mathbf{W} \in \mathbb{C}^{n \times n}} \ \text{tr}\left\{\mathbf{Q}_{ii} \mathbf{W}_i\right\} \tag{4.28}$$
$$\text{s.t.} \quad \text{tr}\left\{\mathbf{Q}_{ij} \mathbf{W}_i\right\} \leq c_{ij}, \tag{4.29}$$
$$\text{tr}\left\{\mathbf{W}_i\right\} \leq 1, \tag{4.30}$$
$$\mathbf{W}_i \succeq \mathbf{0}, \tag{4.31}$$
$$\text{rank}\{\mathbf{W}_i\} = 1. \tag{4.32}$$

The objective function (4.28), the constraints (4.29) and (4.30) are linear. The cone of positive semidefinite matrices (4.31) is convex. But the rank constraint (4.32)

is non-convex. Dropping it, the remaining problem (4.28)–(4.31) is a semidefinite programming (SDP) problem, which can be solved efficiently. Due to the absence of (4.32), the SDP problem will not necessarily return rank-1 optimal matrices. We experienced through extensive simulations that it actually *does* yield rank-1 matrices.

4.1.3 Numerical illustrations

We present extensive simulation results to evaluate the performance of the algorithm we propose. We focus on the case of statistical CSI, but also provide some results for instantaneous CSI. Firstly we explain how we generate CSI (i.e., channel covariance matrices or channel vectors) for simulation purposes. Then we compare the outcome of the algorithm to the NE, ZF, SR, and NBS, followed by showing exemplary bargaining trajectories. Finally, in we illustrate the SAPHYRE gain. Throughout the simulations, we assume that the transmitters use $n = 5$ antennas. We allow our algorithm run up to $N = 20$ iterations. The results reported in Figure 4.2–4.7 are averages over 100 Monte-Carlo (MC) runs. Figure 4.2–4.7 illustrate the sum of the transmission rates, i.e., $R_1 + R_2$. Figure 4.8–4.11 show examples of achievable rate regions, i.e., for a single CSI realization.

We generate the direct and the crosstalk channels in two different ways, to model the scenarios of weak or strong spatial correlation. Specifically, in the case of instantaneous CSI and weak correlation, we generate the channel vectors $\mathbf{h}_{ii}$ and $\mathbf{h}_{ij}$ drawing independent samples from $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. For the scenario of strong correlation, we use the formula

$$\mathbf{h}_{ij} = \mu_i \mathbf{h}_{ii} + \sqrt{1 - \mu_i^2}\tilde{\mathbf{h}}_{ij}, \tag{4.33}$$

where $\mathbf{h}_{ii}$ and $\tilde{\mathbf{h}}_{ij}$ are drawn from $\mathcal{CN}(\mathbf{0}, \mathbf{I})$, and $\mu_i \in [0, 1]$. A value of μ_i close to 1 refers to the case of strong interference.

In the case of statistical CSI, we construct the covariance matrices, of rank r, randomly as

$$\mathbf{Q} = \sum_{k=1}^{r} \mathbf{q}_k \mathbf{q}_k^H, \tag{4.34}$$

where $\mathbf{q}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. For the scenario of weak correlation, we generate the covariance matrices $\mathbf{Q}_{ii}$ and $\mathbf{Q}_{ij}$ independently according to (4.34). For the scenario of strong correlation, we construct the matrices such that the angle between the eigenvectors of the direct matrix and the eigenvectors of the crosstalk matrix is small. Assuming that $r_{ii} \leq r_{ij}$, we first generate $\mathbf{Q}_{ii}$ as in (4.34). Then, we construct the vectors $\{\mathbf{q}_{ij,k}\}_k$ that define $\mathbf{Q}_{ij}$ as

$$\begin{cases} \mathbf{q}_{ij,k} = \mu_i \mathbf{q}_{ii,k} + \sqrt{1 - \mu_i^2}\tilde{\mathbf{q}}_{ij,k}, & k \leq r_{ii} \\ \mathbf{q}_{ij,k} = \tilde{\mathbf{q}}_{ij,k}, & k > r_{ii} \end{cases} \tag{4.35}$$

where $\tilde{\mathbf{q}}_{ij,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mu_i \in [0, 1]$. If $r_{ii} > r_{ij}$, the matrices are constructed the other way around.

Here we provide results for statistical CSI, both for weak and strong spatial correlation. Also, we distinguish among the cases of having full-rank and low-rank covariance matrices. In the low-rank scenario, the covariance matrices of the direct-channels have rank $r_{11} = r_{22} = 2$, and covariance matrices of the cross-talk channels have rank $r_{12} = r_{21} = 4$. For strong correlation, we use $\mu_i = 0.8$.

In Figure 4.2 and 4.3, we study the full-rank scenario. First, we note that the ZF sum rate is equal to 0 since full-rank crosstalk matrices correspond to $\mathbf{w}_i^{\mathrm{ZF}} = \mathbf{0}$. Second, we see that the sum rates for the proposed algorithm, the NBS, and the NE saturate for high SNR. The reason is that when the SNR is high, interference is the main limiting factor. Since the interference cannot become zero, except for the SU-points, there should be a limitation. Third, since there is no interference at the SU points, the corresponding rates will grow unbounded with SNR and the SR will be found at a SU point for high SNR.

In Figure 4.4 and 4.5, we illustrate the low-rank scenario. Here, all points but the NE converge to the same sum rate at high SNR. The difference is that, for strong correlation they converge at higher SNR value than for weak correlation. Also, we see that the rates grow almost linearly with the SNR. In general, there exists a non-trivial zero-forcing point for the case of low-rank matrices. Using this, the noise is the only limitation. When the noise decreases, the rates increase. At low SNR, the ZF starts growing later for strong correlation than for weak correlation.

Furthermore, we evidence that weak correlation (Figure 4.2 and 4.4) gives higher rates for the proposed algorithm, the NE, and the NBS, than strong correlation (Figure 4.3 and 4.5). As a general remark, low SNR means operation in the noise-limited regime and all the rates but the ZF are almost the same.

Concluding, we see that the performance of the proposed algorithm is slightly below the NBS and close to the SR, except for full-rank matrices and high SNR. Most important, the algorithm performs consistently much better than the NE, which would be the outcome if there was no cooperation.

In Figure 4.6 and 4.7 we report the results for weak and strong correlation ($\mu_i = 0.9$), respectively. We note that the curves behave similarly to the ones in Figure 4.4 and 4.5. The reason for this is that the case of instantaneous CSI can be regarded as a specific instance of the low-rank statistical CSI when all covariance matrices are rank-1.

Then we give examples of the bargaining trajectory, i.e., the rate points (marked with stars) reached at every iteration of the proposed algorithm. Here, the maximum number of iterations used is $N = 10$. Figure 4.8 and 4.9 illustrate the trajectories for statistical CSI with full-rank covariance matrices and SNR equal to 0 and 10 dB, respectively. The Pareto boundary is calculated using the technique proposed in [35]. Figure 4.10 and 4.11 illustrate the trajectories for instantaneous CSI and

SNR equal to 0 and 10 dB, respectively. The Pareto boundary is calculated using the technique proposed in [32].

For statistical CSI and full-rank matrices, the algorithm is always initialized with the NE, since a non-trivial ZF point does not exist. For instantaneous CSI and low SNR it is initialized with the NE point, while for high SNR with the ZF point. We note that the final outcome of the bargaining algorithm is close to the Pareto boundary, but does not necessarily lie on it. On one hand, the final outcome depends on the stepsize of the algorithm. On the other hand, the algorithm terminates when either of the rates stops increasing, i.e., when the tangent of the trajectory stops being positive. More on, the outcome is close to NBS, but generally far from SR. We note that the SR does not imply that both systems have increased their rates compared to NE, while the proposed algorithm and NBS guarantee that both systems get at least their NE rates. Finally, in all figures we show what the bargaining trajectory would look like if the algorithm went the entire way from one extreme point (NE or ZF) to the other with small steps. Note that for statistical CSI and full-rank matrices the ZF point corresponds to the origin of the rate region.

At last, we illustrate the SAPHYRE gain of the proposed algorithm over the non-sharing situation as a function of the SNR. In Figure 4.2–4.7 we illustrate the sum-rate of the orthogonal (TDMA) sharing. We see, Figure 4.2 and 4.3, that TDMA sharing performs better than the proposed algorithm for the case of statistical CSI and full-rank covariance matrices and high SNR. Also, for statistical CSI with low-rank covariance matrices and medium SNR, Figure 4.5, the TDMA scheme gives higher rates than our algorithm. In Figure 4.12 and 4.13 we illustrate the absolute and fractional SAPHYRE gains, respectively. The illustrations are made for the weak spatial correlation scenario. For each SNR value we have made 100 Monte-Carlo simulations. The absolute SAPHYRE gain is defined as in (1.1), but for $K = 2$ users. We see that the absolute gain increases with increasing SNR for the scenarios of instantaneous CSI and statistical CSI with low-rank matrices. For statistical CSI the gain starts to decrease at SNR $\approx$ 5dB and becomes negative at SNR $\approx$ 18dB. We define the fractional SAPHYRE gain as in (1.2). In Figure 4.13, we see that the more we know about the channels, the higher is the fractional SAPHYRE gain. For the scenario of instantaneous CSI, we have a fractional gain close to 2.0, i.e., the rate is doubled. Again, we see that scenario of high SNR and statistical CSI with full-rank covariance matrices yield a fractional SAPHYRE gain $\Xi_{\mathrm{F}} < 1$.

From the rate regions in Figure 4.8, 4.10, and 4.11 we see that the orthogonal (TDMA) sharing region lies inside the non-orthogonal sharing region. Also, we notice that the outcome of the proposed algorithm lies outside the TDMA region. Studying Figure 4.9 (statistical CSI and high SNR), we see that the TDMA region does not entirely contained in the non-orthogonal sharing region. The outcome of the algorithm lies south-west of the TDMA boundary. This illustrates why we get $\Xi_{\mathrm{F}} < 1$.

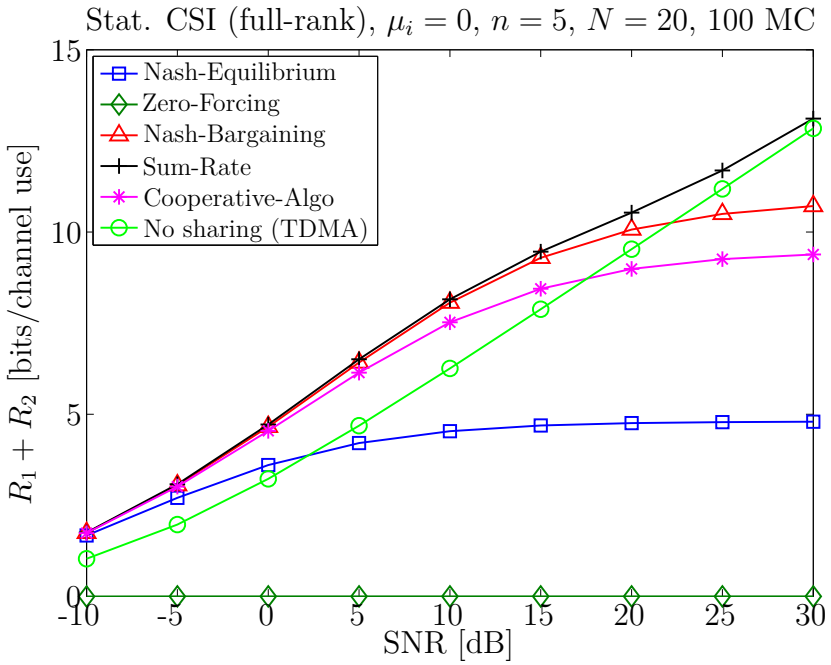


Figure 4.2: Sum rate; stat. CSI (full-rank), $\mu_i = 0$

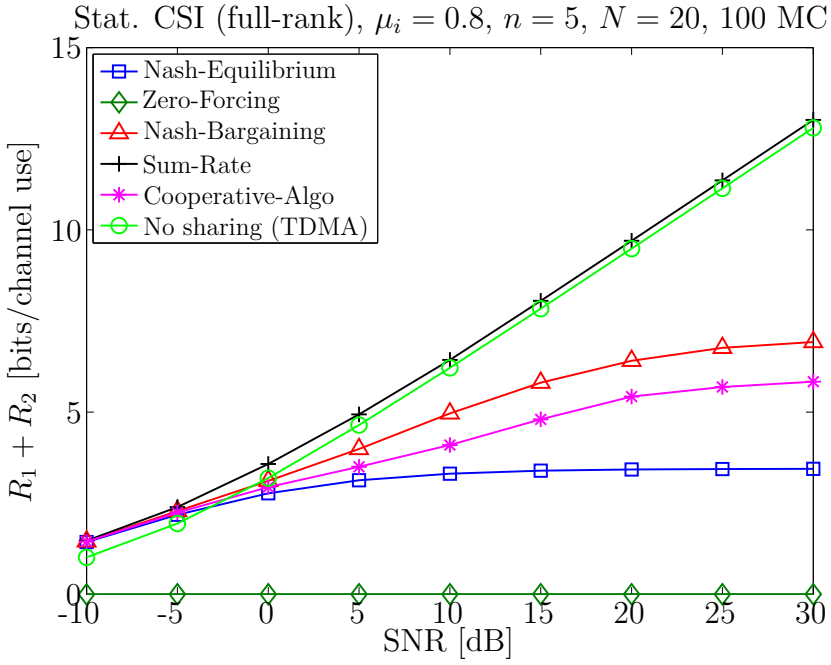


Figure 4.3: Sum rate; stat. CSI (full-rank), $\mu_i = 0.8$

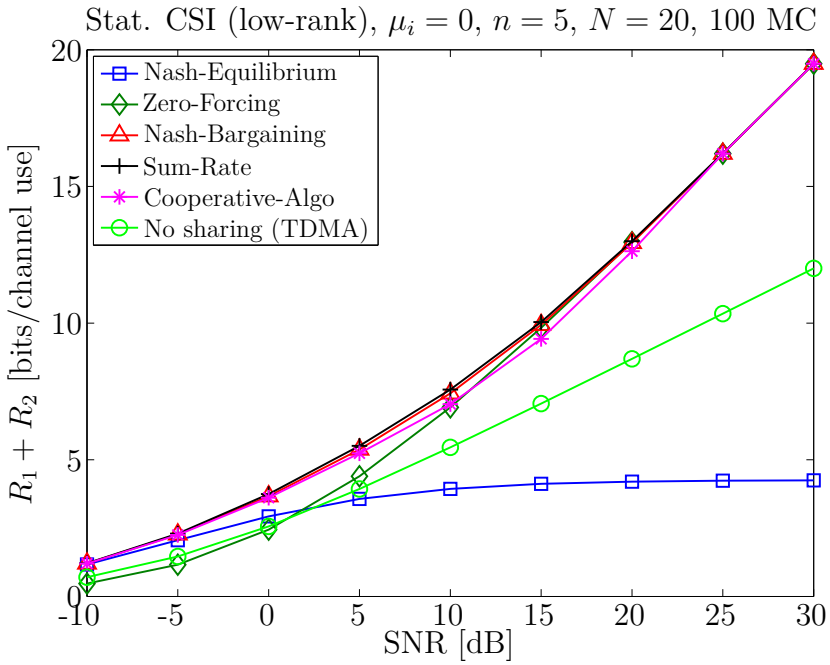


Figure 4.4: Sum rate; stat. CSI (low-rank), $\mu_i = 0$

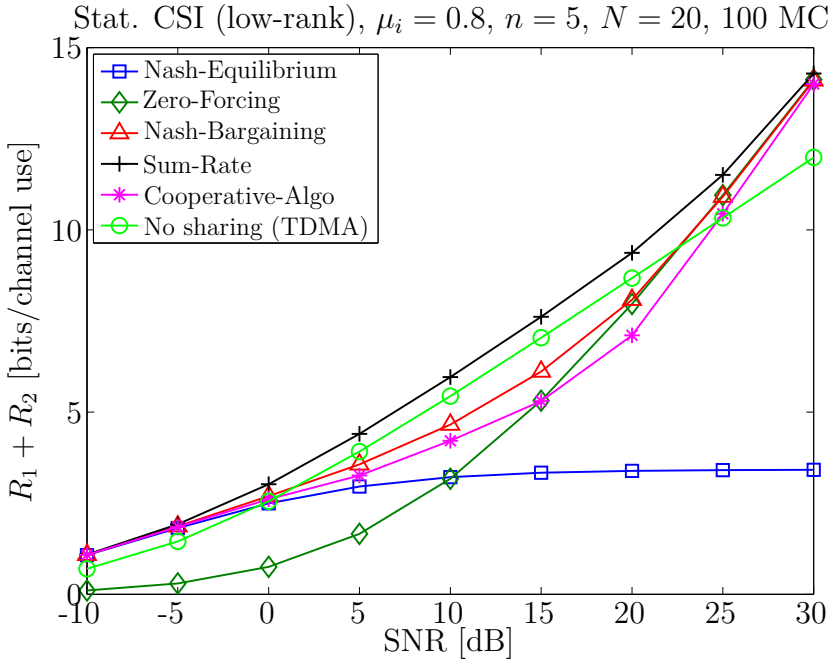


Figure 4.5: Sum rate; stat. CSI (low-rank), $\mu_i = 0.8$

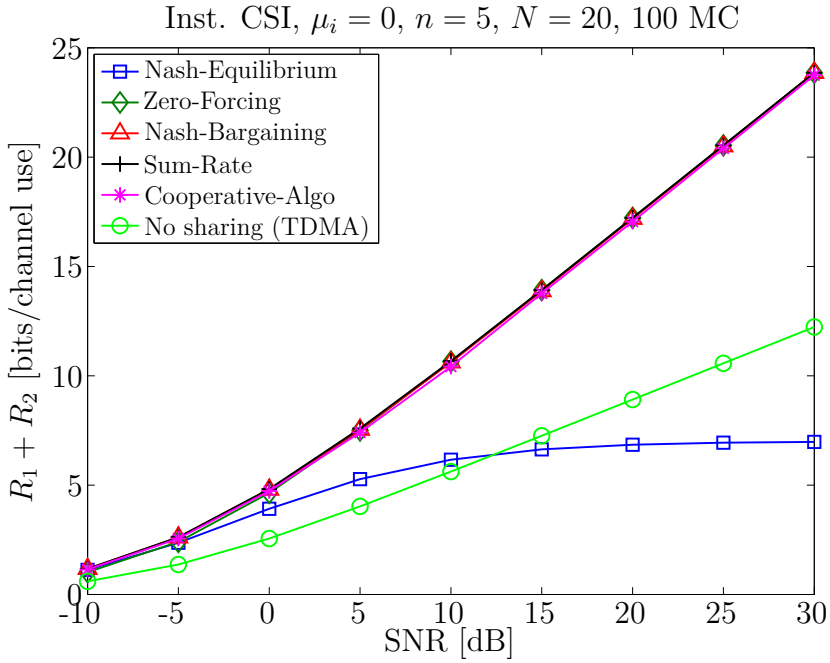


Figure 4.6: Sum rate; inst. CSI, $\mu_i = 0$

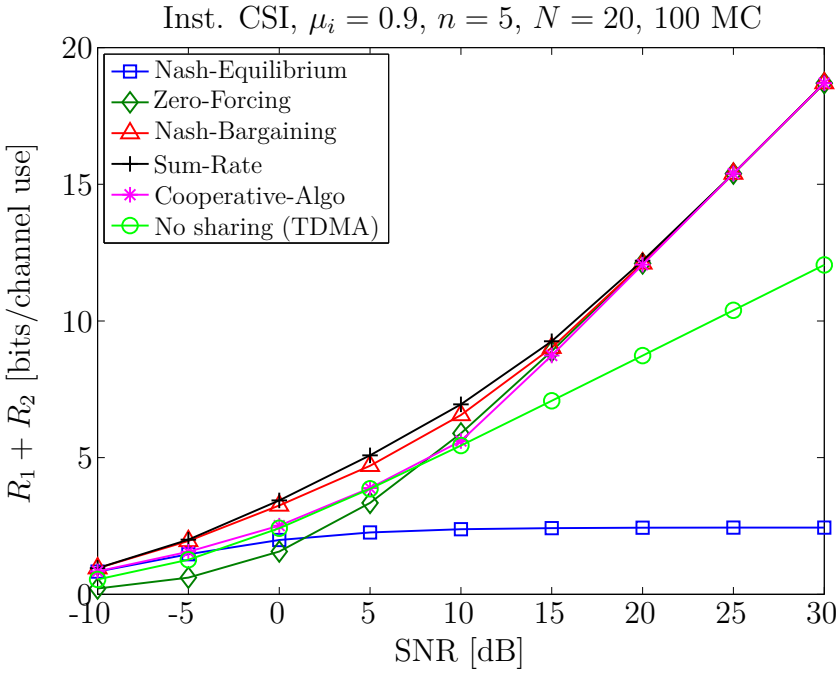


Figure 4.7: Sum rate; inst. CSI, $\mu_i = 0.9$

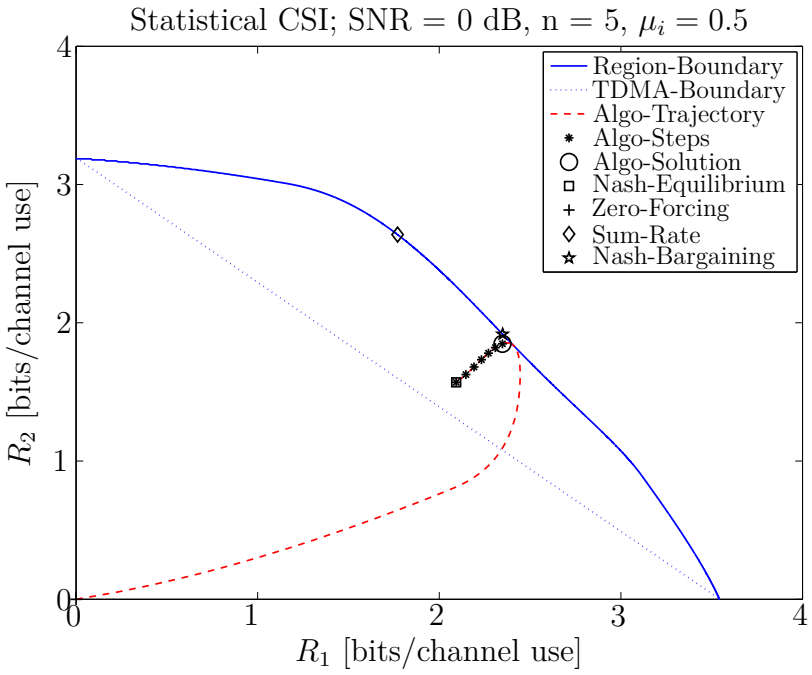


Figure 4.8: Bargaining trajectory; stat. CSI, SNR = 0 dB

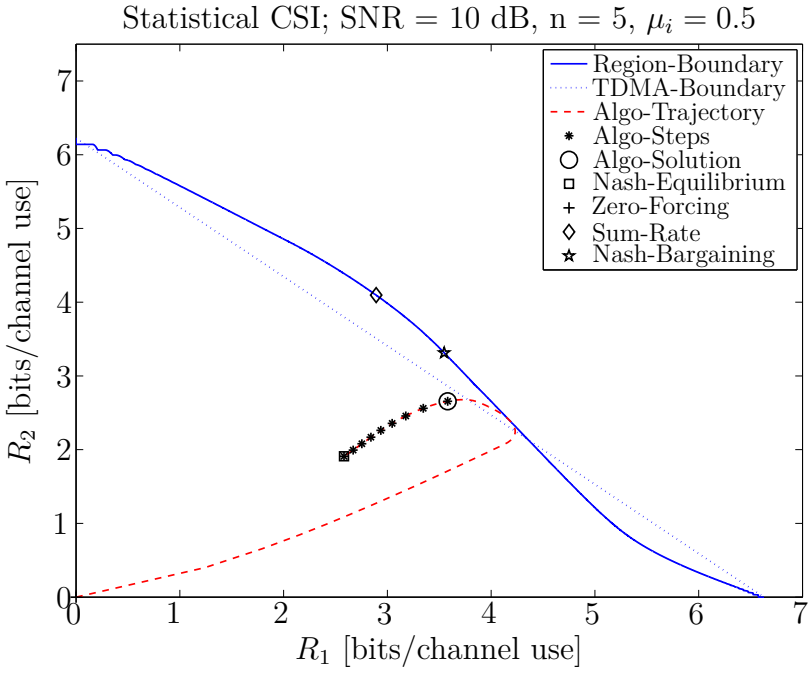


Figure 4.9: Bargaining trajectory; stat. CSI, SNR = 10 dB

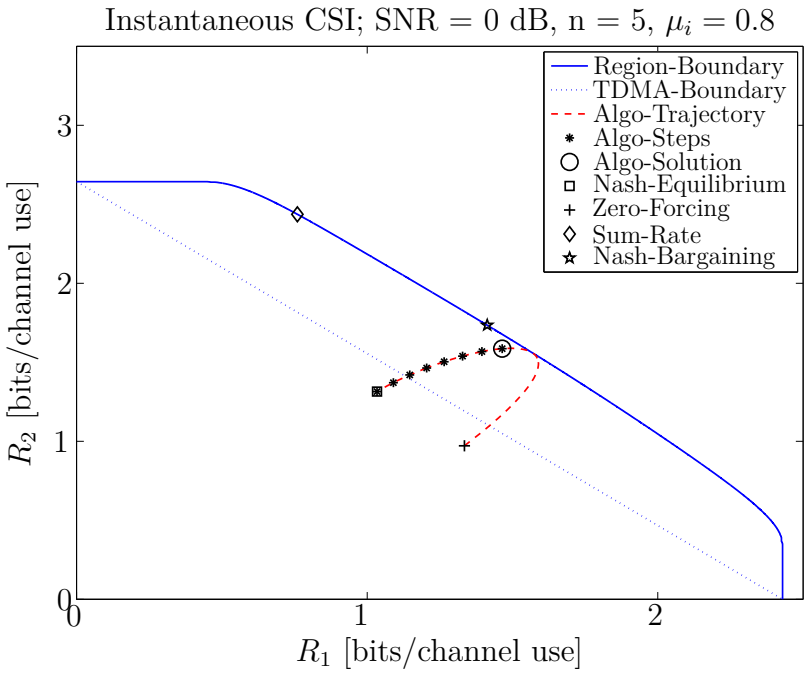


Figure 4.10: Bargaining trajectory; inst. CSI, SNR = 0 dB

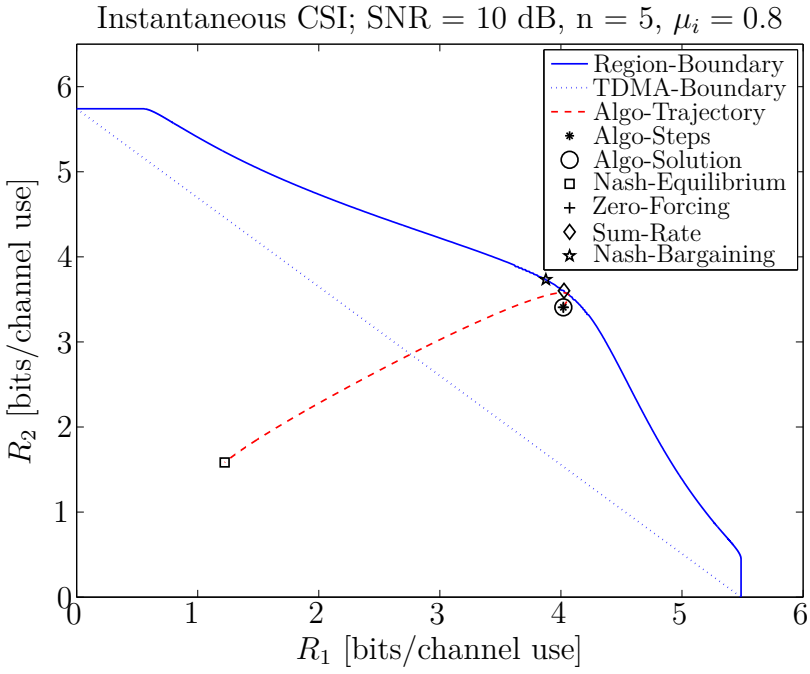


Figure 4.11: Bargaining trajectory; inst. CSI, SNR = 10 dB

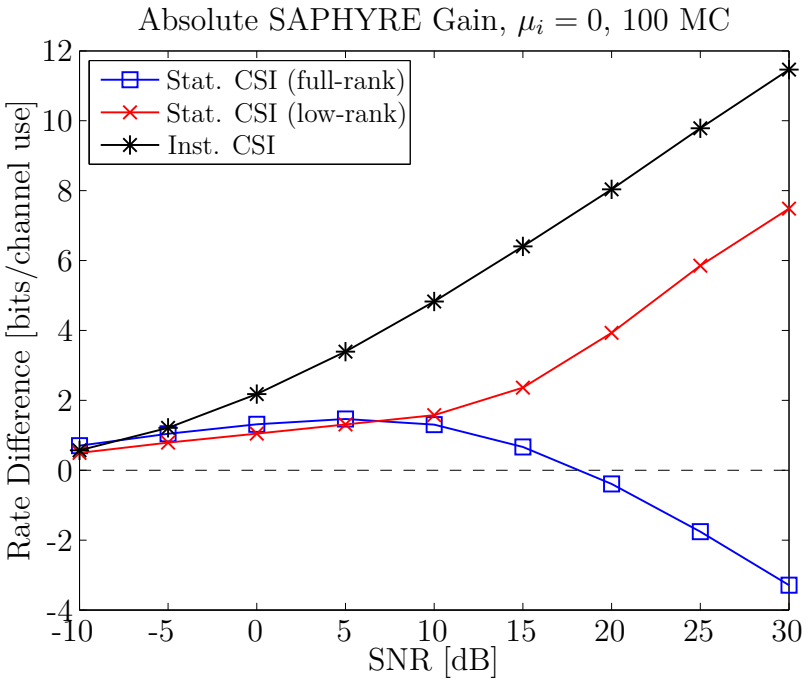


Figure 4.12: Absolute SAPHYRE Gain

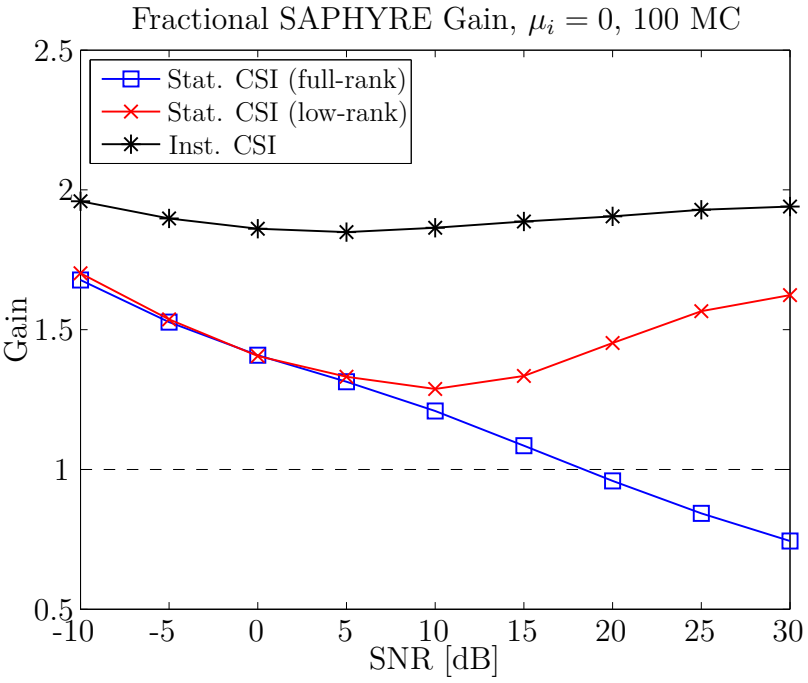


Figure 4.13: Fractional SAPHYRE Gain

4.2 Resource Optimization in Multicell OFDMA Networks

In this section, we consider the downlink of multicell orthogonal frequency-division multiple-access (OFDMA) networks and address the adaptive allocation of spectrum, power, and rate. We assume networks with frequency reuse one and discrete-level rates. We formulate the joint allocation problem as a nonlinear mixed integer program (MIP), which is computationally intractable to solve optimally for practical problem sizes. We exploit the capability of the receivers to measure the perceived interference-plus-noise on every subcarrier and accordingly decompose the joint allocation problem into subproblems that are solved by individual base stations. Each subproblem is a linear MIP and its optimal solution can be obtained by means of standard branch-and-cut solvers, at least for small to medium problem sizes. In order to further reduce the complexity, we propose a distributed iterative algorithm that capitalizes on the subgradient method. We demonstrate the merit of the proposed algorithm with numerical comparisons of its performance with the solution of the individual MIPs and the iterative waterfilling algorithm.

4.2.1 System model and problem formulation

We consider downlink transmission in a multicell OFDMA network with a set $\mathcal{L} \triangleq \{i : i = 1, \ldots, L\}$ of BSs and a set $\mathcal{K} \triangleq \{k : k = 1, \ldots, K\}$ of receivers, where every BS is assumed to serve the same number of receivers, i.e. K/L. The ith BS serves the receivers within the set $\mathcal{K}_i \triangleq \{(i-1)K/L + 1, \ldots, iK/L\}$. The network bandwidth is shared by all BSs and it is divided into a set $\mathcal{N} \triangleq \{n : n = 1, \ldots, N\}$ of orthogonal subcarriers. The channel of each subcarrier is flat, since its bandwidth is chosen small enough compared to the coherence bandwidth. The number of bits loaded on each subcarrier is chosen from a finite set $\mathcal{Q} \triangleq \{q : q = 1, \ldots, Q\}$.

We define the binary allocation variables $x_k^{n,q}$, where $x_k^{n,q} = 1$ if subcarrier n is assigned to receiver k with rate q and $x_k^{n,q} = 0$ otherwise. To avoid intracell interference, each subcarrier can be used by *at most* one receiver per cell. Hence, for BS_i and subcarrier n we have the constraint

$$\sum_{k \in \mathcal{K}_i} \sum_{q=1}^{Q} x_k^{n,q} \leq 1. \tag{4.36}$$

The sum in the left-hand side of (4.36) equals to zero when BS_i does not allocate any receiver to subcarrier n.

Let $G_{i,k}^n$ denote the gain of the channel between BS_i and receiver k on the nth subcarrier. The signal-to-interference-plus-noise ratio (SINR) of receiver k, served by BS_i on subcarrier n with transmit power p_i^n, is

$$\gamma_k^n \triangleq \frac{G_{i,k}^n p_i^n}{I_k^n(p_{-i}^n)}. \tag{4.37}$$

In (4.37), the interference generated by simultaneous transmissions throughout the network on subcarrier n plus the AWGN noise variance σ_k^2 is denoted

$$I_k^n(p_{-i}^n) \triangleq \sum_{j=1, j\neq i}^{L} G_{j,k}^n p_j^n + \sigma_k^2, \tag{4.38}$$

where $p_{-i}^n \triangleq [p_1^n, \ldots, p_{i-1}^n, p_{i+1}^n, \ldots, p_L^n]$ is the vector of all interfering transmit powers.

Assuming Gaussian signaling, let T_q denote the threshold that the SINR should reach to load q bits, i.e., $\log_2(1+\gamma) = q \Leftrightarrow \gamma = 2^q - 1 \triangleq T_q$. When BS_i decides to serve receiver $k \in \mathcal{K}_i$ with q bits on subcarrier n, i.e. $x_k^{n,q} = 1$, then, due to (4.37), in order to have $\gamma_k^n = T_q$ the required transmit power is $p_i^n = I_k^n(p_{-i}^n) T_q / G_{i,k}^n$. Due to (4.36), this power is given, for an arbitrary subcarrier and bit allocation, by

$$p_i^n = \sum_{k\in\mathcal{K}_i} \sum_{q=1}^{Q} x_k^{n,q} I_k^n(p_{-i}^n) T_q / G_{i,k}^n. \tag{4.39}$$

Moreover, we assume that the total transmit power of every BS cannot exceed the maximum budget P, i.e. $\sum_{n=1}^{N} p_i^n \leq P$.

The objective is to maximize the achievable sum-rate in the network, i.e., the sum of bit rates of all subcarriers over all cells, subject to the aforementioned constraints. Consequently, the joint resource allocation problem is stated as

$$\max_{\mathbf{X},\mathbf{P}} \sum_{i=1}^{L} \sum_{n=1}^{N} \sum_{k\in\mathcal{K}_i} \sum_{q=1}^{Q} q x_k^{n,q} \tag{4.40}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}_i} \sum_{q=1}^{Q} x_k^{n,q} \leq 1 \qquad \forall i \in \mathcal{L}, \forall n \in \mathcal{N}, \tag{4.41a}$$

$$p_i^n = \sum_{k\in\mathcal{K}_i} \sum_{q=1}^{Q} x_k^{n,q} I_k^n(p_{-i}^n) T_q / G_{i,k}^n \qquad \forall i \in \mathcal{L}, \forall n \in \mathcal{N}, \tag{4.41b}$$

$$\sum_{n=1}^{N} p_i^n \leq P \qquad \forall i \in \mathcal{L}. \tag{4.41c}$$

Problem (4.40)–(4.41) is a MIP with KNQ binary allocation variables $\mathbf{X} = \{x_k^{n,q} \in \{0,1\}\}_{k\in\mathcal{K}}^{n\in\mathcal{N},\ q\in\mathcal{Q}}$ and LN continuous power variables $\mathbf{P} = \{p_i^n \in \mathbb{R}_+\}_{i\in\mathcal{L}}^{n\in\mathcal{N}}$. This problem is NP-hard in general [37]. The formulation is nonlinear due to the right-hand side of (4.41b) which involves, due to (4.38), bilinear products of the optimization variables. Finding the optimal solution requires an exhaustive search with worst-case complexity exponential in the total number of variables. The complexity is prohibitive for modern broadband networks which have hundreds of subcarriers. This motivates the low-complexity distributed approach that we are proposing in Section 4.2.3.

4.2.2 Single-cell resource allocation

The most significant challenge in the solution of problem (4.40)–(4.41) is due to the interference-plus-noise terms $\{I_k^n(p_{-i}^n)\}$ in (4.41b) that couple the resource allocation performed in different cells. However, the fact that each receiver is able to sense and measure the interference-plus-noise on subcarriers motivates us to decompose the global problem into subproblems solved by individual BSs. In other words, BS_i takes as input the values $\{I_k^n\}_{k\in\mathcal{K}_i}^{n\in\mathcal{N}}$ collecting them from the receivers in its cell, when the other BSs have already performed the resource allocation. Hence, the coupling among the resource allocation problems in different cells is eliminated. Consequently, the joint problem (4.40)–(4.41) decouples into L sub-problems, each solved separately by a different BS. The problem corresponding to BS_i is

$$\max_{\mathbf{X}_i,\mathbf{P}_i} \sum_{n=1}^{N}\sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} q x_k^{n,q} \tag{4.42}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} x_k^{n,q} \le 1 \qquad \forall n\in\mathcal{N}, \tag{4.43a}$$

$$p_i^n = \sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} x_k^{n,q} I_k^n T_q / G_{i,k}^n \qquad \forall n\in\mathcal{N}, \tag{4.43b}$$

$$\sum_{n=1}^{N} p_i^n \le P. \tag{4.43c}$$

Problem (4.42)–(4.43) is a MIP with KNQ/L binary variables $\mathbf{X}_i = \{x_k^{n,q} \in \{0,1\}\}_{k\in\mathcal{K}_i}^{n\in\mathcal{N},\ q\in\mathcal{Q}}$ and N continuous variables $\mathbf{P}_i = \{p_i^n \in \mathbb{R}_+\}^{n\in\mathcal{N}}$. Not only this problem has L times smaller dimension than the joint one, but also it is linear, since the constraints (4.43b) have now, for given $\{I_k^n\}$, become linear. There exist several solvers, implementing branch-and-cut techniques, that find the optimal solution of linear MIP problems frequently avoiding exhaustive search. However, the worst-case complexity of these techniques still increases exponentially with the number of variables and becomes impractical for large problem sizes, motivating us to investigate low-complexity solutions to (4.42)–(4.43). Due to the binary variables, this problem is non-convex. Existing solutions to this problem of single-cell resource allocation typically exploits the relaxation of binary variables so that the problem can be solved using convex linear programming [38]. The disadvantage is that rounding off the variables into binary ones takes the solution far from the optimal solution. Herein, we take advantage of the seminal contribution on multicarrier systems in [39], which has shown that, using dual optimization, the duality gap decreases as the number of subcarriers increases. The large number of subcarriers in practical OFDMA networks therefore motivates us to solve (4.42)–(4.43) in the dual domain.

In the following, we focus on the resource allocation problem in the ith cell, assuming that the allocation has been already performed in the other cells, i.e. for some given $\{I_k^n\}_{k\in\mathcal{K}_i}^{n\in\mathcal{N}}$. Inspecting (4.43b), we observe that the transmit powers $\mathbf{P}_i$ depend entirely on the variables $\mathbf{X}_i$, provided that they also meet the bound (4.43c). Hence, substituting (4.43b) into (4.43c), we can rewrite problem (4.42)–(4.43), with respect to only the optimization variables $\mathbf{X}_i$, as

$$\max_{\mathbf{X}_i} \sum_{n=1}^{N}\sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} q x_k^{n,q} \tag{4.44}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} x_k^{n,q} \le 1 \qquad \forall n \in \mathcal{N}, \tag{4.45a}$$

$$\sum_{n=1}^{N}\sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} x_k^{n,q} I_k^n T_q / G_{i,k}^n \le P. \tag{4.45b}$$

This enables us to solve the MIP (4.42)–(4.43) in two steps. First, we solve the linear binary problem (4.44)–(4.45) to determine the subcarrier and bit level allocation, and then plug the solution into (4.43b) to compute the transmit powers.

The solution to (4.44)–(4.45) would be straightforward if we decouple the power budget constraint in (4.45b) and perform the optimization per subcarrier. This motivates the incorporation of (4.45b) into the objective function and form a Lagrangian function as

$$\begin{aligned} L_i(\mathbf{X}_i, \lambda_i) = &\sum_{n=1}^{N}\sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} q x_k^{n,q} \\ &- \lambda_i \left(\sum_{n=1}^{N}\sum_{k\in\mathcal{K}_i}\sum_{q=1}^{Q} x_k^{n,q} I_k^n T_q / G_{i,k}^n - P \right) \end{aligned} \tag{4.46}$$

and the corresponding dual function as

$$D_i(\lambda_i) = \sup_{\mathbf{X}_i} \left\{ L_i(\mathbf{X}_i, \lambda_i) : (4.45a) \right\}, \tag{4.47}$$

where λ_i is the Lagrange multiplier. This multiplier is obtained in the dual domain for a given $\mathbf{X}_i$ by solving the corresponding dual problem

$$\min_{\lambda_i \ge 0} \; D_i(\lambda_i). \tag{4.48}$$

This problem can be solved by the subgradient method, i.e., beginning with an initial $\lambda_i(0)$, given $\lambda_i(t)$ at iteration t, we obtain $\mathbf{P}_i$ from $\mathbf{X}_i$ using (4.43b). We then update the Lagrange multiplier as

$$\lambda_i(t+1) = \left[\lambda_i(t) - \alpha \left(P - \sum_{n=1}^{N} p_i^n \right) \right]^+, \tag{4.49}$$

where $P - \sum_{n=1}^{N} p_i^n$ is the subgradient of $D_i(\lambda_i)$ with respect to λ_i and α is a step size that should be small enough to ensure the convergence [40]. The aforementioned approach therefore enables individual BSs to contribute the solution of the original problem separately.

To evaluate $D_i(\lambda_i)$ for a given λ_i in (4.47), BS_i substitutes $L_i(\mathbf{X}_i, \lambda_i)$ in (4.47) with (4.46) and forms an optimization problem represented by

$$\max_{\mathbf{X}_i} \quad \sum_{n=1}^{N} \sum_{k \in \mathcal{K}_i} \sum_{q=1}^{Q} x_k^{n,q} f_k^{n,q} \tag{4.50}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}_i} \sum_{q=1}^{Q} x_k^{n,q} \leq 1 \qquad \forall n \in \mathcal{N}, \tag{4.51}$$

where $f_k^{n,q} \triangleq q - \lambda_i I_k^n T_q / G_{i,k}^n$. Due to (4.51), each subcarrier can be used by at most one receiver, with a single bit rate, within cell i. This statement along with the decomposable form of (4.50)–(4.51) enables separate allocation for each individual subcarrier. The solution is therefore obtained by assigning each subcarrier n to receiver $k_n \in \mathcal{K}_i$ with bit rate q_n as

$$(k_n, q_n) = \underset{(k,q):k \in \mathcal{K}_i, q \in \mathcal{Q}}{\arg\max} \; f_k^{n,q} \tag{4.52}$$

provided that $f_{k_n}^{n,q_n} > 0$. In other words, for each subcarrier in cell i, we go over the QK/L possible receiver-bit assignments and select the one giving the largest positive value. Hence, $x_k^{n,q} = 1$ if $k = k_n$ and $q = q_n$, otherwise $x_k^{n,q} = 0$. Due to (4.43b), the transmit power is $p_i^n = I_k^n T_q / G_{i,k}^n$ in the former case and $p_i^n = 0$ in the latter case. However when $f_{k_n}^{n,q_n} \leq 0$, then $x_k^{n,q} = 0$ for all $k \in \mathcal{K}_i$, $q \in \mathcal{Q}$, and accordingly $p_i^n = 0$.

4.2.3 Distributed resource allocation algorithm

Given the solution for the allocation problem of each BS, presented in Section 4.2.2, in the sequel we propose a distributed subcarrier, power, and bit level (DSPB) allocation algorithm for the downlink of multicell OFDMA networks. The DSPB algorithm is based on the iterative update of the Lagrange multiplier in (4.49). We assume that there is a network coordinator, which synchronizes the BSs so that they know their order in the algorithm. This coordinator, also, terminates the algorithm upon satisfaction of the convergence condition. During the algorithm iterations, the channel gains are assumed to be constant. In addition, there is a mechanism to feedback, for all subcarriers, the channel gains and perceived interference from all receivers in each cell to the corresponding BS.

At first, every BS initializes the Lagrange multiplier and distributes uniformly a part of the power budget on all subcarriers (step 1, where $\delta < 1$). The network coordinator continues the iterations till the aggregate differential power in the network

Algorithm 3 Distributed Subcarrier, Power, and Bit level allocation (DSPB)

1: Initialization: $t = 0$, $\lambda_i(0) = \lambda_{\text{init}} \quad \forall i \in \mathcal{L}$, $p_i^n = \delta P/N \quad \forall i \in \mathcal{L}$, $\forall n \in \mathcal{N}$
2: **while** $\sum_{i \in \mathcal{L}} |\sum_{n \in N} p_i^n - P| \geq \epsilon$ **do**
3: $t = t + 1$
4: Network coordinator chooses BS_j in a round-robin order.
5: BS_j measures $\{I_k^n(p_{-j}^n)\}_{k \in \mathcal{K}_j}^{n \in \mathcal{N}}$.
6: BS_j determines $\mathbf{X}_j$ and $\mathbf{P}_j$ using (4.52) and (4.43b) respectively.
7: BS_j updates $\lambda_j(t)$ using (4.49).
8: **end while**

would be less than an accuracy threshold ϵ (step 2). This condition characterizes the satisfaction of the power constraints (4.43c). At each iteration, a BS is chosen in a round-robin manner to update its subcarrier, power, and bit level allocation subject to the measured interference from the other BSs (steps 4, 5, and 6). Using the new power settings, the chosen BS updates its Lagrange multiplier (step 7).

The DSPB algorithm takes advantage of two decomposition levels to overcome the exponential complexity of exhaustive search methods over the NQK binary variables. First, decoupling the original resource allocation problem (4.40)–(4.41) into subproblems, we decrease the exponential complexity to be linear in L. The linearity is due to the Lagrange multiplier update in (4.49). Second, the complexity $O((QK/L)^N)$ of subcarrier and rate allocation within each cell is decreased to $O(NQK/L)$ by dual decomposition in Section 4.2.2, as we came up with an optimization per subcarrier. In overall, the complexity is $O(NQK)$, linear in the number of subcarriers, bit levels, and users. On the other hand, the algorithm burdens some signalling overhead. The network coordinator notifies the BSs of their order in the algorithm and finally terminates the algorithm. At the end of every iteration, the chosen BS has to send to the coordinator its updated aggregate transmit power.

4.2.4 Performance evaluation

We consider downlink transmission in a network with four cells of radius $R = 1$ Km and 8 users. Every BS, located at the center of the corresponding cell, serves 2 users, randomly placed within the cell. The path loss (in dB) at a distance d from a BS is given by $L(d) = L(d_0) + 10\alpha \log_{10}(d/d_0)$, where for the reference point it is $d_0 = 50$ m, $L(50) = 0$, and the path loss exponent is $\alpha = 3.5$. The shadowing effect is modeled as an independent log-normal random variable with 8 dB standard deviation. The channel on each link is assumed to be Rayleigh fading, modeled by a six-tap impulse response with exponential power delay profile indicated by $ge^{-(l-1)}$, where $g = 1$ is the first path's average power gain and l is the path index. Moreover, the root-mean-square delay spread is 0.9 μs. The transmission budget of each BS is $P = 5$ W and the noise variance is assumed to be $\sigma_k^2 = -90$ dBm for all receivers.

The bit level on each subcarrier is chosen from the set $\mathcal{Q} = \{1, 2, ..., 5\}$, so that the corresponding SINR thresholds are $T_q = \{1, 3, 7, 15, 31\}$, respectively.

Firstly, to investigate the performance of DSPB for a typical number of subcarriers, e.g. $N = 64$, we show in Figure 4.14 the sum-rate achievement (in bits per OFDM symbol) of each cell versus the iteration number. It is seen that with the convergence of the transmit powers ($\epsilon = 0.1$), the cell sum rates attain their final values.

In the following, we compare, in the aforementioned setup, the performance of DSPB with the result obtained by solving the individual MIP (IMIP) (4.42)–(4.43) at individual BSs. The optimal solution in the primal domain of each IMIP is obtained calling the GNU linear programming kit (GLPK)[41]. In this scheme, similar to DSPB, beginning with uniform power allocation, the individual problems at BSs are solved optimally in a round-robin manner. As a lower bound, we also include the sum-rate values achieved from the iterative waterfilling algorithm (IWF) [42, 43] customized to OFDMA systems using joint subcarrier and power allocation as in [44] and [45]. Since the subcarrier rates in IWF are assumed to be continuous, we round off each achievable rate to the largest integer value not greater than that rate. We compare the overall sum-rate of the aforementioned schemes with different number of subcarriers, i.e. N. For each value of N, we obtain the sum rates for 50 realizations of the fading channel gains and show the average sum rates in Figure 4.16. We observe that DSPB outperforms both IMIP and IWF schemes. The performance gap between DSPB and IWF becomes larger as the number of subcarriers increases. This is due to the degradation effect of the rounding operation in IWF which increases with the number of subcarriers.

The performance difference between IMIP and DSPB is due to the fact that, in IMIP, each BS adopts the optimal solution in (4.42)–(4.43) to maximize its own sum rate. This optimal strategy most likely generates a large interference and therefore degrades the performance of other BSs significantly. However, in DSPB, each BS assigns each subcarrier as in (4.52), where $f_k^{n,q}$ can be written as $f_k^{n,q} = (q - \lambda_i p_i^n)$. In other words, in addition to the achieved rate q, DSPB also takes the required transmit power p_i^n into account in subcarrier allocation via the Lagrange multiplier acting as power price. Apparently, DSPB tends to minimize the generated interference on the other cells and therefore they undergo small rate degradation at the last iteration.

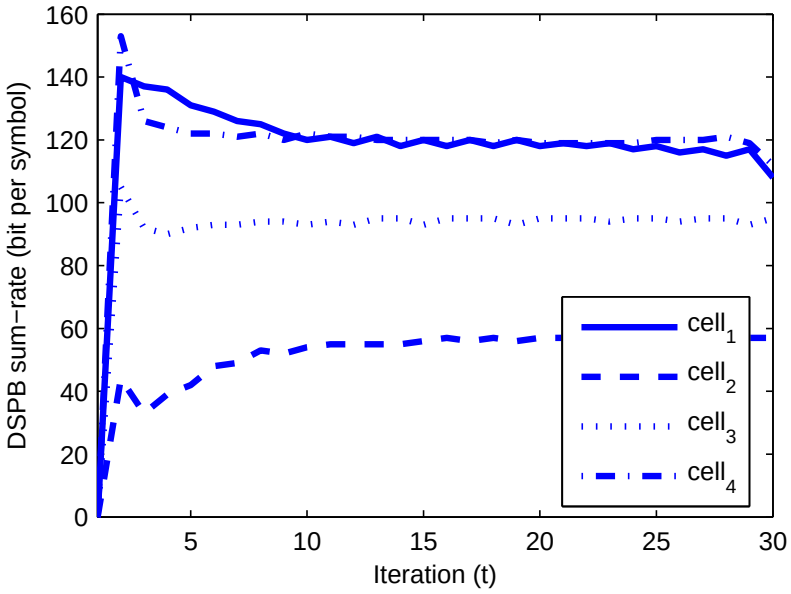


Figure 4.14: Sum-rate variation in DSPB

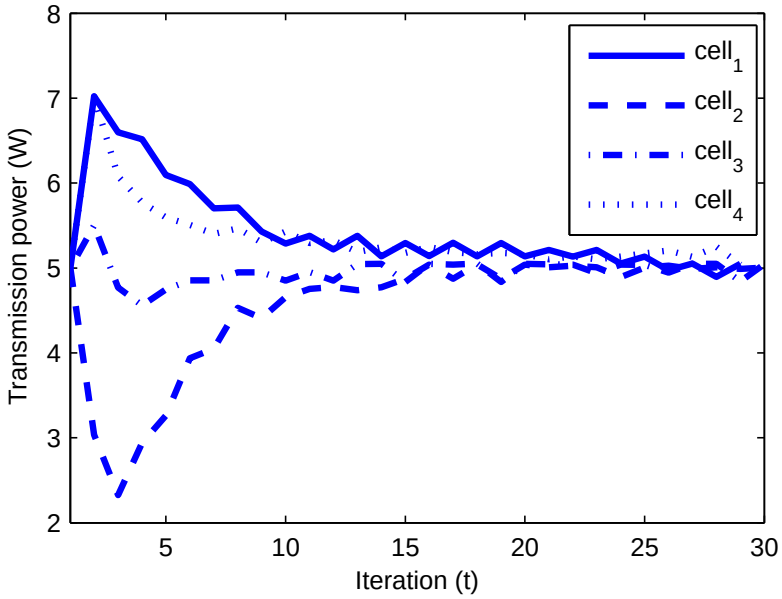


Figure 4.15: Power variation in DSPB

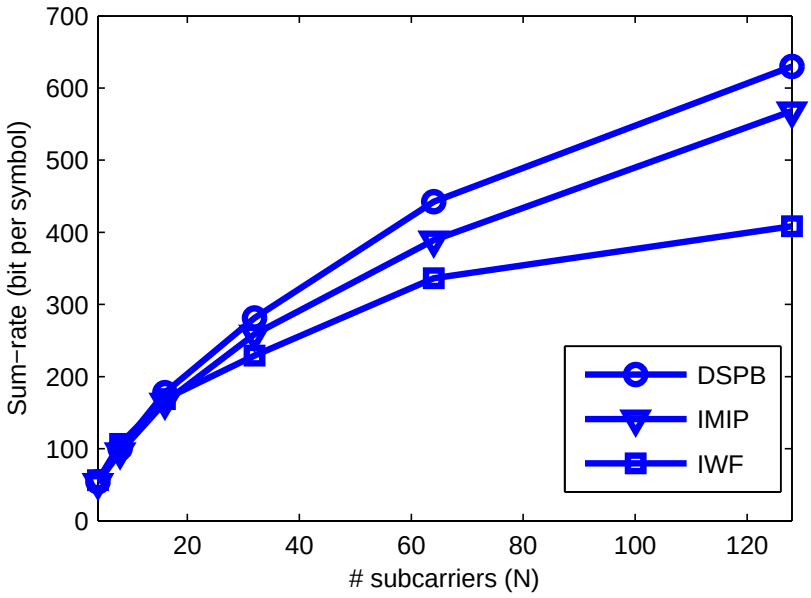


Figure 4.16: Average sum-rate in DSPB, IMIP, and IWF

5 Relay Assisted Infrastructure Sharing and Spectrum Sharing

5.1 Interference Neutralization

In this Section, we focus on a multi-user wireless network equipped with multiple relay nodes where some relays are more intelligent than other relay nodes. The intelligent relays are able to gather channel state information, perform linear processing and forward signals whereas the dumb relays is only able to serve as amplifiers. As the dumb relays are oblivious to the source and destination nodes, the wireless network can be modeled as a relay network with *smart instantaneous relay* only: the signals of the source-destination link arrive at the same time as the source-relay-destination link. Recently, instantaneous relaying is shown to improve the degrees-of-freedom of the network as compared to the classical cut-set bound. In the following, we study an achievable rate region and its boundary of the instantaneous interference relay channel in the scenario of (a) uninformed non-cooperative source-destination nodes (source and destination nodes are not aware of the existence of the relay and are non-cooperative) and (b) informed and cooperative source-destination nodes. Further, we examine the performance of interference neutralization: a relay strategy which is able to cancel interference signals at each destination node in the air. We observe that interference neutralization, although promised to achieve desired degrees-of-freedom, may not be feasible if the relay has limited power. Simulation results show that the optimal relay strategies improve the achievable rate region and provide better user-fairness in both uninformed non-cooperative and informed cooperative scenarios.

In a SISO interference relay channel (IRC), an example of two sources and two destinations shown in Figure 5.1, we denote the sources as S_i and destinations as D_i, $i = 1, \ldots, K$. The multi-antenna relay node is denoted as R. Denote the complex channel from S_i to D_j as h_{ji} and the complex channel vector from S_i to R as $\boldsymbol{g}_{ri}$ and from R to D_j as $\boldsymbol{g}_{jr}$. All channels are assumed to be independent identically distributed complex Gaussian variates, $\boldsymbol{g}_{ri}, \boldsymbol{g}_{jr} \in \mathbb{C}^{M\times 1}$, where M is the number of antennas at the relay. We assume a memoryless instantaneous relay [46, 47, 48] which has a linear processing matrix $\boldsymbol{R} \in \mathbb{C}^{M\times M}$. The signal received at R is:

$$\mathbf{y}_r = \sum_{j=1}^{K} \boldsymbol{g}_{rj} x_j + \mathbf{n}_r \tag{5.1}$$

where x_j are the transmit symbols from S_j which is assumed to be zero mean

proper Gaussian variable and has power constraint P_s^{max}, $\mathbb{E}\left[|x_j|^2\right] = P_j \leq P_s^{max}, j = 1, \ldots, K$. The noise at the relay is denoted as $\mathbf{n}_r$ which is assumed to be independent identically distributed proper Gaussian variables with zero mean and unit variance. The assumption of circularity for the transmit symbols simplifies the derivation as the achievable rate is the Shannon rate. For the degrees-of-freedom and achievable rates improvement with improper Gaussian signaling, please refer to [49, 50] respectively. The received signal at $D_j, j = 1, \ldots, K$, is

$$y_j = \sum_{l=1}^{K} \left(h_{jl} + \boldsymbol{g}_{jr}^H \boldsymbol{R} \boldsymbol{g}_{rl}\right) x_l + \boldsymbol{g}_{jr}^H \boldsymbol{R} \mathbf{n}_r + n_j. \tag{5.2}$$

For brevity, denote $\mathbf{p} = [P_1, \ldots, P_K]^T \in \mathbb{R}_+^{K \times 1}$. The Signal-to-Interference-and-Noise ratio at destination j is

$$\mathrm{SINR}_j(\boldsymbol{R}, \mathbf{p}) = \frac{|h_{jj} + \boldsymbol{g}_{jr}^H \boldsymbol{R} \boldsymbol{g}_{rj}|^2 P_j}{\sum_{l=1, l \neq j}^{K} |h_{jl} + \boldsymbol{g}_{jr}^H \boldsymbol{R} \boldsymbol{g}_{rl}|^2 P_l + \|\boldsymbol{g}_{jr}^H \boldsymbol{R}\|^2 + 1} \tag{5.3}$$

where $\|\boldsymbol{g}_{jr}^H \boldsymbol{R}\|^2$ is the amplified noise from R to D_j. The power constraint at the relay is

$$\mathbb{E}_{\mathbf{y}_r}\left[\mathrm{tr}\left(\boldsymbol{R} \mathbf{y}_r \mathbf{y}_r^H \boldsymbol{R}^H\right)\right] \leq P_r^{max}. \tag{5.4}$$

Note that $\mathbb{E}_{\mathbf{n}_r, x_j, j=1,\ldots,K}\left[\mathbf{y}_r \mathbf{y}_r^H\right] = \sum_{j=1}^{K} \boldsymbol{g}_{rj} \boldsymbol{g}_{rj}^H P_j + \mathbf{I}$. The power constraint is therefore rewritten as the following:

$$\mathrm{tr}\left(\boldsymbol{R} \boldsymbol{Q} \boldsymbol{R}^H\right) \leq P_r^{max} \tag{5.5}$$

where $\boldsymbol{Q} = \sum_{j=1}^{K} \boldsymbol{g}_{rj} \boldsymbol{g}_{rj}^H P_j + \mathbf{I}$.

Interference neutralization

In order to neutralize interference, the following $K(K-1)$ equations have to be satisfied at the same time:

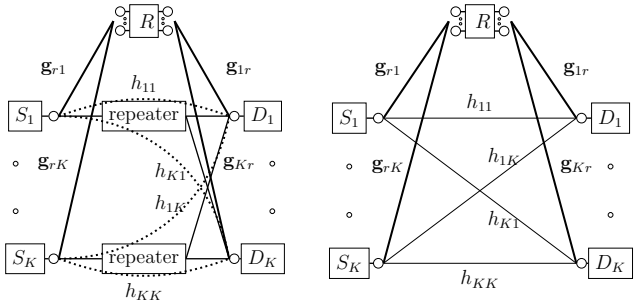
$$h_{ij} + \boldsymbol{g}_{ir}^H \boldsymbol{R} \boldsymbol{g}_{rj} = 0, \qquad i, j = 1, \ldots, K, i \neq j. \tag{5.6}$$

Let $\boldsymbol{G}_{dr} = [\boldsymbol{g}_{1r}, \boldsymbol{g}_{2r}, \ldots, \boldsymbol{g}_{Kr}]$ and $\boldsymbol{G}_{rt} = [\boldsymbol{g}_{r1}, \boldsymbol{g}_{r2}, \ldots, \boldsymbol{g}_{rK}]$. Denote $s_l = \boldsymbol{g}_{lr}^H \boldsymbol{R} \boldsymbol{g}_{rl}$. We have the interference neutralization requirement,

$$\boldsymbol{G}_{dr}^H \boldsymbol{R} \boldsymbol{G}_{rt} = \boldsymbol{S} \tag{5.7}$$

with

$$\boldsymbol{S} = \begin{bmatrix} s_1 & \ldots & -h_{1K} \\ -h_{21} & s_2 & \ldots \\ & \ddots & \\ \ldots & -h_{K(K-1)} & s_K \end{bmatrix}. \tag{5.8}$$



(a) a relay-assisted network (b) instantaneous relay network

Figure 5.1: The wireless relay-assisted network with layer one repeaters and one smart relay is shown in subfigure (a). The dotted lines demonstrate the equivalent links between source the destination taking into account of the presence of the repeaters. All paths from source to destination nodes take two time slots and links from source to relay and relay to destination take one time slot. The equivalent channel is established in subbfigure (b) by replacing the relay as an instantaneous relay and information through instantaneous relay arrive at destinations the same time as the direct links.

Note that $\boldsymbol{S}$ is a matrix with off-diagonal elements as the channel coefficients of the interference channel and diagonal elements as the optimization variables $\boldsymbol{s}$. As we sill show later, the optimization can be facilitated if the optimization variable is $\boldsymbol{S}$ instead of $\boldsymbol{s}$. This is due to the linear relationship between $\boldsymbol{R}$ and $\boldsymbol{S}$. However, we must stress that only the diagonal elements of $\boldsymbol{S}$, $\boldsymbol{s}$, are free to be optimized as $\boldsymbol{S}$ is constrained to the form in (5.8). Eqt. (5.8) can be rewritten to a more comprehensive form. We introduce a row selection matrix $\boldsymbol{T}$ which select the off-diagonal elements of $\boldsymbol{S}$ from the vector $\mathrm{vec}(\boldsymbol{S})$. For example when $K = 2$, we have $\boldsymbol{T} = [0, 1, 0, 0; 0, 0, 1, 0]$ and $\boldsymbol{T}\,\mathrm{vec}(\boldsymbol{S}) = [-h_{21}, -h_{12}]^T$. We have

$$\boldsymbol{T}\,\mathrm{vec}(\boldsymbol{S}) = -\boldsymbol{T}\,\mathrm{vec}(\mathbf{H}). \tag{5.9}$$

If IN is feasible (see later in Theorem Theorem 5.1 for feasibility issue), we can choose a relay matrix $\boldsymbol{R}$, which is a function of complex coefficients $\boldsymbol{s}$, that satisfies the IN constraint (5.7) and achieves the following SINR,

$$\mathrm{SINR}_j^{\mathrm{IN}}(\boldsymbol{S}, P_j) = \frac{|h_{jj} + \boldsymbol{g}_{jr}^H \boldsymbol{R}(\boldsymbol{S}) \boldsymbol{g}_{rj}|^2 P_j}{\|\boldsymbol{g}_{jr}^H \boldsymbol{R}(\boldsymbol{S})\|^2 + 1}. \tag{5.10}$$

In the following, we use the matrices $\boldsymbol{R}$ and $\boldsymbol{R}(\boldsymbol{S})$ interchangeably when we wish to emphasize the optimization parameter $\boldsymbol{S}$. Please see the next lemma for the formulation of $\boldsymbol{R}(\boldsymbol{S})$. In order to make sure the requirements in (5.7) are not over-

determined, we find the minimum number of antennas required in the relay such that (5.7) is feasible.

Lemma 5.1. *A sufficient condition of IN in* (5.7) *on the minimum number of antennas at the relay is $M \geq K$. For some target signal coefficients $s_1, \ldots, s_K \in \mathbb{C}$, the relay processing matrix $\boldsymbol{R}$ that satisfies the interference neutralization requirement* (5.7) *is determined by*

$$\mathrm{vec}(\boldsymbol{R}) = \left(\boldsymbol{G}_{rt}^T \otimes \boldsymbol{G}_{dr}^H\right)^{\dagger} \mathrm{vec}(\boldsymbol{S}) \tag{5.11}$$

where $\boldsymbol{G}_{rt}^T \otimes \boldsymbol{G}_{dr}^H$ is assumed to have full row rank[1]*,* $\mathrm{rank}\left(\boldsymbol{G}_{rt}^T \otimes \boldsymbol{G}_{dr}^H\right) = K^2$.

Note that when $M < K$, the system in (5.7) has more equations than unknowns and the consistency of the system depends heavily on the particular channel realizations. For $M \geq K$, we write $\mathrm{vec}(\boldsymbol{R})$ as

$$\mathrm{vec}(\boldsymbol{R}) = \left(\boldsymbol{G}_{rt}^T \otimes \boldsymbol{G}_{dr}^H\right)^{\dagger} \mathrm{vec}(\boldsymbol{S}) + \boldsymbol{b} \tag{5.12}$$

where $\boldsymbol{b} \in \mathbb{C}^{M^2}$ is in the null space of $\boldsymbol{G}_{rt}^T \otimes \boldsymbol{G}_{dr}^H$. With (5.12), one must optimize both $\boldsymbol{b}$ and $\mathrm{vec}(\boldsymbol{S})$ in order to optimize $\mathrm{vec}(\boldsymbol{R})$. For simplicity of presentation, from now on, we set $M = K$ and obtain [51, P.35]

$$\mathrm{vec}(\boldsymbol{R}) = \left(\boldsymbol{G}_{rt}^T \otimes \boldsymbol{G}_{dr}^H\right)^{-1} \mathrm{vec}(\boldsymbol{S}). \tag{5.13}$$

If there is no power constraint at the relay, given any target signal coefficients s_i, we can construct a relay matrix $\boldsymbol{R}$ as in (5.13) such that the IN requirement is satisfied. Otherwise, any performance metrics, e.g. sum rate, subject to the IN and power constraints, are optimized in a dimension of K by optimizing the complex coefficients $\boldsymbol{s} = [s_1, \ldots, s_K]^T$.

5.1.1 The Pareto boundary optimization problem

The achievable rate region of an instantaneous AF relay is defined to be the set of rate tuple achieved by all possible relay processing matrix $\boldsymbol{R}$ satisfying the power constraint (5.5):

$$\mathcal{R} = \bigcup_{\boldsymbol{R}:\mathrm{tr}(\boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}^H) \leq P_r} \left(C(\mathrm{SINR}_1(\boldsymbol{R})), \ldots, C(\mathrm{SINR}_K(\boldsymbol{R}))\right) \tag{5.14}$$

where $C(x) = \log_2(1 + x)$. Similarly, we define the achievable rate region of an instantaneous AF relay with IN to be the set of rate tuple achieved by all possible relay processing matrix $\boldsymbol{R}$ satisfying the IN constraint (5.13) and the power

[1] The assumption of full column rank is not difficult to fulfill as the elements in the channel matrix are statistically independent.

constraint (5.5):

$$\mathcal{R}^{\mathrm{IN}} = \bigcup_{\substack{\boldsymbol{R}:\mathrm{tr}\left(\boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}^H\right)\le P_r,\\ \mathrm{vec}(\boldsymbol{R})=\left(\boldsymbol{G}_{rt}^T\otimes\boldsymbol{G}_{dr}^H\right)^{-1}\mathrm{vec}(\boldsymbol{S}),\\ \boldsymbol{T}\,\mathrm{vec}(\boldsymbol{S})=-\boldsymbol{T}\,\mathrm{vec}(\mathbf{H})}} \left(C(\mathrm{SINR}_1(\boldsymbol{R})),\ldots,C(\mathrm{SINR}_K(\boldsymbol{R}))\right). \tag{5.15}$$

Note that the IN requirement gives a smaller feasible set and thus $\mathcal{R}^{IN} \subset \mathcal{R}$. However, the motivation of the study of $\mathcal{R}^{\mathrm{IN}}$ is two-fold: (a) intuitively when both transmit power at S and R is very large, the optimal relay strategy is to neutralize interference, so as to transmit at the maximum DOF; but the performance of other SNR regimes is yet to be studied. (b) the characterization of $\mathcal{R}^{\mathrm{IN}}$ in the instantaneous IRC with direct link between S-D pairs is still an open problem. The outer boundary of $\mathcal{R}$ ($\mathcal{R}^{\mathrm{IN}}$) – the Pareto boundary (PB) of $\mathcal{R}$ ($\mathcal{R}^{\mathrm{IN}}$) – is a set of operating points at which one user cannot increase its own rate without simultaneously decreasing other users rate.

Definition 5.1 ([52, 53])**.** A rate-tuple $(r_1,\ldots,r_K)$ is Pareto optimal if there is no other rate tuple $(q_1,\ldots,q_K)$ such that $(q_1,\ldots,q_K) \ge (r_1,\ldots,r_K)$ and $(q_1,\ldots,q_K) \ne (r_1,\ldots,r_K)$[2].

By definition, the operation points on the PB can be evaluated by maximizing one user's rate while keeping other users' rates constant. Other optimization techniques for PB evaluation has been proposed [54, 55]. Here, the PB problem is formulated as the maximization of SINR_1 subject to the constraints on $\mathrm{SINR}_j \ge \gamma_j$ for some pre-determined target SINR values γ_j, $j = 2,\ldots,K$.

Problem Statement 1. The Pareto boundary of $\mathcal{R}$ (5.14) is a set of rate tuple $\left(C(\gamma_1^{\#}), C(\gamma_2),\ldots,C(\gamma_K)\right)$ where $\gamma_1^{\#}$ is the optimal objective value and $\gamma_j, j = 2,\ldots K$ are the constraints of the following optimization problem.

$$(\mathrm{PB})\begin{cases} \max\limits_{\boldsymbol{R}\in\mathbb{C}^M,\mathbf{p}\in\mathbb{R}_+^{K\times 1}} & \mathrm{SINR}_1(\boldsymbol{R},\mathbf{p}) \\ \text{s.t.} & \mathrm{SINR}_j(\boldsymbol{R},\mathbf{p}) \ge \gamma_j, \quad j=2,\ldots,K, \\ & \mathrm{tr}\left(\boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}^H\right) \le P_r^{max}. \end{cases} \tag{5.16}$$

Similarly, we formulate the PB optimization problem with IN in the following. To this end, we combine the first two constraints in the feasibility set of (5.15). Using the fact that $\mathrm{tr}\left(\boldsymbol{A}\,\mathbf{B}\,\boldsymbol{C}\boldsymbol{D}\right) = \mathrm{vec}(\boldsymbol{D}^T)^T\left(\boldsymbol{C}^T\otimes\boldsymbol{A}\right)\mathrm{vec}(\mathbf{B})$ [51], we have

$$\begin{aligned}\mathrm{tr}(\boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}^H) &= \mathrm{vec}(\boldsymbol{R})^H\left(\boldsymbol{Q}^T\otimes\mathcal{I}\right)\mathrm{vec}(\boldsymbol{R}) \\ &\overset{(a)}{=} \mathrm{vec}(\boldsymbol{S})^H \underbrace{\left(\left(\mathrm{diag}(\mathbf{p}) + \boldsymbol{G}_{rt}^{-*}\boldsymbol{G}_{rt}^{-T}\right)\otimes\boldsymbol{G}_{dr}^{-1}\boldsymbol{G}_{dr}^{-H}\right)}_{\tilde{\boldsymbol{Q}}}\mathrm{vec}(\boldsymbol{S})\end{aligned} \tag{5.17}$$

[2]The inequality is component-wise.

where (a) is due to (5.13) and $\boldsymbol{G}_{rt}^{-*}\boldsymbol{Q}^T\boldsymbol{G}_{rt}^{-T} = \mathrm{diag}(\mathbf{p}) + \boldsymbol{G}_{rt}^{-*}\boldsymbol{G}_{rt}^{-T}$.

Problem Statement 2. The Pareto boundary of $\mathcal{R}^{\mathrm{IN}}$ (5.15) is a set of rate tuple $\left(C(\gamma_1^{\#}), C(\gamma_2), \ldots, C(\gamma_K)\right)$ where $\gamma_1^{\#}$ is the optimal objective value and $\gamma_j, j = 2, \ldots K$ are the constraints of the following optimization problem.

$$
\begin{aligned}
(\mathrm{PB}-\mathrm{IN}): \quad \max_{\boldsymbol{S}\in\mathbb{C}^K, \mathbf{p}\in\mathbb{R}_+^{K\times 1}} \quad & \mathrm{SINR}_1^{\mathrm{IN}}(\boldsymbol{S}, P_1) && \text{(5.18a)}\\
\text{s.t.} \quad & \mathrm{SINR}_j^{\mathrm{IN}}(\boldsymbol{S}, P_j) \geq \gamma_j,\ j = 2, \ldots, K, && \text{(5.18b)}\\
& \mathrm{vec}(\boldsymbol{S})^H \tilde{\boldsymbol{Q}}\, \mathrm{vec}(\boldsymbol{S}) \leq P_r^{max}, && \text{(5.18c)}\\
& \boldsymbol{T}\,\mathrm{vec}(\boldsymbol{S}) = -\boldsymbol{T}\,\mathrm{vec}(\mathbf{H}). && \text{(5.18d)}
\end{aligned}
\tag{5.18}
$$

Before we show the optimization methods of the aforementioned problems, the feasibility issue of (5.18a) needs to be addressed.

Theorem 5.1. *A necessary and sufficient condition for the feasibility of interference neutralization - satisfying* (5.18c) *and* (5.18d) *simultaneously - is*

$$
\left(\boldsymbol{T}\,\mathrm{vec}(\mathbf{H})\right)^H \left(\boldsymbol{T}\tilde{\boldsymbol{Q}}^{-1}\boldsymbol{T}^H\right)^{-1} \boldsymbol{T}\,\mathrm{vec}(\mathbf{H}) \leq P_r^{max}, \tag{5.19}
$$

where $\tilde{\boldsymbol{Q}} = \left(\mathrm{diag}(\mathbf{p}) + \boldsymbol{G}_{rt}^{-*}\boldsymbol{G}_{rt}^{-T}\right) \otimes \left(\boldsymbol{G}_{dr}^{-1}\boldsymbol{G}_{dr}^{-H}\right)$. *A feasible solution is*

$$
\mathrm{vec}(\boldsymbol{S}) = \boldsymbol{F}\left(\boldsymbol{x}_n + (\boldsymbol{T}\boldsymbol{F})^{\dagger}\boldsymbol{T}\,\mathrm{vec}(\mathbf{H})\right) \tag{5.20}
$$

where $\boldsymbol{x}_n \in \mathcal{N}(\boldsymbol{T}\boldsymbol{F})$ *and* $\tilde{\boldsymbol{Q}}^{-1} = \boldsymbol{F}\boldsymbol{F}^H$.

Note that the left hand side of the condition (5.19) is the power of the relay matrix which only performs IN (with $\boldsymbol{x}_n = \mathbf{0}$ in (5.20)) and only depends on channel coefficients whereas the right hand side is the relay power constraint. The importance of Theorem Theorem 5.1 lies in the practicality of IN feasibility verification. With the information of channel qualities $\mathbf{H}, \boldsymbol{G}_{rt}, \boldsymbol{G}_{dr}$, transmit power at S $\mathbf{p}$ and transmit power constraint at relay P_r^{max}, we can immediately check whether IN is feasible or not. Further, if IN is feasible, there is always one feasible solution in (5.20) by setting $\boldsymbol{x}_n = \mathbf{0}_{K\times 1}$; if there is excess power at relay, we can optimize $\boldsymbol{x}_n$, and in turn $\mathrm{vec}(\boldsymbol{S})$, to improve system performance, which is the objective of the succeeding section.

5.1.2 The optimal relay strategies with uninformed S-D nodes

In this section, we analyze the PB problems taking into consideration that the S and D nodes are uninformed of the presence of the relay nodes and therefore do not optimize their transmit power values: $\mathbf{p} = \mathbf{p}_0$ for some pre-determined power values $\mathbf{p}_0$. If the S-D pairs are non-cooperative, then each S transmits at full power and

thus $\mathbf{p} = \mathbf{1}\, P_s^{max}$. Given the transmit power values of the source nodes, we maximize the achievable rate by choosing the relax matrix. The PB problems are formulated into quadratically constrained quadratic programs (QCQP) and are then relaxed to semi-definite programs (SDP). The relaxed problems are convex and can be solved using efficient convex optimization tools such as CVX [56]. We show that in some scenarios, the optimal solution of the relaxed problems attains the optimality of the original problems. In other words, the convex optimization methods solve the original problem efficiently in such scenarios (Please refer to Corollary Remark 5.1 and Lemma Lemma 5.4 for details). The procedures used to obtain the PB are summarized in Algorithm 4.

Algorithm 4 The pseudo-code for the Pareto boundary optimization in (5.16) and (5.18a).

1: **for** $j = 2 \to K$ **do**
2: Compute the single-user-point of user j: $\gamma_j^{max} = \max_{\boldsymbol{R}} \frac{|h_{jj}+\boldsymbol{g}_{jr}^H \boldsymbol{R} \boldsymbol{g}_{rj}|^2 P_j}{\|\boldsymbol{g}_{jr}^H \boldsymbol{R}\|^2+1}$, in which the user j is the only user in the system with no interference.
3: For a predefined integer N, let $\mathbb{V}_j = \left\{0, \frac{\gamma_j^{max}}{N-1}, \frac{2\gamma_j^{max}}{N-1}, \ldots, \gamma_j^{max}\right\}$.
4: **end for**
5: Define a tuple $\boldsymbol{s}$ to be a vector of possible values of γ_j, $\boldsymbol{s} = [\gamma_2, \ldots, \gamma_K]$ and $\boldsymbol{s} \in \mathbb{S} = \mathbb{V}_2 \times \mathbb{V}_3 \times \cdots \times \mathbb{V}_K$.
6: **for** each $\boldsymbol{s} \in \mathbb{S}$ **do**
7: With input parameter $\boldsymbol{s}$, matrix $\boldsymbol{S}$ is defined as in (5.8). Solve the optimization problems (5.22) for general relay processing matrix optimization or (5.24) for relay processing matrix optimization with interference neutralization.
8: **if** the optimization problem is feasible **then**
9: the optimal value $\gamma_1^{\dagger}$ and $\boldsymbol{s}$ form a point on the Pareto boundary, which is a $K-1$ dimension hyper-surface, in the K dimensional space.
10: **else**
11: the values of $\boldsymbol{s}$ are unachievable subject to the constraints.
12: **end if**
13: **end for**

The Pareto boundary: general relay optimization

By introducing an auxiliary variable, we formulate the PB optimization problem with general relay processing matrix as a QCQP. The optimization variable is of dimension $M^2 + 1$ as compared to M^2 number of elements in the relay matrix. Nevertheless, this provides a more structural formulation and amends the analysis as shown in the sequel.

Lemma 5.2. *The Pareto boundary of an IRC with instantaneous AF relay* (5.16) *is a rate tuple* $(C(\gamma_1^{\#}), C(\gamma_2), \ldots, C(\gamma_K))$ *in which* $\gamma_1^{\#} = \frac{\bar{\boldsymbol{v}}^H \boldsymbol{X}_{11} \bar{\boldsymbol{v}}}{\bar{\boldsymbol{v}}^H \boldsymbol{X}_{12} \bar{\boldsymbol{v}}}$ *and* $\bar{\boldsymbol{v}}$ *is the op-*

timal solution of the following optimization problem. The values $\gamma_j, j = 2, \ldots, K$ contribute to the constraints of the optimization problem.

$$\begin{cases} \max\limits_{\boldsymbol{v}\in\mathbb{C}^{(M^2+1)\times 1}} & \dfrac{\boldsymbol{v}^H \boldsymbol{X}_{11} \boldsymbol{v}}{\boldsymbol{v}^H \boldsymbol{X}_{12} \boldsymbol{v}} \\ s.t. & \dfrac{\boldsymbol{v}^H \boldsymbol{X}_{j1} \boldsymbol{v}}{\boldsymbol{v}^H \boldsymbol{X}_{j2} \boldsymbol{v}} \geq \gamma_j, \quad j = 2, \ldots, K, \\ & \boldsymbol{v}^H \boldsymbol{X}_3 \boldsymbol{v} \leq 0. \end{cases} \tag{5.21}$$

The matrices, for $i = 1, \ldots, K$, are given by

$$\boldsymbol{X}_{i1} = \begin{bmatrix} (\boldsymbol{g}_{ri} \otimes \boldsymbol{g}_{ir}^*)^* (\boldsymbol{g}_{ri} \otimes \boldsymbol{g}_{ir}^*)^T & (\boldsymbol{g}_{ri} \otimes \boldsymbol{g}_{ir}^*)^T h_{ii} \\ h_{ii}^* (\boldsymbol{g}_{ri} \otimes \boldsymbol{g}_{ir}^*)^T & |h_{ii}|^2 \end{bmatrix} P_i,$$

$$\boldsymbol{X}_{i2} = \sum_{l\neq i}^{K} \begin{bmatrix} (\boldsymbol{g}_{rl} \otimes \boldsymbol{g}_{ir}^*)^* (\boldsymbol{g}_{rl} \otimes \boldsymbol{g}_{ir}^*)^T & (\boldsymbol{g}_{rl} \otimes \boldsymbol{g}_{ir}^*)^T h_{il} \\ h_{il}^* (\boldsymbol{g}_{rl} \otimes \boldsymbol{g}_{ir}^*)^T & |h_{il}|^2 \end{bmatrix} P_l + \begin{bmatrix} \mathcal{I}_M \otimes (\boldsymbol{g}_{ir} \boldsymbol{g}_{ir}^H) & \boldsymbol{0}_{M^2\times 1} \\ \boldsymbol{0}_{1\times M^2} & 1 \end{bmatrix},$$

$$\boldsymbol{X}_3 = \begin{bmatrix} \boldsymbol{Q}^T \otimes \mathcal{I} & \boldsymbol{0}_{M^2\times 1} \\ \boldsymbol{0}_{1\times M^2} & -P_r \end{bmatrix}.$$

Define $\boldsymbol{V} = \boldsymbol{v}\boldsymbol{v}^H$, we obtain the following convex problem after removing the constraint $\text{rank}(\boldsymbol{V}) = 1$:

$$\begin{cases} \max\limits_{\boldsymbol{V}\in\mathbb{C}^{(M^2+1)}, \boldsymbol{V}\succeq 0} & \text{tr}\left(\boldsymbol{X}_{11}\boldsymbol{V}\right) \\ \text{s.t.} & \text{tr}\left(\boldsymbol{X}_{12}\boldsymbol{V}\right) = 1, \\ & \text{tr}\left(\left(\boldsymbol{X}_{j1} - \gamma_j \boldsymbol{X}_{j2}\right)\boldsymbol{V}\right) \geq 0, \quad j = 2, \ldots, K, \\ & \text{tr}\left(\boldsymbol{X}_3 \boldsymbol{V}\right) \leq 0 \end{cases} \tag{5.22}$$

which is a semi-definite program (SDP) as matrices $\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \boldsymbol{X}_3$ are Hermitian and can be solved efficiently using SDP solvers.

Remark 5.1. By [57, Theorem 3.2] the rank of the optimum solution of (5.22) is smaller than $\sqrt{K+1}$. In the scenario of $K = 2$ S-D pairs, the rank of the optimal solution is one which means that the relaxation is tight and the obtained solution is the global optimal solution of (5.21). In occasions when the optimal solution returned by CVX is not rank-one, one can find a vector which satisfies the same constraint and objective values in (5.22) using the rank-one reduction procedures [58, Theorem 2.3]. Such vector is thus one of the global optimal solutions of (5.21). For more S-D pairs, the result of SDP in (5.22) is not rank-one. However, one can apply randomization approximation techniques [59] to approximate the optimal solution of (5.21).

The Pareto boundary with IN

We manipulate the problem (5.18a) in the same fashion as in (5.21). The optimization problem (5.18a) has one more constraint (the IN constraint) and is thus optimized in a smaller dimension K^2+1 rather than M^2+1. Note that the condition given in Theorem Theorem 5.1 is a necessary and sufficient condition for the feasibility of IN (non-empty feasible set in (5.15)). In other words, if the condition is not satisfied, then the optimization problem in (5.23) is not feasible regardless of the target SINR values $\gamma_2, \ldots, \gamma_K$.

Lemma 5.3. *The Pareto boundary of an IRC with instantaneous AF relay and IN* (5.18a) *is a rate tuple* $(C(\gamma_1^{\#}), C(\gamma_2), \ldots, C(\gamma_K))$ *in which* $\gamma_1^{\#} = \bar{\mathbf{y}}^H \hat{\mathbf{B}}_{i1} \bar{\mathbf{y}}$ *and* $\bar{\mathbf{y}}$ *is the optimal solution of the following optimization problem. The values* $\gamma_j, j = 2, \ldots, K$ *contribute to the constraints of the optimization problem.*

$$\begin{cases} \max\limits_{\mathbf{y}\in\mathbb{C}^{(K^2+1)\times 1}} & \mathbf{y}^H \hat{\mathbf{B}}_{11}\, \mathbf{y} \\ \quad s.t. & \mathbf{y}^H \hat{\mathbf{B}}_{12}\, \mathbf{y} = 1, \\ & \mathbf{y}^H \hat{\mathbf{B}}_j\, \mathbf{y} \geq 0, \qquad j = 2, \ldots, K, \\ & \mathbf{y}^H \hat{\boldsymbol{D}}_3\, \mathbf{y} \leq 0, \qquad \mathbf{y}^H \hat{\boldsymbol{D}}_4\, \mathbf{y} = 0. \end{cases} \tag{5.23}$$

where

$$\hat{\mathbf{B}}_{11} = \begin{bmatrix} \boldsymbol{L}^T \mathrm{e}_1 \mathrm{e}_1^T \boldsymbol{L} & \boldsymbol{L}^T \mathrm{e}_1 h_{11} \\ h_{11}^* \mathrm{e}_1^T \boldsymbol{L} & |h_{11}|^2 \end{bmatrix} P_1, \quad \hat{\mathbf{B}}_{12} = \begin{bmatrix} \boldsymbol{G}_{rt}^{-*} \boldsymbol{G}_{rt}^{-T} \otimes \mathrm{e}_1 \mathrm{e}_1^T & \mathbf{0}_{K^2\times 1} \\ \mathbf{0}_{1\times K^2} & 1 \end{bmatrix},$$

$$\hat{\mathbf{B}}_j = \begin{bmatrix} \boldsymbol{L}^T \mathrm{e}_j \mathrm{e}_j^T \boldsymbol{L} - \gamma_j \left(\boldsymbol{G}_{rt}^{-*} \boldsymbol{G}_{rt}^{-T} \otimes \mathrm{e}_j \mathrm{e}_j^T \right) & \boldsymbol{L}^T \mathrm{e}_j\, \mathbf{h}_{jj} \\ \boldsymbol{h}_{jj}^* \mathrm{e}_j^T \boldsymbol{L} & |h_{jj}|^2 - \gamma_j \end{bmatrix}, \hat{\boldsymbol{D}}_3 = \begin{bmatrix} \tilde{\boldsymbol{Q}} & \mathbf{0}_{K^2\times 1} \\ \mathbf{0}_{1\times K^2} & -P_r^{max} \end{bmatrix},$$

$$\hat{\boldsymbol{D}}_4 = \begin{bmatrix} \boldsymbol{T}^H \boldsymbol{T} & \boldsymbol{T}^H \boldsymbol{T}\, \mathrm{vec}(\mathbf{H}) \\ \mathrm{vec}(\mathbf{H})^H \boldsymbol{T}^H \boldsymbol{T} & \mathrm{vec}(\mathbf{H}) \boldsymbol{T}^H \boldsymbol{T}\, \mathrm{vec}(\mathbf{H}) \end{bmatrix}.$$

Let $\boldsymbol{Y} = \mathbf{y}\, \mathbf{y}^H$. Using SDR techniques, the problem in (5.23) can be relaxed to the following problem by dropping the rank one constraint on Y. Problem (5.24) is a convex problem, in particular a semi-definite program, which can be solved efficiently.

$$\begin{cases} \max\limits_{\boldsymbol{Y}\in\mathbb{C}^{K^2+1}, \boldsymbol{Y}\succeq 0} & \mathrm{tr}\left(\hat{\mathbf{B}}_{11} \boldsymbol{Y} \right) \\ \quad \text{s.t.} & \mathrm{tr}\left(\hat{\mathbf{B}}_{12} \boldsymbol{Y} \right) = 1, \\ & \mathrm{tr}\left(\hat{\mathbf{B}}_j \boldsymbol{Y} \right) \geq 0, \qquad j = 2, \ldots, K, \\ & \mathrm{tr}\left(\hat{\boldsymbol{D}}_3 \boldsymbol{Y} \right) \leq 0,\ \mathrm{tr}\left(\hat{\boldsymbol{D}}_4 \boldsymbol{Y} \right) = 0. \end{cases} \tag{5.24}$$

Note that, if the optimal solution in (5.24) is rank one, then such solution solves (5.23) optimally. If the optimal solution in (5.24) is not rank one, then the optimality

of (5.23) can no longer be guaranteed. In the following, we characterize the rank of the optimal solution of (5.24).

Lemma 5.4. *The optimal solution of K-user IRC Pareto boundary problem with IN in* (5.24)*, $\tilde{\boldsymbol{Y}}$, satisfies*

$$\text{rank}(\tilde{\boldsymbol{Y}}) \leq \sqrt{K+1}. \tag{5.25}$$

The optimization methods and results for (5.24) are similar to Corollary Remark 5.1 and shall not be repeated here. In the following section, we provide numerical evidence of performance gain of a relay introduction to a SISO interference channel. In particular, in a setting of uninformed source nodes, we show the rate improvement of solely introducing and optimizing the relay strategy whereas in a setting of informed S-D nodes, we compare the rate performance of the general relay optimization and the relay optimization with IN to the rate region of a SISO IC.

5.1.3 Simulation results

For illustrative purposes, we let $K = M = 2$. We assume that each element in the channel matrices $\mathbf{H}, \boldsymbol{G}_{rt}, \boldsymbol{G}_{dr}$ is an independent identically distributed complex Gaussian variable with zero mean and unit variance. In Section 5.1.3, we simulate the achievable rates of the SISO IC (marked as squares). For comparison, we simulate the achievable rates of the fully connected IRC with the same S-D nodes, by introducing a relay, equipped with 2 antennas, to the aforementioned IC, and the relay can choose to enforce IN (marked with asterisks) or not (marked with triangles). We show that optimized relay strategies improve achievable rate regions. In Section 5.1.3 and 5.1.3, we compare the average sum rate and proportional fairness utility achieved by optimized relay strategies and by power allocation on the IC respectively. In Section 5.1.3, we illustrate the sum rate performance and proportional fairness utility when the transmit power constraint at source nodes and relay nodes vary.

Rate region improvement

In Figure 5.2, we plot the achievable rate region of a two-user SISO IC with transmit power constraint at each source node $P_s^{max} = 10$ dB. Introducing an instantaneous relay, equipped with 2 antennas, we obtain an IRC. We set the relay power constraint as $P_r^{max} = 20$ dB. The achievable rates achieved by general relay strategies and IN outperform the IC case. The black arrows originate from the Nash Equilibrium point: the rate points in which both users transmit with full power. The north-east side of the arrows mark the rate region improved by the relay, in the scenario of uninformed source nodes. This validates our intuition that optimized relay strategies can improve achievable rates of the system even if the source nodes are oblivious to the existence of the relay and do not change their transmit power.

Further note that the single user points achieved in IRC with general relay strategies always outperform the single user points in a SISO IC. It demonstrates that the relay not only is capable of reducing interference in the system but also forwarding the desired signal to the destinations.

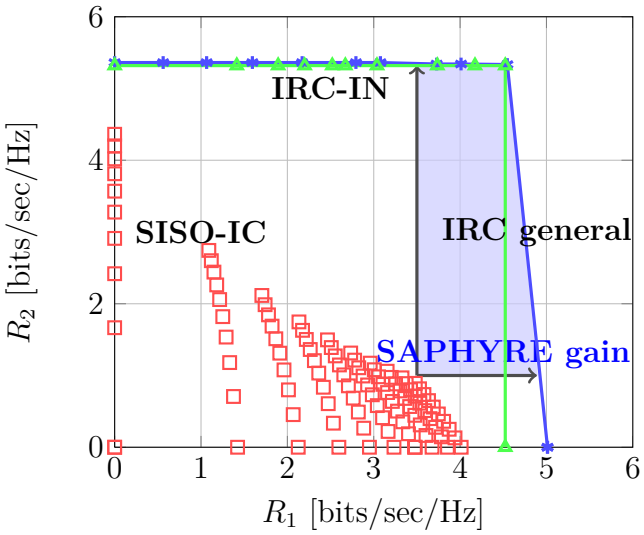


Figure 5.2: The rate improvement of relay optimization on a two-user SISO IRC with $K = M = 2$, $P_r^{max} = 20$ dB, $P_s^{max} = 10$ dB. The arrow marks the increment of rate region by introducing a relay into the system and optimizing the relay strategy over the Nash equilibrium point.

Average sum rate improvement

In Figure 5.3, we show the maximum sum rate achieved by general relay strategies, IN and power allocation on the IC, averaged over 100 independent channel realizations. The power constraint at the source node is assumed to be $P_s^{max} = 10$ dB and we increase the relay transmit power from 5 dB to 25 dB. We observe that the optimized relay strategy without IN always outperform the maximum sum rate of the IC, demonstrating that an instantaneous smart relay can improve average sum rate performance. Further, we observe that the performance of IN is limited by the relay transmit power. Although IN is analytically appealing, there are limitations of the implementation of IN. Such scenarios include strong interference channels in which the receivers have strong interference from other transmitters in the system. In this case, more power at the relay may be required to completely null out interference and if such power is not available to the relay, then IN is not feasible. On the other hand, if the strength of the interference channel is not strong, enforcing IN, the relay loses its optimization degrees of freedom and may not be able to achieve some operating points as the general relay strategies would achieve.

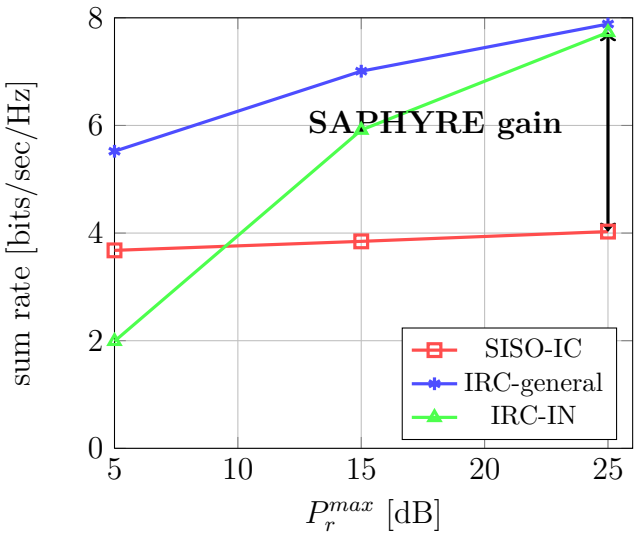


Figure 5.3: The average sum rate a two-user SISO IRC with $K = M = 2$, $P_s^{max} = 10$ dB. The optimized relay strategies improve the average sum rate of the system significantly.

Proportional fairness improvement

While average sum rate is an important system performance measure, user fairness holds importance in many applications. In a 2-user system, the proportional fairness utility is defined as the area of a rectangle with one vertex as the operating point (R_1, R_2) and the opposite vertex as the threat point (or the Nash Equilibrium point) (R_1^{NE}, R_2^{NE}),

$$\left(R_1 - R_1^{NE}\right)\left(R_2 - R_2^{NE}\right). \tag{5.26}$$

The idea is to choose an operating point on the Pareto boundary which provides *good* improvement for both users (or loosely speaking as far away from the threat point as possible). The proportional fairness utility achieved by a scheme with Pareto boundary $\mathcal{P}$ is defined as

$$r_{pf} = \max_{\substack{(R_1,R_2)\in\mathcal{P},\\ R_1\geq R_1^{NE}, R_2\geq R_2^{NE}}} \left(R_1 - R_1^{NE}\right)\left(R_2 - R_2^{NE}\right) \tag{5.27}$$

where $\mathcal{P}$ denotes the Pareto boundary. In Figure 5.4, we illustrate the average proportional fairness utility. We observe that the optimized general relay strategies and optimized IN strategies, provide promising proportional fairness and better sum rate performance compared to IC.

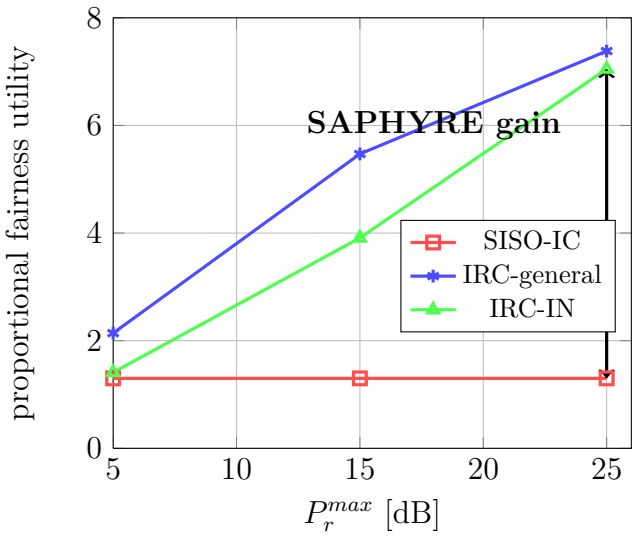


Figure 5.4: The average proportional fairness utility $(R_1 - R_1^{NE})(R_2 - R_2^{NE})$ of a two-user SISO IRC with $K = M = 2$, $P_s^{max} = 10$ dB. The optimized relay strategies improve fairness of the system significantly.

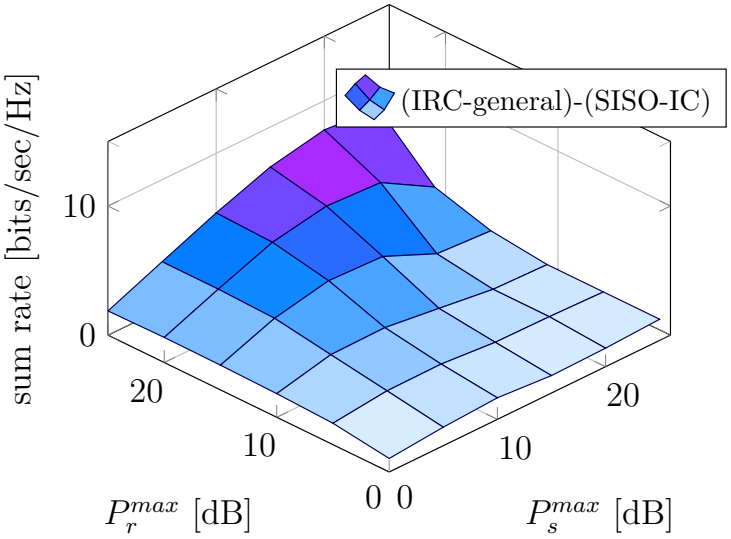
Performance measures in terms of transmit power constraints

It is interesting to observe that the performance of optimized relay strategies depend on both P_s^{max} and P_r^{max}. This is due to the amplify-and-forward nature of the relay. If the transmit power from source nodes is high, the relay can spend less power on amplification of signals due to the relay power constraint.

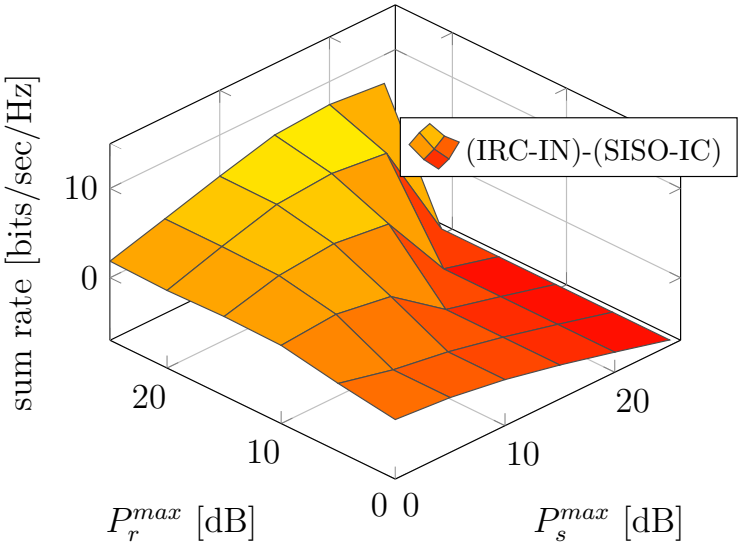
In Figure 5.5, we plot the maximum sum rate achieved by optimized relay strategy minus the maximum sum rate of the SISO-IC with general relay strategy in Figure 5.5(a) and with IN in Figure 5.5(b). Note that the feasibility conditions of IN, shown in Theorem Theorem 5.1, is validated in Figure 5.5. For a fixed relay power P_r^{max}, when the source power increases such that the conditions are violated, IN is not feasible and the achievable rate is zero and consequently the corresponding values in Figure 5.5(b) are negative. For the general relay optimization, for a fixed relay power and increasing source power, the rate performance is not always better because the interference power increases with the source power and the relay may not have enough power to manage interference and amplify desired signals simultaneously.

In Figure 5.6, we show the maximum proportional fairness utility achieved by optimized relay strategies and IC. When both source power and relay power are abundant, the fairness is desirable. However, when the source power (which is also the strength of interference) is relatively stronger than the relay power, the fairness achieved by the relay strategies is overtaken by the proportional fairness utility

achieved by IC.



(a) Improvement of IRC-general over SISO-IC

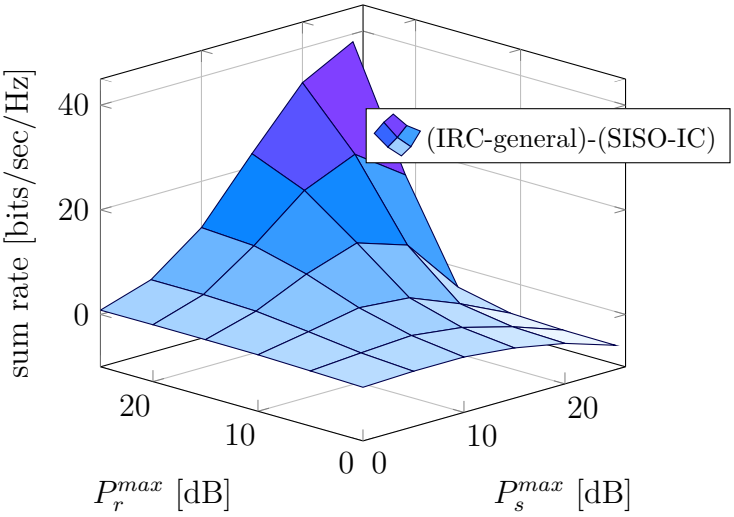


(b) Improvement of IRC-IN over SISO-IC

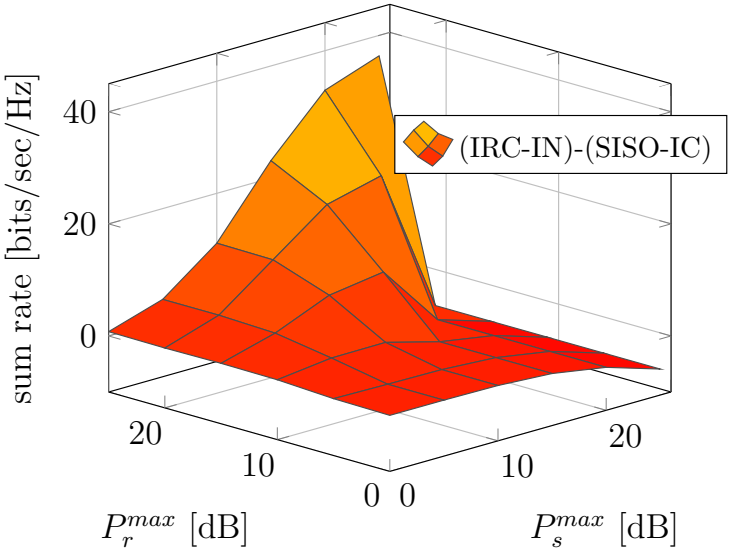
Figure 5.5: The sum rate improvement of (a) IRC-general over SISO-IC and (b) IRC-IN over SISO-IC with a particular channel realization of a two-user SISO IRC with $K = M = 2$.

5.1.4 Conclusion and future research directions

The achievable rate region of a SISO-IC has been an on-going research topic, with recent interest on the question whether a relay introduction to the SISO-IC, obtain-



(a) Improvement of IRC-general over SISO-IC



(b) Improvement of IRC-IN over SISO-IC

Figure 5.6: The proportional fairness improvement of (a) IRC-general over SISO-IC and (b) IRC-IN over SISO-IC of a particular channel realization of a two-user SISO IRC with $K = M = 2$.

ing an interference relay channel, provides any performance gain. In this paper, we study this problem by assuming an instantaneous amplify-and-forward relay with uninformed source and destination nodes in the system. We examine the gain of rate region of the relay introduction by formulating the Pareto boundary problem with optimization over relay processing matrix. The optimization problems, with and without the employment of interference neutralization techniques, are solved

using semi-definite relaxation techniques. The global optimality of the solutions are proved in the scenario of two source and two destination nodes. In the scenario of informed source nodes, we allow the source nodes to optimize their transmit power. The transmit power values at the source nodes which attain the Pareto boundary are obtained in closed-form. Simulation results confirm that instantaneous relay is able to improve the achievable rate region, even in the scenario of uninformed source nodes; improve average sum rate and average proportional fairness of the system.

This paper motivates the study of performance of the IRC with an AF relay which can be implemented easily in practical applications as the relay is only responsible for a simple forward process which does not incur a processing delay as compared to other complicated relays. As a preliminary study, we only allow the relay to choose between IN or no IN. To evaluate the full potential of the AF relay, one may allow the relay to choose different relay strategies, e.g. interference forwarding, signal amplifying, interference neutralization, etc., depending on its power budget and the channel qualities in the system. Another interesting problem is the physical placement of the relay with the goal of rate performance improvement.

5.2 Information Leakage Neutralization

In heterogeneous dense networks where spectrum is shared, users' privacy remains one of the major challenges. On a multi-antenna relay-assisted multi-carrier interference channel, each user shares the frequency and spatial resources with all the other users. When the receivers are not only interested in their own signals but also in eavesdropping other users' signals, the cross talk on the frequency and spatial channels becomes information leakage. In this paper, we propose a novel secrecy rate enhancing relay strategy that utilizes both frequency and spatial resources, termed as *information leakage neutralization*. To this end, the relay matrix is chosen such that the effective channel from the transmitter to the colluding eavesdropper is equal to the negative of the effective channel over the relay to the colluding eavesdropper and thus the information leakage to zero. Interestingly, the optimal relay matrix in general is not block-diagonal which encourages users' encoding over the frequency channels. We proposed two information leakage neutralization strategies, namely *efficient information leakage neutralization* (EFFIN) and *optimized information leakage neutralization* (OPTIN). EFFIN provides a simple and efficient design of relay processing matrix and precoding matrices at the transmitters in the scenario of limited power and computational resources. OPTIN, despite its higher complexity, provides a better sum secrecy rate performance by optimizing the relay processing matrix and the precoding matrices jointly. The proposed methods are shown to improve the sum secrecy rates over several state-of-the-art baseline methods.

In the following subsection, we give an example of a two-user interference relay channel in which the relay has two antennas and all nodes share two frequency

subcarriers. We shall illustrate that the conventional assumption of block diagonal relay matrix (which maximizes achievable rates in peaceful systems) cannot be adopted a-priori when secrecy rates are considered.

5.2.1 An example of two users on two frequencies with two antennas at the relay

Transmitter i, $i = 1, 2$, transmits symbols $\mathbf{x}_i \in \mathbb{C}^{M\times 1}$ which are spread over M frequency subcarriers by precoding matrix $\boldsymbol{P}_i$. For the ease of notation, we assume that precoding matrix $\boldsymbol{P}_i$ is a square matrix $\boldsymbol{P}_i \in \mathbb{C}^M$. When user i transmits $S_i \leq M$ symbols, then zeros are padded in $\mathbf{x}_i$ so that its dimension is always $M \times 1$ and correspondingly zero columns are padded in $\boldsymbol{P}_i$. We assume that the users do not overload the system and therefore S_i is smaller than or equal to the number of frequency subcarriers, here two. Note that $\boldsymbol{P}_i$ may have low row rank when certain subcarriers are not used. For example, if user i transmits one symbol on subcarrier 1 but nothing on subcarrier 2, then $\boldsymbol{P}_i = [a, 0; 0, 0]$ for some complex scalar a. If $\boldsymbol{P}_i$ is diagonal, then each symbol is only sent on one frequency. Denote the m-th transmit symbol of user i as $\mathbf{x}_i(m)$ which is randomly generated, mutually independent and with covariance matrix $\mathcal{I}_2$. The precoding matrix $\boldsymbol{P}_i$ satisfies the transmit power constraint of user i: $\mathrm{tr}\left(\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\right) \leq P_i^{max}$. Denote the channel gain from transmitter (TX) i to receiver (RX) j on frequency m as $h_{ji}(m)$. For simplicity of the example, we let S_i equal two. The received signal of user i is a vector whose m-th element is the received signal on the m-th frequency subcarrier,

$$\mathbf{y}_i = \begin{bmatrix}\mathbf{y}_i(1)\\ \mathbf{y}_i(2)\end{bmatrix} = \sum_{j=1}^{2} \begin{bmatrix} h_{ij}(1) & 0 \\ 0 & h_{ij}(2)\end{bmatrix} \boldsymbol{P}_j \begin{bmatrix} x_j(1)\\ x_j(2)\end{bmatrix} + \begin{bmatrix} n_i(1)\\ n_i(2)\end{bmatrix}. \tag{5.28}$$

The circular Gaussian noise with unit variance received on the m-th subcarrier at RX i is denoted as $n_i(m)$. If a relay with two antennas is introduced into the system, it receives the broadcasting signal from TXs and forwards them to RXs. We denote the received signal at the relay as a stacked vector of the received signal at each frequency m, with $\mathbf{y}_r(m) \in \mathbb{C}^{2\times 1}$ representing the received signal on frequency m and the a-th element in $\mathbf{y}_r(m)$ representing the signal at the a-th antenna:

$$\mathbf{y}_r = \begin{bmatrix}\mathbf{y}_r(1)\\ \mathbf{y}_r(2)\end{bmatrix} = \sum_{j=1}^{2} \begin{bmatrix} \boldsymbol{f}_j(1) & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{2\times 1} & \boldsymbol{f}_j(2)\end{bmatrix} \boldsymbol{P}_j \begin{bmatrix} x_j(1)\\ x_j(2)\end{bmatrix} + \begin{bmatrix} \mathbf{n}_r(1)\\ \mathbf{n}_r(2)\end{bmatrix} \tag{5.29}$$

where $\mathbf{n}_r(m) \in \mathbb{C}^{2\times 1}$ is a circular Gaussian noise vector received at frequency m with identity covariance matrix and $\boldsymbol{f}_j(m)$ is the complex vector channel from user j to the relay on frequency m. The relay processes the received signal $\mathbf{y}_r$ by a multiplication of matrix $\boldsymbol{R} \in \mathbb{C}^4$ and forwards the signal to the RXs. Denote the channel from relay to RX i on frequency m by $\boldsymbol{g}_i(m) \in \mathbb{C}^{2\times 1}$. At RX i, the received

signal is

$$\mathbf{y}_i = \sum_{j=1}^{2} \left(\begin{bmatrix} h_{ij}(1) & 0 \\ 0 & h_{ij}(2) \end{bmatrix} + \begin{bmatrix} \boldsymbol{g}_i^{\mathrm{H}}(1) & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \boldsymbol{g}_i^{\mathrm{H}}(2) \end{bmatrix} \boldsymbol{R} \begin{bmatrix} \boldsymbol{f}_j(1) & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{2\times 1} & \boldsymbol{f}_j(2) \end{bmatrix} \right) \boldsymbol{P}_j \begin{bmatrix} x_j(1) \\ x_j(2) \end{bmatrix} + \begin{bmatrix} \boldsymbol{g}_i^{\mathrm{H}}(1) & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \boldsymbol{g}_i^{\mathrm{H}}(2) \end{bmatrix} \boldsymbol{R} \begin{bmatrix} \mathbf{n}_r(1) \\ \mathbf{n}_r(2) \end{bmatrix} + \begin{bmatrix} n_i(1) \\ n_i(2) \end{bmatrix}. \tag{5.30}$$

Denote channel matrices

$$\boldsymbol{H}_{ij} = \begin{bmatrix} h_{ij}(1) & 0 \\ 0 & h_{ij}(2) \end{bmatrix}, \quad \boldsymbol{G}_i^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{g}_i^{\mathrm{H}}(1) & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \boldsymbol{g}_i^{\mathrm{H}}(2) \end{bmatrix}, \quad \boldsymbol{F}_i = \begin{bmatrix} \boldsymbol{f}_j(1) & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{2\times 1} & \boldsymbol{f}_j(2) \end{bmatrix}$$

and the equivalent channel from TX j to RX i as $\bar{\boldsymbol{H}}_{ij} = \boldsymbol{H}_{ij} + \boldsymbol{G}_i^{\mathrm{H}} \boldsymbol{R} \boldsymbol{F}_j$. An achievable rate of user 1 is

$$r_1(\boldsymbol{R}) = \mathcal{C}\left(\boldsymbol{I}_2 + \bar{\boldsymbol{H}}_{11} \boldsymbol{P}_1 \boldsymbol{P}_1^{\mathrm{H}} \bar{\boldsymbol{H}}_{11}^{\mathrm{H}} \left(\bar{\boldsymbol{H}}_{12} \boldsymbol{P}_2 \boldsymbol{P}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_{12}^{\mathrm{H}} + \boldsymbol{G}_1^{\mathrm{H}} \boldsymbol{R} \boldsymbol{R}^{\mathrm{H}} \boldsymbol{G}_1 + \boldsymbol{I}_2 \right)^{-1} \right). \tag{5.31}$$

Consider that RX 2 is an eavesdropper. We compute the worst-case scenario in which RX 2 decodes all other symbols perfectly before decoding the messages from TX 1 and RX 2 sees a MIMO channel and decodes messages $x_1(1)$ and $x_2(2)$ utilizing both frequencies (with a MMSE receive filter for example).

$$\mathbf{y}_{2\leftarrow 1} = \left(\begin{bmatrix} h_{21}(1) & 0 \\ 0 & h_{21}(2) \end{bmatrix} + \begin{bmatrix} \boldsymbol{g}_2^{\mathrm{H}}(1) & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \boldsymbol{g}_2^{\mathrm{H}}(2) \end{bmatrix} \boldsymbol{R} \begin{bmatrix} \boldsymbol{f}_1(1) & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{2\times 1} & \boldsymbol{f}_1(2) \end{bmatrix} \right) \boldsymbol{P}_1 \begin{bmatrix} x_1(1) \\ x_1(2) \end{bmatrix} + \begin{bmatrix} \boldsymbol{g}_2^{\mathrm{H}}(1) & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \boldsymbol{g}_2^{\mathrm{H}}(2) \end{bmatrix} \boldsymbol{R} \begin{bmatrix} \mathbf{n}_r(1) \\ \mathbf{n}_r(2) \end{bmatrix} + \begin{bmatrix} n_2(1) \\ n_2(2) \end{bmatrix} \tag{5.32}$$

An achievable rate is then $r_{2\leftarrow 1}(\boldsymbol{R}) = \mathcal{C}\left(\boldsymbol{I}_2 + \bar{\boldsymbol{H}}_{21} \boldsymbol{P}_1 \boldsymbol{P}_1^{\mathrm{H}} \bar{\boldsymbol{H}}_{21}^{\mathrm{H}} \left(\boldsymbol{G}_2^{\mathrm{H}} \boldsymbol{R} \boldsymbol{R}^{\mathrm{H}} \boldsymbol{G}_2 + \boldsymbol{I}_2 \right)^{-1} \right)$. An achievable secrecy rate of user 1 is then the achievable rate of user 1 $r_1(\boldsymbol{R})$ minus the leakage rate to user 2 $r_{2\leftarrow 1}(\boldsymbol{R})$:

$$\begin{aligned} r_1^s(\boldsymbol{R}) &= \left(r_1(\boldsymbol{R}) - r_{2\leftarrow 1}(\boldsymbol{R}) \right)^+ \\ &= \Bigg(\mathcal{C}\left(\boldsymbol{I}_2 + \bar{\boldsymbol{H}}_{11} \boldsymbol{P}_1 \boldsymbol{P}_1^{\mathrm{H}} \bar{\boldsymbol{H}}_{11}^{\mathrm{H}} \left(\bar{\boldsymbol{H}}_{12} \boldsymbol{P}_2 \boldsymbol{P}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_{12}^{\mathrm{H}} + \boldsymbol{G}_1^{\mathrm{H}} \boldsymbol{R} \boldsymbol{R}^{\mathrm{H}} \boldsymbol{G}_1 + \boldsymbol{I}_2 \right)^{-1} \right) \\ &\quad - \mathcal{C}\left(\boldsymbol{I}_2 + \bar{\boldsymbol{H}}_{21} \boldsymbol{P}_1 \boldsymbol{P}_1^{\mathrm{H}} \bar{\boldsymbol{H}}_{21}^{\mathrm{H}} \left(\boldsymbol{G}_2^{\mathrm{H}} \boldsymbol{R} \boldsymbol{R}^{\mathrm{H}} \boldsymbol{G}_2 + \boldsymbol{I}_2 \right)^{-1} \right) \Bigg)^+ . \end{aligned} \tag{5.33}$$

The relay processing matrix is defined as

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{bmatrix} \tag{5.34}$$

Table 5.1: Randomly generated channel realizations for a two user two frequency interference relay channel with two antennas at relay and single antenna at TXs and RXs.

$$\boldsymbol{H}_{11} = \begin{bmatrix} 0.5129 + 0.4605i & 0 \\ 0 & 0.3504 + 0.0950i \end{bmatrix}, \quad \boldsymbol{H}_{21} = \begin{bmatrix} 0.4337 + 0.0709i & 0 \\ 0 & 0.1160 + 0.0078i \end{bmatrix}$$

$$\boldsymbol{H}_{12} = \begin{bmatrix} 0.3693 + 0.0336i & 0 \\ 0 & 0.1922 + 0.4714i \end{bmatrix}, \quad \boldsymbol{H}_{22} = \begin{bmatrix} 0.1449 + 0.0718i & 0 \\ 0 & 0.6617 + 0.0432i \end{bmatrix}$$

$$\boldsymbol{G}_1 = \begin{bmatrix} 0.4460 + 0.5281i & 0 \\ 0.5083 + 0.5729i & 0 \\ 0 & 0.3608 + 0.1733i \\ 0 & 0.3365 + 0.0861i \end{bmatrix}, \quad \boldsymbol{G}_2 = \begin{bmatrix} 0.3933 + 0.0111i & 0 \\ 0.8044 + 0.2331i & 0 \\ 0 & 0.9339 + 0.7859i \\ 0 & 0.2268 + 0.4107i \end{bmatrix}$$

$$\boldsymbol{F}_1 = \begin{bmatrix} 0.1194 + 0.8624i & 0 \\ 0.6344 + 0.1582i & 0 \\ 0 & 0.6012 + 0.6261i \\ 0 & 0.1176 + 0.8351i \end{bmatrix}, \quad \boldsymbol{F}_2 = \begin{bmatrix} 0.9404 + 0.2720i & 0 \\ 0.4156 + 0.9280i & 0 \\ 0 & 0.9213 + 0.8129i \\ 0 & 0.5420 + 0.1664i \end{bmatrix}$$

$$\mathbb{R}^{\mathrm{IN}} = \begin{bmatrix} -0.0364 - 0.0035i & -0.1793 - 0.0233i & 0.0234 - 0.0575i & 0.0574 + 0.0596i \\ -0.1046 + 0.0925i & -0.2837 - 0.0390i & -0.0832 - 0.0249i & 0.0029 + 0.1567i \\ 0.2729 + 0.0708i & -0.1376 + 0.1714i & -0.3130 - 0.2977i & 0.2012 - 0.1606i \\ 0.0529 + 0.0099i & -0.1388 + 0.0348i & -0.4690 - 0.3154i & -0.0414 - 0.1751i \end{bmatrix}$$

$$\mathbb{R}^{\mathrm{IN},d} = \begin{bmatrix} -0.0364 - 0.0035i & -0.1793 - 0.0233i & 0 & 0 \\ -0.1046 + 0.0925i & -0.2837 - 0.0390i & 0 & 0 \\ 0 & 0 & -0.3130 - 0.2977i & 0.2012 - 0.1606i \\ 0 & 0 & -0.4690 - 0.3154i & -0.0414 - 0.1751i \end{bmatrix}$$

$$\mathbb{R}^{\mathrm{IN},z} = \begin{bmatrix} -0.2709 + 0.2267i & -0.0820 + 0.1738i & -0.0770 + 0.0704i & -0.1357 + 0.1183i \\ -0.1509 + 0.0212i & -0.3225 - 0.4885i & -0.2088 - 0.0485i & 0.6810 + 0.1046i \\ 0.2459 + 0.1223i & -0.1315 + 0.0682i & -0.2702 - 0.2781i & 0.2683 - 0.2842i \\ -0.0155 + 0.1640i & -0.2285 - 0.0472i & -0.5114 - 0.2436i & -0.0346 - 0.1960i \end{bmatrix}$$

where each submatrix block $\boldsymbol{R}_{mn}$ forwards signals from frequency n to frequency m. In a peaceful MIMO IRC, $\boldsymbol{R}$ bares a block diagonal structure, $\boldsymbol{R}_{12} = \boldsymbol{R}_{21} = \boldsymbol{0}_2$. The intuition is that relays should not generate cross talk over frequency channels. However, it is not trivial to examine the effect of $\boldsymbol{R}_{12}$ and $\boldsymbol{R}_{21}$ on secrecy rates as illustrated below and the conventional block diagonal structure should not be a-priori assumed.

As a numerical example, we compute the secrecy rates with the following randomly generated channels given in Table 5.1. We set the precoding matrices of TX 1 and

TX 2 to be

$$\boldsymbol{P}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{P}_2 = \begin{bmatrix} 1 & 4 \\ -4 & 1 \end{bmatrix}$$

which means that TX 1 transmits only one data stream on both subcarriers and TX 2 transmits two data streams spread over both frequency subcarriers with orthogonal sequences. With relay matrix $\boldsymbol{R}^{\mathrm{IN}}$ (see Table 5.1) a sum secrecy rate of 3.4104 is achievable whereas with block diagonal matrix $\boldsymbol{R}^{\mathrm{IN},d}$ the sum secrecy rate is 3.1881. A block diagonal relay matrix does not always improve secrecy rate and therefore in the following we assume a general non-block-diagonal structure $\boldsymbol{R}$. In fact, the relay matrix $\boldsymbol{R}^{\mathrm{IN}}$ is chosen such that the secrecy leakage is zero: $(\boldsymbol{H}_{12} + \mathcal{G}_1^{\mathrm{H}} \boldsymbol{R} \boldsymbol{F}_2)\boldsymbol{P}_2 = \boldsymbol{0}$ and $(\boldsymbol{H}_{21} + \mathcal{G}_2^{\mathrm{H}} \boldsymbol{R} \boldsymbol{F}_1)\boldsymbol{P}_1 = \boldsymbol{0}$. Thus, the secrecy rate from (5.33) can be simplified to the following

$$r_i^s = \mathcal{C}\left(\boldsymbol{I}_2 + \bar{\boldsymbol{H}}_{11}\boldsymbol{P}_1\boldsymbol{P}_1^{\mathrm{H}}\bar{\boldsymbol{H}}_{11}^{\mathrm{H}}\left(\boldsymbol{G}_1^{\mathrm{H}}\boldsymbol{R}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{G}_1 + \boldsymbol{I}_2\right)^{-1}\right). \tag{5.35}$$

This motivates our following proposition on information leakage neutralization techniques. Interestingly, with information leakage neutralization, we can simplify the optimization problem significantly. The idea is to set the information leakage from each user at each frequency to zero, in particular, by setting the equivalent channel of $\mathbf{x}_1$ from TX 1 to RX 2 and vice verse in (5.32) to zero,

$$\begin{cases} \left(\boldsymbol{H}_{12} + \boldsymbol{G}_1^{\mathrm{H}}\boldsymbol{R}\boldsymbol{F}_2\right)\boldsymbol{P}_2 = \boldsymbol{0} \\ \left(\boldsymbol{H}_{21} + \boldsymbol{G}_2^{\mathrm{H}}\boldsymbol{R}\boldsymbol{F}_1\right)\boldsymbol{P}_1 = \boldsymbol{0}\,. \end{cases} \tag{5.36}$$

With the properties of the Kronecker product, (5.36) can be written as

$$\begin{bmatrix} \left((\boldsymbol{F}_2\boldsymbol{P}_2)^{\mathrm{T}} \otimes \boldsymbol{G}_1^{\mathrm{H}}\right) \\ \left((\boldsymbol{F}_1\boldsymbol{P}_1)^{\mathrm{T}} \otimes \boldsymbol{G}_2^{\mathrm{H}}\right) \end{bmatrix} \mathrm{vec}(\boldsymbol{R}) = \mathbf{B}\,\mathrm{vec}(\boldsymbol{R}) = \begin{bmatrix} -\mathrm{vec}(\boldsymbol{H}_{12}\boldsymbol{P}_2) \\ -\mathrm{vec}(\boldsymbol{H}_{21}\boldsymbol{P}_1) \end{bmatrix} = \boldsymbol{b} \tag{5.37}$$

The stacked matrix $\mathbf{B}$ in the above equation is a fat matrix[3]. We obtain the relay matrix that can perform information leakage neutralization:

$$\mathrm{vec}(\boldsymbol{R}) = \mathbf{B}^{\mathrm{H}}\left(\mathbf{B}\,\mathbf{B}^{\mathrm{H}}\right)^{-1}\boldsymbol{b}. \tag{5.38}$$

Substitute the channel realizations in Table 5.1 into the above equation and reverse the vectorization operation, we obtain the relay matrix $\boldsymbol{R}^{\mathrm{IN}}$ (please refer to the table for numerical values).

Remark 5.2. If the precoding matrices $\{\boldsymbol{P}_i\}$ are invertible, then the relay matrix $\boldsymbol{R}$ obtained using (5.38) is block diagonal. A block diagonal relay matrix means that the relay sets cross talk over frequency subcarriers to zero and due to the interference

[3]Care must be taken when users send less than M data streams (when $\boldsymbol{P}_i$ has zero columns. More discussion is provided later in Proposition 2).

leakage neutralization, the interference from users on the same frequency is also zero. This results in KM parallel channels without interference. We propose in Section 5.2.3 a suboptimal but very efficient algorithm which optimizes the achievable rates in this case[4].

In fact, the matrix in (5.38) is not unique, any matrix which is a sum of $\mathrm{vec}(\boldsymbol{R})$ in (5.38) and a vector in the null space of $\mathbf{B}$ can also neutralize information leakage,

$$\mathrm{vec}(\boldsymbol{R}) = \mathbf{B}^{\mathrm{H}} \left(\mathbf{B}\,\mathbf{B}^{\mathrm{H}}\right)^{-1} \boldsymbol{b} + \boldsymbol{z}, \tag{5.39}$$

where $\boldsymbol{z} \in \mathcal{N}(\mathbf{B})$. With the channel realizations given in Table 5.1, we can generate another matrix $\boldsymbol{R}^{\mathrm{IN},\boldsymbol{z}}$ which achieves a higher secrecy rate 4.1553, a 17.8% increase of secrecy rate by optimization over $\boldsymbol{z}$. This motivates us to investigate an efficient method to find $\boldsymbol{z}$ and consequently $\boldsymbol{R}$ which neutralizes information leakage and optimizes the secrecy rate at the same time.

Remark 5.3. With the optimization over $\boldsymbol{z}$, the relay matrix is no longer block diagonal which couples the frequency channels. Although the problem is more complicated, we have shown in the above example that one can get a better secrecy rate performance. In Section 5.2.3, we propose an iterative sum secrecy rates optimization over the relay matrix $\boldsymbol{R}$ and the precoding matrices $\{\boldsymbol{P}_i\}$.

In the following section, we illustrate how the relay matrix can be chosen carefully to amplify the desired signal strength and at the same time neutralize information leakage in the multi-user scenario.

5.2.2 General multi-user multi-antenna multi-carrier scenario

In this section, we let the number of TXs and RXs be $K \geq 2$. The TXs and RXs have single antenna and the relay has N antennas. Let the number of frequency subcarriers be M. Denote the complex channel from TX i to RX j, as a diagonal matrix $\boldsymbol{H}_{ji} \in \mathbb{C}^{M}$ and the complex channel from TX i to relay as $\boldsymbol{F}_i \in \mathbb{C}^{NM \times M}$ and from relay to RX j as $\boldsymbol{G}_j \in \mathbb{C}^{MN \times M}$. The signal received at the relay is,

$$\mathbf{y}_r = \sum_{i=1}^{K} \boldsymbol{F}_i \boldsymbol{P}_i \mathbf{x}_i + \mathbf{n}_r \tag{5.40}$$

where $\boldsymbol{F}_i = \mathrm{diag}\left(\boldsymbol{f}_i(1), \ldots, \boldsymbol{f}_i(M)\right)$ and $\mathbf{x}_i \in \mathbb{C}^{M \times 1}$ are the circular Gaussian transmit symbols from TX i, with zero mean and identity covariance matrix. The matrix $\boldsymbol{P}_i \in \mathbb{C}^{M}$ satisfies the power constraint:

$$\mathrm{tr}\left(\boldsymbol{P}_i \boldsymbol{P}_i^{\mathrm{H}}\right) \leq P_i^{max}. \tag{5.41}$$

[4]The achievable rates here are secrecy rates as the information leakage is zero.

With AF strategy, the relay multiplies the received signal $\mathbf{y}_r$ on the left by processing matrix $\boldsymbol{R}$ and transmits $\boldsymbol{R}\,\mathbf{y}_r$. The transmit power of the relay is constrained by P_r^{max},

$$\mathrm{tr}\left(\boldsymbol{R}\left(\sum_{i=1}^{K}\boldsymbol{F}_i\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{F}_i^{\mathrm{H}}+\boldsymbol{I}_{\mathrm{MN}}\right)\boldsymbol{R}^{H}\right)\leq P_r^{max}. \tag{5.42}$$

The received signal at RX j is

$$\mathbf{y}_j=\sum_{i=1}^{K}\left(\boldsymbol{H}_{ji}+\boldsymbol{G}_j^{\mathrm{H}}\boldsymbol{R}\boldsymbol{F}_i\right)\boldsymbol{P}_i\mathbf{x}_i+\boldsymbol{G}_j^{\mathrm{H}}\boldsymbol{R}\mathbf{n}_r+\boldsymbol{n}_j \tag{5.43}$$

where $\mathbf{n}_j$ is the circular Gaussian noise at RX j with zero mean and identity covariance matrix and $\boldsymbol{G}_j=\mathrm{diag}(\boldsymbol{g}_j(1),\ldots,\boldsymbol{g}_j(M))$. For the ease of notation, we define the equivalent channel from i to j as

$$\bar{\boldsymbol{H}}_{ji}=\boldsymbol{H}_{ji}+\boldsymbol{G}_j^{H}\boldsymbol{R}\boldsymbol{F}_i \tag{5.44}$$

and its (f,m)-element is $[\bar{\boldsymbol{H}}_{ji}]_{fm}=h_{ji}+\boldsymbol{g}_j^{\mathrm{H}}(f)\boldsymbol{R}_{fm}\boldsymbol{f}_i(m)$ which is the equivalent channel from user i frequency m to user j frequency f.

Each RX is not only interested in decoding its own signal but also eavesdropping from other TXs. In the following, we define the worst case achievable secrecy rate with colluding eavesdroppers. For messages $\mathbf{x}_i$, all RXs except RX i collaborate to form an eavesdropper with multiple antennas and the message $\mathbf{x}_i$ goes through a multi-carrier MIMO channel to the colluding eavesdroppers. A worst case secrecy rate is then to assume that all other messages $\mathbf{x}_j, j\neq i$ are decoded perfectly and subtracted before decoding $\mathbf{x}_i$. The received signals at RX i and the colluding eavesdroppers are

$$\begin{cases}\mathbf{y}_i=\displaystyle\sum_{k=1}^{K}\bar{\boldsymbol{H}}_{ik}\boldsymbol{P}_k\mathbf{x}_k+\boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\mathbf{n}_r+\mathbf{n}_i\\ \mathbf{y}_{-i}=\begin{bmatrix}\bar{\boldsymbol{H}}_{1i}\\ \vdots\\ \bar{\boldsymbol{H}}_{(i-1)i}\\ \bar{\boldsymbol{H}}_{(i+1)i}\\ \vdots\\ \bar{\boldsymbol{H}}_{Ki}\end{bmatrix}\boldsymbol{P}_i\mathbf{x}_i+\begin{bmatrix}\boldsymbol{G}_1^{\mathrm{H}}\\ \vdots\\ \boldsymbol{G}_{i-1}^{\mathrm{H}}\\ \boldsymbol{G}_{i+1}^{\mathrm{H}}\\ \vdots\\ \boldsymbol{G}_k^{\mathrm{H}}\end{bmatrix}\boldsymbol{R}\,\mathbf{n}_r+\begin{bmatrix}\mathbf{n}_1\\ \vdots\\ \mathbf{n}_{i-1}\\ \mathbf{n}_{i+1}\\ \vdots\\ \mathbf{n}_K\end{bmatrix}\\ \quad=\bar{\boldsymbol{H}}_{-i}\boldsymbol{P}_i\mathbf{x}_i+\boldsymbol{G}_{-i}^{\mathrm{H}}\boldsymbol{R}\,\mathbf{n}_r+\mathbf{n}_{-i}\,.\end{cases} \tag{5.45}$$

The secrecy rate of user i is,

$$r_i^s = \left(\mathcal{C}\left(\boldsymbol{I}_M + \bar{\boldsymbol{H}}_{ii}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\bar{\boldsymbol{H}}_{ii}^{\mathrm{H}} \left(\sum_{j\neq i} \bar{\boldsymbol{H}}_{ij}\boldsymbol{P}_j\boldsymbol{P}_j^{\mathrm{H}}\bar{\boldsymbol{H}}_{ij}^{\mathrm{H}} + \boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{G}_i + \boldsymbol{I}_M \right)^{-1} \right) \right. \\ \left. - \mathcal{C}\left(\boldsymbol{I}_{M(K-1)} + \bar{\boldsymbol{H}}_{-i}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\bar{\boldsymbol{H}}_{-i}^{\mathrm{H}} \left(\boldsymbol{G}_{-i}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{G}_{-i} + \boldsymbol{I}_{M(K-1)} \right)^{-1} \right) \right)^{+}. \tag{5.46}$$

Recall from (5.44) that the equivalent channel from Tx j to Rx i $\bar{\boldsymbol{H}}_{ij}$ is a function of the relay processing matrix $\boldsymbol{R}$, $\bar{\boldsymbol{H}}_{ij} = \boldsymbol{H}_{ij} + \boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\boldsymbol{F}_j$. The optimization of the aforementioned secrecy rates is highly complicated due to their non-convex structure. In the following, we propose the information leakage neutralization technique [60] which is able to neutralize all information leakage to all eavesdroppers in the air by choosing the relay strategy in a careful manner. As illustrated in the previous section, with information leakage neutralization, the secrecy rate expression (5.46) can be simplified to

$$r_i^s = \mathcal{C}\left(\boldsymbol{I}_M + \bar{\boldsymbol{H}}_{ii}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\bar{\boldsymbol{H}}_{ii}^{\mathrm{H}} \left(\boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{G}_i + \mathcal{I}_M \right)^{-1} \right). \tag{5.47}$$

In the following section, we illustrate how we can choose $\boldsymbol{R}$ to achieve a secrecy rate as such.

5.2.3 Information leakage neutralization algorithms

In the previous section, we have shown that secrecy rates (5.47) are achievable by information leakage neutralization. Also, in order to implement information leakage neutralization, the number of antennas at the relay, the number of frequency subcarriers and the number of users in the system must satisfy the relation in Proposition 1. In Proposition 2, we computed the minimum relay power required in order to perform information leakage neutralization. With more power available at the relay, we can improve the achievable secrecy rates by optimizing the relay matrix and the precoding matrices. The optimization of sum secrecy rates can be written formally in the following:

$$\begin{aligned} \max_{\boldsymbol{R},\{\boldsymbol{P}_i\}} \quad & \sum_{i=1}^{K} \mathcal{C}\left(\boldsymbol{I}_M + \bar{\boldsymbol{H}}_{ii}\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\bar{\boldsymbol{H}}_{ii}^{\mathrm{H}} \left(\boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{G}_i + \boldsymbol{I}_M \right)^{-1} \right) \\ \text{s.t.} \quad & \mathrm{tr}\left(\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}} \right) \leq P_i^{max} \\ & \mathrm{tr}\left(\boldsymbol{R}\left(\sum_{i=1}^{K} \boldsymbol{F}_i\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{F}_i^{\mathrm{H}} \right)\boldsymbol{R}^{\mathrm{H}} \right) \leq P_r^{max}. \end{aligned} \tag{5.48}$$

In the following, we propose two algorithms. The first algorithm EFFIN, in Section 5.2.3, considers the scenario where all users transmit the maximum number of data

streams allowed $S_i = M$. We observe that in this situation, information leakage neutralization decomposes the system into KM parallel channels and consequently both the relay processing matrix $\boldsymbol{R}$ and the precoding matrix $\boldsymbol{P}_i$ can be computed very efficiently. The second algorithm OPTIN, in Section 5.2.3, investigates a systematic method for the computation of $\boldsymbol{R}$ and $\boldsymbol{P}_i$ when there is enough transmit power budget at the relay to allow further optimization of secrecy rates.

Efficient Information Leakage Neutralization (EFFIN)

When every user transmits $S_i = M$ data streams and $\boldsymbol{P}_i$ is invertible, we propose the following algorithm that decomposes the K user interference relay channels with M frequency subcarriers and N antennas at the relay to KM parallel secure channels *with no interference and no information leakage*. The information leakage neutralization criteria $\left(\boldsymbol{H}_{ij} + \boldsymbol{G}_i^{\mathrm{H}} \boldsymbol{R} \boldsymbol{F}_j\right) \boldsymbol{P}_i = \boldsymbol{0}$, when $\boldsymbol{P}_i$ is invertible, is equivalent to

$$\boldsymbol{H}_{ij} + \boldsymbol{G}_i^{\mathrm{H}} \boldsymbol{R} \boldsymbol{F}_j = \boldsymbol{0}\,. \tag{5.49}$$

Due to the block diagonal structure of $\boldsymbol{H}_{ij}$, $\boldsymbol{G}_i$ and $\boldsymbol{F}_j$, one feasible solution of the above equation is a block diagonal $\boldsymbol{R}$. With the block diagonal structure, the resulting secrecy rates may be suboptimal, but the information leakage neutralization constraint can be broken down to the optimization over the diagonal blocks $\boldsymbol{R}_{mm}$ in $\boldsymbol{R}$:

$$h_{ji}(m) + \boldsymbol{g}_j^{\mathrm{H}}(m) \boldsymbol{R}_{mm} \boldsymbol{f}_i(m) = 0, \quad i, j = 1, \ldots, K, i \neq j. \tag{5.50}$$

Following the same approach as before, we stack the constraints for all $j \neq i$ and define

$$\begin{aligned} \boldsymbol{h}_{-i}(m) &= \left[h_{1i}^{\mathrm{H}}(m), \ldots, h_{(i-1)i}^{\mathrm{H}}(m), h_{(i+1)i}^{\mathrm{H}}(m), \ldots, h_{Ki}^{\mathrm{H}}(m)\right]^{\mathrm{H}} \\ \boldsymbol{G}_{-i}(m) &= \left[\boldsymbol{g}_1(m), \ldots, \boldsymbol{g}_{i-1}(m), \boldsymbol{g}_{i+1}(m), \ldots, \boldsymbol{g}_K(m)\right]. \end{aligned}$$

We obtain $\boldsymbol{h}_{-i}(m) + \boldsymbol{G}_{-i}^{\mathrm{H}}(m) \boldsymbol{R}_{mm} \boldsymbol{f}_i(m) = \boldsymbol{0}_{(K-1)\times 1}$ which is equivalent to

$$\left(\boldsymbol{f}_i^{\mathrm{T}}(m) \otimes \boldsymbol{G}_{-i}^{\mathrm{H}}(m)\right) \mathrm{vec}\left(\boldsymbol{R}_{mm}\right) = -\boldsymbol{h}_{-i}(m).$$

Stacking constraints for all i, we have

$$\boldsymbol{A}(m) = \begin{bmatrix} \left(\boldsymbol{f}_1^{\mathrm{T}}(m) \otimes \boldsymbol{G}_{-1}^{\mathrm{H}}(m)\right) \\ \vdots \\ \left(\boldsymbol{f}_K^{\mathrm{T}}(m) \otimes \boldsymbol{G}_{-K}^{\mathrm{H}}(m)\right) \end{bmatrix}, \quad \boldsymbol{b}(m) = \begin{bmatrix} -\boldsymbol{h}_{-1}(m) \\ \vdots \\ -\boldsymbol{h}_{-K}(m) \end{bmatrix}.$$

With a limited power budget at relay, we propose to implement information leakage neutralization with the least relay transmit power and utilize the result from Proposition 2, the relay matrix has the m-th diagonal block equal to

$$\boldsymbol{R}_{mm} = \mathrm{vec}^{-1}\left(\left(\boldsymbol{A}(m)\right)^{\dagger} \boldsymbol{b}(m)\right) \tag{5.51}$$

where $\text{vec}(.)^{-1}$ is to reverse the vectorization of a vector column-wise to a $M \times M$ matrix. After the computation of the relay matrix in (5.51), $\boldsymbol{R} = \text{diag}\left(\boldsymbol{R}_{11}, \ldots, \boldsymbol{R}_{MM}\right)$, the optimal precoding matrices $\{\boldsymbol{P}_i\}$ are computed by solving $\mathcal{Q}_1$.

$$\begin{aligned} \mathcal{Q}_1 : \max_{\{\boldsymbol{Q}_i\}, \boldsymbol{Q}_i \succeq 0} \quad & \sum_{i=1}^{K} \mathcal{C}\left(\boldsymbol{I}_M + \boldsymbol{Q}_i \boldsymbol{W}_i\right) \\ \text{s.t.} \quad & \text{tr}\left(\boldsymbol{Q}_i\right) \leq P b_i^{max}, \;\; i = 1, \ldots, K, \\ & \sum_{i=1}^{K} \text{tr}\left(\boldsymbol{Q}_i \boldsymbol{X}_i\right) \leq \bar{P}_r^{max}. \end{aligned} \tag{5.52}$$

where we replace $\boldsymbol{P}_i \boldsymbol{P}_i^{\text{H}}$ by positive semi-definite variable $\boldsymbol{Q}_i$ and denote the following matrices

$$\begin{aligned} \boldsymbol{W}_i &= \left(\boldsymbol{H}_{ii} + \boldsymbol{G}_i^{\text{H}} \boldsymbol{R} \boldsymbol{F}_i\right)^{\text{H}} \left(\boldsymbol{G}_i^{\text{H}} \boldsymbol{R} \boldsymbol{R}^{\text{H}} \boldsymbol{G}_i + \boldsymbol{I}_M\right)^{-1} \left(\boldsymbol{H}_{ii} + \boldsymbol{G}_i^{\text{H}} \boldsymbol{R} \boldsymbol{F}_i\right), \\ \boldsymbol{X}_i &= \boldsymbol{F}_i^{\text{H}} \boldsymbol{R}^{\text{H}} \boldsymbol{R} \boldsymbol{F}_i, \\ \bar{P}_r^{max} &= P_r^{max} - \text{tr}\left(\boldsymbol{R} \boldsymbol{R}^{\text{H}}\right). \end{aligned} \tag{5.53}$$

The objective in $\mathcal{Q}_1$ is concave in $\boldsymbol{Q}_i$ as $\boldsymbol{W}_i$ is positive semi-definite and the constraints are linear in $\boldsymbol{Q}_i$. Thus, $\mathcal{Q}_1$ is a semi-definite program and can be solved readily using convex optimization solvers, e.g. CVX[5]. The optimal $\boldsymbol{P}_i$ is obtained by performing eigenvalue decomposition on $\boldsymbol{Q}_i = \boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{U}_i^{\text{H}}$ and $\boldsymbol{P}_i = \boldsymbol{U}_i \boldsymbol{D}_i^{1/2}$. The psuedocode of the EFFIN is given in Algorithm 5.

Optimized Information Leakage Neutralization (OPTIN)

In the previous subsection, we have discussed a simple, efficient and power saving solution of the relay matrix and precoding matrices for secure transmission. One drawback of the efficient method is that its performance may be suboptimal. In this subsection, we discuss how to choose the relay and precoding matrices such that the sum secrecy rates are optimized while ensuring zero information leakage.

To this end, we rewrite the information leakage neutralization constraint (5.49) to promote the optimization of secrecy rates,

$$\left(\boldsymbol{H} + \boldsymbol{G}^{\text{H}} \boldsymbol{R} \boldsymbol{F}\right) \boldsymbol{P} = \boldsymbol{T} \tag{5.54}$$

where $\boldsymbol{H} = [\boldsymbol{H}_{11}, \ldots, \boldsymbol{H}_{1K}; \ldots; \boldsymbol{H}_{K1}, \ldots, \boldsymbol{H}_{KK}]$, $\boldsymbol{G}^{\text{H}} = \left[\boldsymbol{G}_1^{\text{H}}; \ldots; \boldsymbol{G}_K^{\text{H}}\right]$, $\boldsymbol{F} = [\boldsymbol{F}_1, \ldots, \boldsymbol{F}_K]$ and $\boldsymbol{P} = \text{diag}(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_K)$. The block diagonal matrix $\boldsymbol{T} =$

[5]Given block diagonal $\boldsymbol{R}$ in (5.51), the equivalent channel $\boldsymbol{W}_i$ and matrix $\boldsymbol{X}_i$ are also block diagonal. It is possible to solve $\mathcal{Q}_1$ using water-filling with $K+1$ Lagrange multipliers. For large problem size, it may be more computational efficient using a tailor made water-filling method. For medium size problems and illustrative purposes, we propose here to solve by semi-definite programming.

Algorithm 5 The pseudo-code for Efficient Information Leakage Neutralization (EFFIN)

1: **for** $m = 1 \to M$ **do**
2: Compute $\boldsymbol{R}_{mm} = \text{vec}^{-1}\left((\boldsymbol{A}(m))^{\dagger}\,\boldsymbol{b}(m)\right)$ with

$$\boldsymbol{A}(m) = \begin{bmatrix} \left(\boldsymbol{f}_1^{\mathrm{T}}(m) \otimes \boldsymbol{G}_{-1}^{\mathrm{H}}(m)\right) \\ \vdots \\ \left(\boldsymbol{f}_K^{\mathrm{T}}(m) \otimes \boldsymbol{G}_{-K}^{\mathrm{H}}(m)\right) \end{bmatrix}, \quad \boldsymbol{b}(m) = \begin{bmatrix} -\boldsymbol{h}_{-1}(m) \\ \vdots \\ -\boldsymbol{h}_{-K}(m) \end{bmatrix}.$$

3: **end for**
4: The relay processing matrix is $\boldsymbol{R} = \text{diag}\,(\boldsymbol{R}_{11}, \ldots, \boldsymbol{R}_{MM})$.
5: Solve $\mathcal{Q}_1$ using convex optimization solvers and obtain optimal $\{\boldsymbol{Q}_i\}$.
6: **for** $i = 1 \to K$ **do**
7: Perform eigen-value decomposition, $\boldsymbol{Q}_i = \boldsymbol{U}_i\boldsymbol{D}_i\boldsymbol{U}_i^{\mathrm{H}}$. Set $\boldsymbol{P}_i = \boldsymbol{U}_i\boldsymbol{D}_i^{1/2}$.
8: **end for**

$\text{diag}(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_K)$ is the new optimization variable. $\boldsymbol{T}_i$ is the equivalent desired channel from TX i to RX i as $\boldsymbol{T}_i = (\boldsymbol{H}_{ii} + \boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\boldsymbol{F}_i)\boldsymbol{P}_i$. By applying pseudo-inverses[6] of $\boldsymbol{G}^{\mathrm{H}}$ and $\boldsymbol{F}\boldsymbol{P}$ ($\boldsymbol{G}^{\mathrm{H}\,\dagger}$ and $(\boldsymbol{F}\boldsymbol{P})^{\dagger}$ respectively), one can rewrite (5.54) to the following

$$\boldsymbol{R} = \boldsymbol{G}^{\mathrm{H}\,\dagger}\left(\boldsymbol{T} - \boldsymbol{H}\boldsymbol{P}\right)\left(\boldsymbol{F}\boldsymbol{P}\right)^{\dagger}. \tag{5.55}$$

The maximum achievable sum secrecy rate is the solution of the following problem

$$\max_{\boldsymbol{R},\boldsymbol{T},\{\boldsymbol{P}_i\}} \quad \sum_{i=1}^{K} \mathcal{C}\left(\mathcal{I}_M + \boldsymbol{T}_i\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\boldsymbol{T}_i^{\mathrm{H}}\left(\boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{G}_i + \boldsymbol{I}_M\right)^{-1}\right) \tag{5.56a}$$

$$\text{s.t.} \quad \text{tr}\left(\boldsymbol{P}_i\boldsymbol{P}_i^{\mathrm{H}}\right) \le P_i^{max}, \;\; i = 1, \ldots, K, \tag{5.56b}$$

$$\left(\boldsymbol{H} + \boldsymbol{G}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{F}\right)\boldsymbol{P} = \boldsymbol{T}, \tag{5.56c}$$

$$\text{tr}\left(\boldsymbol{R}\left(\boldsymbol{F}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}} + \boldsymbol{I}_{\mathrm{MN}}\right)\boldsymbol{R}^{\mathrm{H}}\right) \le P_r^{max} \tag{5.56d}$$

$$\boldsymbol{T} = \text{diag}\,(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_K). \tag{5.56e}$$

Note that in the objective function, the information leakage is neutralized for each user. Constraints (5.56b) and (5.56d) are the transmit power constraints at the TXs and at the relay respectively. The information leakage neutralization constraint is written as (5.56c). The optimization is not jointly convex in $\boldsymbol{R}$, $\boldsymbol{T}$ and $\{\boldsymbol{P}_i\}$. To simplify the optimization problem, we propose the following iterative optimization algorithm. Given $\boldsymbol{R}$ and $\boldsymbol{T}$, we solve $\boldsymbol{P}_i$ optimally using $\mathcal{Q}_1$ in EFFIN. The second part of the iterative algorithm is to compute the optimal relay strategy $\boldsymbol{R}$ and the

[6]Note that $\mathcal{G}^{\mathrm{H}}$ has dimension $MK \times MN$ and $\boldsymbol{F}\boldsymbol{P}$ has dimension $MN \times KM$. If $MN \ge MK$, then $\boldsymbol{G}^{\mathrm{H}\,\dagger} = \boldsymbol{G}\left(\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}\right)^{-1}$ and $(\boldsymbol{F}\boldsymbol{P})^{\dagger} = \left((\boldsymbol{F}\boldsymbol{P})^{\mathrm{H}}(\boldsymbol{F}\boldsymbol{P})\right)^{-1}(\boldsymbol{F}\boldsymbol{P})^{\mathrm{H}}$. If $MN < KM$, then $\boldsymbol{G}^{\mathrm{H}\,\dagger} = \left(\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\right)^{-1}\boldsymbol{G}$ and $(\boldsymbol{F}\boldsymbol{P})^{\dagger} = (\boldsymbol{F}\boldsymbol{P})^{\mathrm{H}}\left(\boldsymbol{F}\boldsymbol{P}(\boldsymbol{F}\boldsymbol{P})^{\mathrm{H}}\right)^{-1}$.

auxiliary variable $\boldsymbol{T}$ (by solving $\mathcal{Q}_2$) if the precoding matrices $\boldsymbol{P}_i$ as the solutions of $\mathcal{Q}_1$ are given.

$$\begin{aligned} \mathcal{Q}_2: \quad \max_{\boldsymbol{R},\boldsymbol{T}} \quad & \sum_{i=1}^{K} \mathcal{C}\left(\boldsymbol{I}_M + \boldsymbol{T}_i\boldsymbol{T}_i^{\mathrm{H}}\left(\boldsymbol{G}_i^{\mathrm{H}}\boldsymbol{R}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{G}_i + \boldsymbol{I}_M\right)^{-1}\right) \\ \text{s.t.} \quad & \boldsymbol{R} = \boldsymbol{G}^{\mathrm{H}\,\dagger}\left(\boldsymbol{T} - \boldsymbol{H}\boldsymbol{P}\right)\left(\boldsymbol{F}\boldsymbol{P}\right)^{\dagger}, \\ & \mathrm{tr}\left(\boldsymbol{R}\left(\boldsymbol{F}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}} + \boldsymbol{I}_{\mathrm{MN}}\right)\boldsymbol{R}^{\mathrm{H}}\right) \leq P_r^{max}, \\ & \boldsymbol{T} = \mathrm{diag}(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_K). \end{aligned} \tag{5.57}$$

Problem $\mathcal{Q}_2$ is non-convex. The major challenge is due to the sum of log-determinants in the objective function and the equality constraints. In the following, we utilize the first equality constraint and replace $\boldsymbol{R}$ as a function of $\boldsymbol{T}$. The optimization problem $\mathcal{Q}_2$ can be written as,

$$\begin{aligned} \mathcal{Q}_2': \quad \max_{\boldsymbol{T}} \quad & \sum_{i=1}^{K}\left(\mathcal{C}\left(\boldsymbol{X}_i + \bar{\boldsymbol{T}}_i\boldsymbol{Z}_i\bar{\boldsymbol{T}}_i^{H}\right) - \mathcal{C}\left(\boldsymbol{X}_i + \bar{\boldsymbol{T}}_i\boldsymbol{Y}_i\bar{\boldsymbol{T}}_i^{\mathrm{H}}\right)\right) \\ \text{s.t.} \quad & \mathrm{tr}\left(\boldsymbol{G}^{\mathrm{H}\,\dagger}\left(\boldsymbol{T} - \boldsymbol{H}\boldsymbol{P}\right)\left(\tilde{\boldsymbol{F}} + \boldsymbol{I}_{MK}\right)\left(\boldsymbol{T} - \boldsymbol{H}\boldsymbol{P}\right)^{\mathrm{H}}\boldsymbol{G}^{\dagger}\right) \leq P_r^{max}, \\ & \bar{\boldsymbol{T}}_i = \left[\boldsymbol{T}_i, \boldsymbol{I}_M\right], \\ & \boldsymbol{T} = \mathrm{diag}(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_K). \end{aligned} \tag{5.58}$$

Although the optimization problem is simplified, it is still non-convex in $\boldsymbol{T}$. We propose to solve $\mathcal{Q}_2'$ with gradient descent method. We summarize in Algorithm 6 the proposed iterative algorithm on sum secrecy rate optimization.

5.2.4 Simulation results

To illustrate the effectiveness of the proposed algorithms, we provide in this section numerical simulations for different system settings. As an example, we simulate the secrecy rates of a relay assisted network with $K = 2$ users, $M = 8$ frequency subcarriers and $N = 2$ antennas at the relay, unless otherwise stated. To examine the performance of the algorithms with respect to system signal-to-noise ratio, we vary the transmit power constraint at relay from 0 to 30 dB while keeping the transmit power constraint at TXs as 10 dB (see Figure 5.7.) Similarly, we examine the algorithms by varying the transmit power constraint at TXs from 0 to 30 dB while keeping the transmit power at relay constraint at $23, 27, 30$ dB. Note that by varying the power constraints, we do not force the power of the optimized precoding matrices and the relay processing matrix to be equal to the power constraints. In the following, we compare algorithms:

Algorithm 6 The pseudo-code for Optimized Information Leakage Neutralization (OPTIN)

while do
  Initialize $\{\boldsymbol{P}_i\}$ and $\boldsymbol{R}$ as the solutions of EFFIN.
  Solve $\mathcal{Q}_2'$ using gradient descent method with gradient of the Lagrangian of $\mathcal{Q}_2'$ and obtain optimal solution $\boldsymbol{T}$. Obtain relay processing matrix $\boldsymbol{R}$ from $\boldsymbol{T}$ using (5.55).
  With $\boldsymbol{R}$ and $\boldsymbol{T}$ above, solve $\mathcal{Q}_1$ using convex optimization solvers and obtain optimal $\{\boldsymbol{Q}_i\}$.
  for $i = 1 \rightarrow K$ **do**
    Perform eigen-value decomposition, $\boldsymbol{Q}_i = \boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{U}_i^{\mathrm{H}}$. Set $\boldsymbol{P}_i = \boldsymbol{U}_i \boldsymbol{D}_i^{1/2}$.
  end for
  if sum secrecy rate improvement is less than a predefined threshold **then**
    Convergence reached. Break.
  end if
end while

- Baseline 1 (Repeater): the relay is a layer 1 relay and is only able to forward signals without additional signal processing. This corresponds to setting $\boldsymbol{R} = \mathcal{I}_{MN}\sqrt{\frac{P_r^{max}}{MN}}$.
- Baseline 2 (IC): the relay shuts down, i.e. $\boldsymbol{R} = \boldsymbol{0}_{MN}$, and we obtain an interference channel where users eavesdrop each other.
- Proposed algorithm EFFIN: an efficient relay and precoding matrices optimization algorithm outlined in Algorithm 5.
- Proposed algorithm OPTIN: an optimized algorithm whose performance exceeds EFFIN with a price of higher complexity. OPTIN is outlined in Algorithm 6.

For each baseline algorithm, we examine the effect of spectrum sharing on achievable secrecy rates by employing either one of the following spectrum sharing methods:

- Full spectrum sharing (FS): users are allowed to use the entire spectrum. Each TX measures the channel qualities of the direct channel and the channel from itself to other RXs. Based on the measured channel qualities, each TX excludes frequency subcarriers with zero secrecy rates and transmits on the channels with non-zero secrecy rates. For subcarriers at which more than one user would like to transmit, we assume that the TXs coordinate so that the TX with a high secrecy rate would transmit on that subcarrier. Despite such coordination, each user eavesdrops other users on each subcarrier.
- Orthogonal spectrum sharing (OS): users are assigned exclusive portion of spectrum. Each TX excludes subcarriers with zero secrecy rates and transmits on the channels with non-zero secrecy rates. Each user eavesdrops other users

on each subcarrier.

Secrecy rates with increasing relay power

In Figure 5.7, we show achievable sum secrecy rates over varying the transmit power constraint at the relay from 0 to 30 dB while keeping the transmit power constraint at the TXs at 10 dB. As the IC does not utilize the relay, the achievable sum secrecy rates (plotted with triangles) are constant as the relay power constraint increases. As expected from intuition, the performance of IC with FS is better than OS because OS has an additional constraint of subcarrier assignment. The achievable sum secrecy rates achieved by a repeater decreases with relay transmit power. This is due to the increased amplification noise in AF relaying. Interestingly, the non-intelligent relaying scheme, e.g. a repeater, may decrease the secrecy rate significantly, even worse than switching off the relay. However, utilizing an intelligent relay and choosing the relaying scheme, one can improve the achievable secrecy rate significantly, about 550% over a simple repeater and about 200% over IC. Although EFFIN is very simple and efficient, it achieves 94.5% of the sum secrecy rate achieved by the more complicated algorithm OPTIN.

Secrecy rates with increasing TX power

In Figure 5.8, we simulate the achievable sum secrecy rate by the transmit power constraint at TXs from 0 to 30 dB while keeping the transmit power at relay constraint at $23, 27, 30$ dB. As the transmit power at TX increases, the sum secrecy rates saturate in both baseline algorithms, Repeater and IC. With the proposed information leakage neutralization, we see that the sum secrecy rates grow unbounded with the TX power as each user enjoys a leakage free frequency channel. Note that the sum secrecy rates achieved by relay with power constraint at $23, 27, 30$ dB are plotted in dotted, dashed and solid lines respectively. When there is only 23 dB available, there is only enough power for information leakage neutralization, but not enough to further optimize the system performance. Hence, the achievable sum secrecy rates of EFFIN and OPTIN overlap. With more power available, it is possible to optimize the sum secrecy rates while neutralizing information leakage and the performance of OPTIN is better than EFFIN.

Secrecy rates with larger systems

In Figure 5.9, we examine the performance of the proposed algorithms in a slightly larger systems with $N = 4$ antennas at the relay and $M = 16$ frequency subcarriers. The relay processing matrix is therefore a 64×64 matrix. The proposed scheme EFFIN and OPTIN outperform baseline algorithms Repeater and IC by 200% whereas

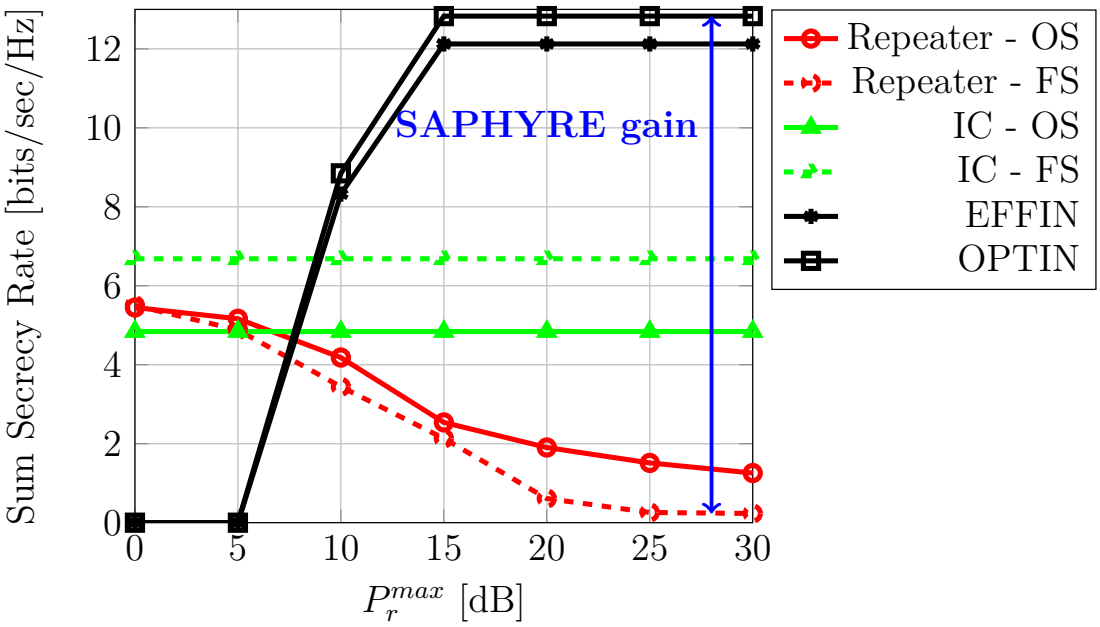


Figure 5.7: The achievable secrecy rates of a two-user IRC with 8 frequency subcarriers is shown with varying relay power constraint. The TX power constraints are 10 dB and there are two antennas at the relay. The proposed scheme EFFIN and OPTIN outperform baseline algorithms Repeater and IC by 550% and 200% respectively.

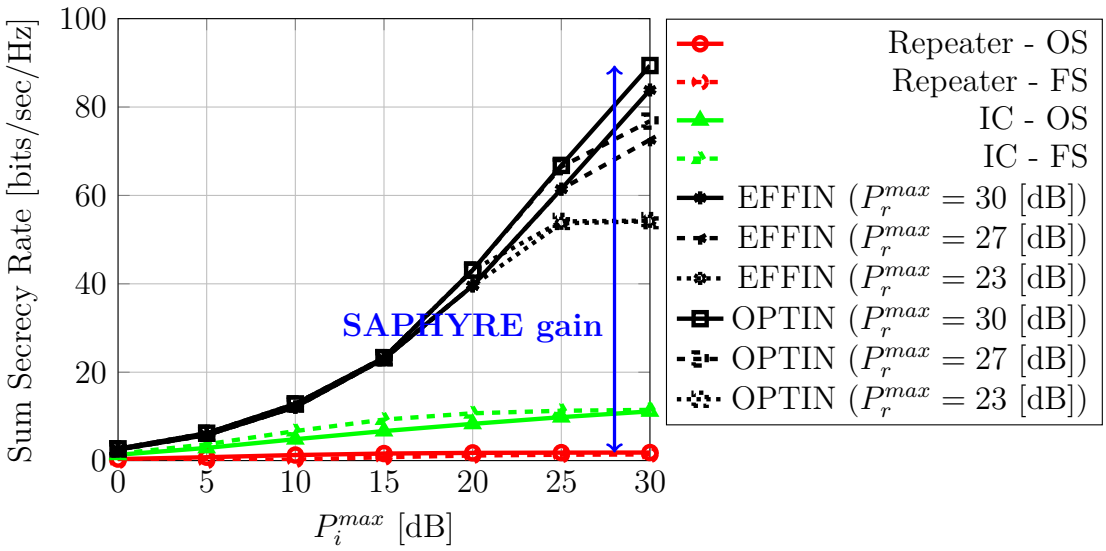


Figure 5.8: The achievable secrecy rates of a two-user IRC with 8 frequency subcarriers is shown with varying transmitter power constraints. The relay power constraint is 30 dB and there are two antennas at the relay. The secrecy rates achieved by EFFIN and OPTIN grows unbounded with the transmit power at TX whereas the secrecy rates achieved by baseline algorithms saturate in high SNR regime.

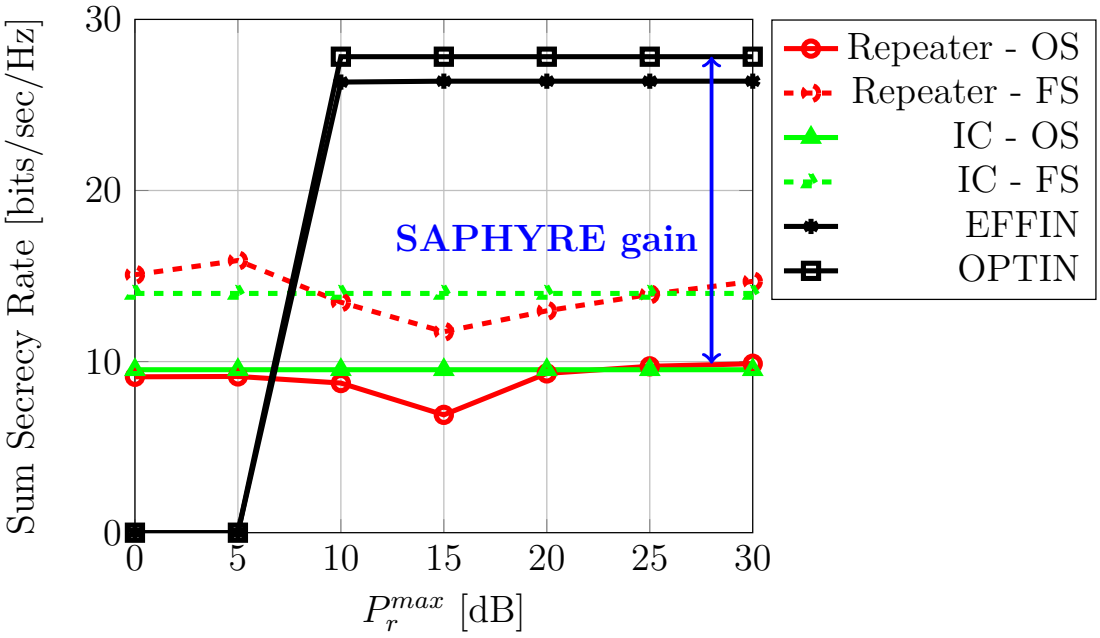


Figure 5.9: The achievable sum secrecy rates of a two-user IRC with 16 frequency subcarriers and 4 antennas at the relay is shown with varying relay power constraint. The TX power constraints are 10 dB and there are two antennas at the relay. The proposed scheme EFFIN and OPTIN outperform baseline algorithms Repeater and IC by 200%. EFFIN achieves 94.86% of the sum secrecy rate performance by OPTIN.

the efficient EFFIN algorithm achieves 94.86% of the sum secrecy rate performance by OPTIN.

5.3 Energy Efficient Relay Sharing with One-Way Amplify-and-Forward MIMO Relays

A multi-user system with a MIMO AF relay relay is considered as shown in Fig. 5.10, where K BSs transmit data to their user terminals (UTs) with the assistance of a shared AF relay which operates in half-duplex mode. Each BS and UT is equipped with $M_{\mathrm{T},k}$ and $M_{\mathrm{U},k}$ antennas, where $k = 1, \ldots, K$. The total number of antennas at the BSs and the UTs are denoted as $M_{\mathrm{T}} = \sum_{k=1}^{K} M_{\mathrm{T},k}$ and $M_{\mathrm{U}} = \sum_{k=1}^{K} M_{\mathrm{U},k}$, respectively. The relay has M_{R} antennas. In this section, the direct links between BSs and UTs are not used since we assume that they are weak due to large path loss or shadowing.

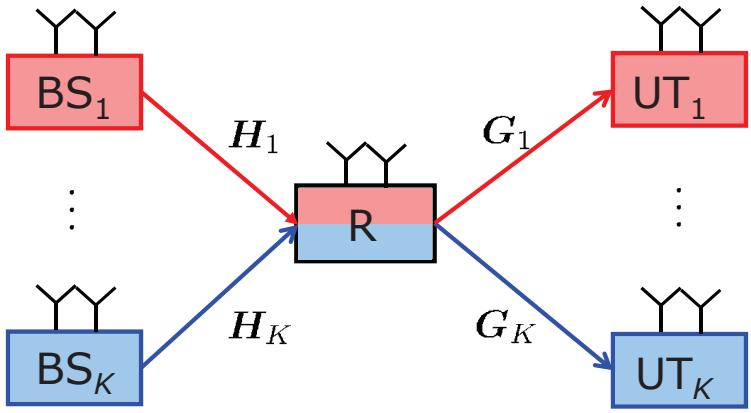


Figure 5.10: System Model for relay sharing between multiple operators

The transmission process consists of two phases. During the multiple access (MAC) phase, the received signal vector at the relay from multiple BSs is

$$\boldsymbol{y}_{\mathrm{R}} = \boldsymbol{H}\tilde{\boldsymbol{s}} + \boldsymbol{n}_{\mathrm{R}},$$

where $\boldsymbol{H} = \left[\boldsymbol{H}_1, \cdots, \boldsymbol{H}_K\right] \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{T}}}$. The matrices $\boldsymbol{H}_k \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{T},k}}$ denote the channels between each BS and the relay. The transmit signal at each BS is $\tilde{\boldsymbol{s}}_k = \boldsymbol{F}_k \boldsymbol{s}_k \in \mathbb{C}^{M_{\mathrm{T},k}}$, where $\boldsymbol{F}_k$ are the precoding matrices at the BSs. In compact form, the transmit signal is written as $\tilde{\boldsymbol{s}} = \left[\tilde{\boldsymbol{s}}_1^{\mathrm{T}}, \cdots, \tilde{\boldsymbol{s}}_K^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{M_{\mathrm{T}}}$. The transmit power at each BS is constrained by P_{T}, i.e., $\mathbb{E}\{\mathrm{tr}(\tilde{\boldsymbol{s}}_k \tilde{\boldsymbol{s}}_k^{\mathrm{H}})\} \leq P_{\mathrm{T}}$. In the broadcasting (BC) phase, the relay amplifies the received signal $\boldsymbol{y}_{\mathrm{R}}$ and forwards it to the UTs. The transmitted signal of the shared relay is

$$\boldsymbol{x}_{\mathrm{R}} = \boldsymbol{F}_{\mathrm{R}} \boldsymbol{y}_{\mathrm{R}} = \boldsymbol{F}_{\mathrm{R}} \boldsymbol{H} \tilde{\boldsymbol{s}} + \boldsymbol{F}_{\mathrm{R}} \boldsymbol{n}_{\mathrm{R}} = \sum_{k=1}^{K} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{H}_k \tilde{\boldsymbol{s}}_k + \boldsymbol{F}_{\mathrm{R}} \boldsymbol{n}_{\mathrm{R}},$$

where $\boldsymbol{F}_{\mathrm{R}} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{R}}}$ denotes the relay amplification matrix. The signals received at the UTs are given by

$$\boldsymbol{y}_k = \boldsymbol{G}_k \boldsymbol{F}_{\mathrm{R}} \boldsymbol{H}_k \tilde{\boldsymbol{s}}_k + \left(\sum_{j=1, j \neq k}^{K} \boldsymbol{G}_k \boldsymbol{F}_{\mathrm{R}} \boldsymbol{H}_j \tilde{\boldsymbol{s}}_j \right) + \boldsymbol{G}_k \boldsymbol{F}_{\mathrm{R}} \boldsymbol{n}_{\mathrm{R}} + \boldsymbol{n}_k \in \mathbb{C}^{M_{\mathrm{U},k}}, k = 1, 2, \cdots, K$$

where $\boldsymbol{G} = \left[\boldsymbol{G}_1^{\mathrm{T}}, \cdots, \boldsymbol{G}_K^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{M_{\mathrm{U}} \times M_{\mathrm{R}}}$ with $\boldsymbol{G}_k \in \mathbb{C}^{M_{\mathrm{U},k} \times M_{\mathrm{R}}}$ denotes the channel between the relay and each UT. The first term denotes the desired signal while the second term stands for the interference that needs to be mitigated. All the remaining terms are the effective noise. The noise at the relay $\boldsymbol{n}_{\mathrm{R}}$ and that at the UTs $\boldsymbol{n}_k$ for $k = 1, 2, \ldots, K$ contain independent, identically distributed complex additive white Gaussian noise samples with variance σ_n^2, i.e., $\mathbb{E}\{\boldsymbol{n}_{\mathrm{R}} \boldsymbol{n}_{\mathrm{R}}^{\mathrm{H}}\} = \sigma_n^2 \boldsymbol{I}_{M_{\mathrm{R}}}$

and $\mathbb{E}\{|n_k|^2\} = \sigma_n^2$. It is assumed that the relay has perfect CSI, i.e., it knows all $\boldsymbol{H}_k$ and $\boldsymbol{G}_k$ perfectly.

In this section, a power efficient transmission is considered for the multiple operator AF relay sharing system. We study the special case that the BSs and the UTs are equipped with single antennas, where only single stream transmission is possible. Further, it is assumed that the transmit signals of all BSs are zero mean and statistically independent with transmit covariance matrix equal to $\mathbb{E}\{\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^{\mathrm{H}}\} = \mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\} = P_{\mathrm{T}}\boldsymbol{I}_{M_{\mathrm{T}}}$. Our objective is to design the relay precoder so as to minimize the relay transmit power $P_{\mathrm{R}} = \mathbb{E}\{\mathrm{tr}(\boldsymbol{x}_{\mathrm{R}}\boldsymbol{x}_{\mathrm{R}}^{\mathrm{H}})\}$ while meeting the signal to interference plus noise ratio (SINR) requirements at each user terminal.

Recently, a method named zero-forcing beamforming (ZFBF) [61] has been designed for relay transmit power minimization subject to an SINR constraint per user, which achieves the local optimum due to the use of the convex optimization tool. It first applies zero-forcing to cancel the interferences between the multiple operators completely. By doing this, the SINR constraint is degraded to an SNR requirement per user, which is easy for the relay precoder design. However, some degrees of freedom are lost, which makes ZFBF a suboptimal solution. Another drawback is that the implementation of ZFBF is not efficient enough due to the use of the convex optimization tool. Therefore, we propose several methods for further improvements in terms of two aspects. Either a global optimum solution is obtained for performance enhancement or the computational complexity is reduced without degrading the power efficiency compared to ZFBF.

This section is structured as follows. A global optimum solution is firstly derived in Subsection 5.3.1, which uses a convex optimization tool to exploit the structure of the relay precoder matrix. Taking this as a benchmark, several suboptimal beamforming algorithms are proposed in Subsection 5.3.2 and Subsection 5.3.3 to find a compromise between the achievable power efficiency and the computational complexity, which are based on zero-forcing and block diagonalization [62] techniques for interference mitigation. The power efficiency performance and the complexity of all the schemes will be compared in the simulation results including the state-of-the-art work in [61]. Following that, the SAPHYRE sharing gain is investigated in terms of the required relay transmit power and the conclusion remarks are made at last.

5.3.1 Optimum precoder design

We first consider an optimum solution of the relay precoder design in the multiple operator AF relay sharing system, where the total relay transmit power is to be minimized while guaranteeing a prescribed SINR constraint at each UT.

As described in the system model, the relay transmit power is expressed as

$$\begin{aligned} P_{\mathrm{R}} &= \mathbb{E}\{\mathrm{tr}(\boldsymbol{x}_{\mathrm{R}}\boldsymbol{x}_{\mathrm{R}}^{\mathrm{H}})\} \\ &= \mathrm{tr}(\boldsymbol{F}_{\mathrm{R}}\boldsymbol{H}\boldsymbol{R}_{\tilde{\boldsymbol{s}}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}) + \sigma_n^2\mathrm{tr}(\boldsymbol{F}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}), \end{aligned} \tag{5.59}$$

where the transmit covariance matrix $\boldsymbol{R}_{\tilde{\boldsymbol{s}}} = \mathbb{E}\{\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^{\mathrm{H}}\} = \mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\} = P_{\mathrm{T}}\boldsymbol{I}_{M_{\mathrm{T}}}$. By inserting this into (5.59) and making use of the property of vec-operator [63], i.e.,

$$\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = (\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}), \tag{5.60}$$

the relay transmit power P_{R} can be further written as

$$\begin{aligned} P_{\mathrm{R}} &= P_{\mathrm{T}}\mathrm{tr}(\boldsymbol{F}_{\mathrm{R}}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}) + \sigma_n^2\mathrm{tr}(\boldsymbol{F}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}) \\ &= P_{\mathrm{T}}\mathrm{tr}\Big\{ (\underbrace{\boldsymbol{H}^{\mathrm{T}} \otimes \boldsymbol{I}_{M_{\mathrm{R}}}}_{\boldsymbol{P}}) \underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{f}_{\mathrm{R}}} \mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})^{\mathrm{H}}(\boldsymbol{H}^{\mathrm{T}} \otimes \boldsymbol{I}_{M_{\mathrm{R}}})^{\mathrm{H}} \Big\} \\ &= P_{\mathrm{T}}\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{f}_{\mathrm{R}}, \end{aligned} \tag{5.61}$$

which is a function of the vectorized relay amplification matrix $\boldsymbol{f}_{\mathrm{R}}$.

At the receiver side, the received signal at UT_k is obtained as

$$y_k = \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{h}_k s_k + \sum_{j\neq k} \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{h}_j s_j + \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{n}_{\mathrm{R}} + n_k, \tag{5.62}$$

where the first term on the right hand side (RHS) denotes the desired signal while the second term represents the inter-operator interference received at user UT_k. The effective noise is given by the remaining RHS terms. Based on (5.62), we express the SINR constraint at UT_k in terms of $\boldsymbol{f}_{\mathrm{R}}$ in the following.

Firstly, the power of the desired signal is calculated as

$$\mathbb{E}\Big\{ \left|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{h}_k s_k\right|^2 \Big\} = P_{\mathrm{T}}\mathrm{tr}\Big\{ \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{g}_k^{*} \Big\} \tag{5.63}$$

with $\mathbb{E}\{|s_k|^2\} = P_{\mathrm{T}}$. Then by use of (5.60), (5.63) is further written as

$$\begin{aligned} P_{\mathrm{T}}\mathrm{tr}\Big\{ \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{g}_k^{*} \Big\} &= P_{\mathrm{T}}\mathrm{tr}\Big\{ (\underbrace{\boldsymbol{h}_k^{\mathrm{T}} \otimes \boldsymbol{g}_k^{\mathrm{T}}}_{\boldsymbol{a}^{\mathrm{T}}}) \underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{f}_{\mathrm{R}}} \mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})^{\mathrm{H}}(\boldsymbol{h}_k^{\mathrm{T}} \otimes \boldsymbol{g}_k^{\mathrm{T}})^{\mathrm{H}} \Big\} \\ &= P_{\mathrm{T}}\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{a}^{*}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{R}} \end{aligned} \tag{5.64}$$

The power of the inter-operator interference is obtained as

$$\mathbb{E}\Big\{ \left|\sum_{j\neq k} \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{h}_j s_j\right|^2 \Big\} = P_{\mathrm{T}}\mathrm{tr}\Big\{ \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\tilde{\boldsymbol{H}}_k\tilde{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{g}_k^{*} \Big\}, \tag{5.65}$$

where the interference matrix for UT_k is defined as $\tilde{\boldsymbol{H}}_k = \left[\boldsymbol{h}_1, \cdots, \boldsymbol{h}_{k-1}, \boldsymbol{h}_{k+1}, \cdots, \boldsymbol{h}_K\right]$. Similarly as (5.64), (5.65) is transformed to

$$P_{\mathrm{T}}\mathrm{tr}\left\{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\tilde{\boldsymbol{H}}_k\tilde{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{g}_k^*\right\} = P_{\mathrm{T}}\mathrm{tr}\left\{\underbrace{(\tilde{\boldsymbol{H}}_k^{\mathrm{T}}\otimes\boldsymbol{g}_k^{\mathrm{T}})}_{\boldsymbol{A}}\underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{f}_{\mathrm{R}}}\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})(\tilde{\boldsymbol{H}}_k^{\mathrm{T}}\otimes\boldsymbol{g}_k^{\mathrm{T}})^{\mathrm{H}}\right\}$$
$$= P_{\mathrm{T}}\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{f}_{\mathrm{R}}. \tag{5.66}$$

Regarding the power of the effective noise,

$$\mathbb{E}\left\{\left|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{n}_{\mathrm{R}}+n_k\right|^2\right\} = \sigma_n^2\mathrm{tr}\left\{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{g}_k^*\right\}+\sigma_n^2, \tag{5.67}$$

where the noise at the relay and the UTs contain the zero-mean circularly symmetric complex Gaussian (ZMCSCG) noise samples with $\mathbb{E}\{\boldsymbol{n}_{\mathrm{R}}\boldsymbol{n}_{\mathrm{R}}^{\mathrm{H}}\} = \sigma_n^2\boldsymbol{I}_{M_{\mathrm{R}}}$ and $\mathbb{E}\{|n_k|^2\} = \sigma_n^2$. (5.67) is further written as

$$\sigma_n^2\mathrm{tr}\left\{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{g}_k^*\right\}+\sigma_n^2 = \sigma_n^2\mathrm{tr}\left\{\underbrace{(\boldsymbol{I}_{M_{\mathrm{R}}}\otimes\boldsymbol{g}_k^{\mathrm{T}})}_{\boldsymbol{B}}\underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{f}_{\mathrm{R}}}\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})(\boldsymbol{I}_{M_{\mathrm{R}}}\otimes\boldsymbol{g}_k^{\mathrm{T}})^{\mathrm{H}}\right\}+\sigma_n^2$$
$$= \sigma_n^2\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{f}_{\mathrm{R}}+\sigma_n^2. \tag{5.68}$$

Combing (5.64), (5.66) and (5.68), the SINR constraint for UT_k is expressed by

$$\frac{P_{\mathrm{T}}\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{a}^*\boldsymbol{a}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{R}}}{P_{\mathrm{T}}\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{f}_{\mathrm{R}}+\sigma_n^2\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{f}_{\mathrm{R}}+\sigma_n^2} \geq \gamma_k, \tag{5.69}$$

which can be written as

$$\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}}\left(P_{\mathrm{T}}\boldsymbol{a}^*\boldsymbol{a}^{\mathrm{T}}-\gamma_k P_{\mathrm{T}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}-\gamma_k\sigma_n^2\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\right)\boldsymbol{f}_{\mathrm{R}} \geq \gamma_k\sigma_n^2. \tag{5.70}$$

By defining $\boldsymbol{W} = \boldsymbol{f}_{\mathrm{R}}\boldsymbol{f}_{\mathrm{R}}^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{R}}^2}$, the problem of minimizing the relay transmit power under SINR constraint is formulated as

$$\begin{aligned}\min_{\boldsymbol{W}} \quad & \mathrm{tr}\{P_{\mathrm{T}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{W}\}\\ \text{s.t.} \quad & \mathrm{tr}\left\{\left(P_{\mathrm{T}}\boldsymbol{a}^*\boldsymbol{a}^{\mathrm{T}}-\gamma_k P_{\mathrm{T}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}-\gamma_k\sigma_n^2\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\right)\boldsymbol{W}\right\} \geq \gamma_k\sigma_n^2, k=1,\cdots,K\\ & \mathrm{rank}(\boldsymbol{W})=1,\end{aligned} \tag{5.71}$$

where

$$\begin{aligned}\boldsymbol{P} &= \boldsymbol{H}^{\mathrm{T}}\otimes\boldsymbol{I}_{M_{\mathrm{R}}},\\ \boldsymbol{a} &= (\boldsymbol{h}_k^{\mathrm{T}}\otimes\boldsymbol{g}_k^{\mathrm{T}})^{\mathrm{T}},\\ \boldsymbol{A} &= \tilde{\boldsymbol{H}}_k^{\mathrm{T}}\otimes\boldsymbol{g}_k^{\mathrm{T}},\\ \boldsymbol{B} &= \boldsymbol{I}_{M_{\mathrm{R}}}\otimes\boldsymbol{g}_k^{\mathrm{T}}.\end{aligned}$$

The original problem in (5.71) is a non-convex quadratically constrained quadratic program (QCQP). By relaxing the non-convex constraint $\text{rank}(\boldsymbol{W}) = 1$ in (5.71), the original problem turns out to be convex in $\boldsymbol{W}$ and can be solved effectively by semi-definite relaxation (SDR) [64, 65] using the convex optimization toolbox cvx [66, 67]. To retrieve $\boldsymbol{f}_{\mathrm{R}}$ from $\boldsymbol{W}$, the rank-1 approximation is performed using the randomization method, which is introduced briefly in the following.

Consider the randomization method, the singular value decomposition (SVD) is first computed $\boldsymbol{W} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}$. To initialize, $\boldsymbol{w}^{(i)}$ is set to $\boldsymbol{w}^{(i)} = \boldsymbol{U}\sqrt{\boldsymbol{\Sigma}}\boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{C}^{M_{\mathrm{R}}^2}$ is a randomly generated zero-mean cyclic symmetric complex Gaussian (ZMCSCG) vector and i is the iteration index. The corresponding $\boldsymbol{W}^{(i)}$ at the i-th iteration is obtained as $\boldsymbol{W}^{(i)} = \boldsymbol{w}^{(i)}\boldsymbol{w}^{(i)\mathrm{H}}$. Note that we need to find a scaling factor $\alpha_k^{(i)}$ for each user to fulfill its SINR requirement exactly, where k, i represents the index of user and the iteration. The coefficient $\alpha_k^{(i)}$ is calculated as

$$\alpha_k^{(i)} = \frac{\gamma_k \sigma_n^2}{\mathrm{tr}\left\{\left(P_{\mathrm{T}}\boldsymbol{a}^*\boldsymbol{a}^{\mathrm{T}} - \gamma_k P_{\mathrm{T}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} - \gamma_k\sigma_n^2\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\right)\boldsymbol{W}\right\}}.$$

Then the coefficient $\alpha^{(i)}$ at the i-th iteration, is selected from $\alpha_k^{(i)}$ such that the minimum SINR for all the users is satisfied. After rescaling $\boldsymbol{w}^{(i)} = \sqrt{\alpha^{(i)}}\boldsymbol{w}^{(i)}$ and $\boldsymbol{W}^{(i)} = \alpha^{(i)}\boldsymbol{w}^{(i)}\boldsymbol{w}^{(i)\mathrm{H}}$, the relay transmit power at the i-th iteration is obtained as $P_{\mathrm{R}}^{(i)} = \alpha^{(i)}\mathrm{tr}\left\{P_{\mathrm{T}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{W}\right\}$ and compare to that from the last iteration $P_{\mathrm{R}}^{(i-1)}$. If $P_{\mathrm{R}}^{(i)}$ is smaller than $P_{\mathrm{R}}^{(i-1)}$, $\boldsymbol{w} = \boldsymbol{w}^{(i)}$ and $P_{\mathrm{R}}^{(i)}$ is used as a new threshold for the next iteration for comparison. Otherwise, $\boldsymbol{w}$ does not change and $P_{\mathrm{R}}^{(i-1)}$ is set as the benchmark for the next iteration. At initialization, a predefined value $P_{\mathrm{R}}^{(0)} = \delta > 0$ is used to start the iterative process. The iteration continues until all the number of iterations L is complete. In the simulation, we set $L = 50$. Tab. 5.2 gives a summary on the randomization method for the rank-1 approximation.

5.3.2 Efficient resource sharing power minimization (EReSh-PM)

In [61], a zero-forcing beamforming (ZFBF) is designed for relay transmit power minimization subject to the SINR constraint per user, which achieves the local optimum. This ZFBF first applies zero-forcing to cancel the inter-operator interferences completely, degrading the SINR constraint to SNR requirement per user. Then the problem turns out to minimize the total relay transmit power with multiple SNR constraints. However, this method needs to be implemented by using the convex optimization tools and iterative procedures are unavoidable. In order to overcome this, we seek a closed-form solution named efficient relay sharing power minimization (EReSh-PM) algorithm. Meanwhile, we will show that the proposed EReSh-PM scheme approaches the performance of the local optimum ZFBF method and no iterations are required.

Table 5.2: Rank-1 relaxation of the optimum solution for single stream transmission

Compute SVD $\boldsymbol{W} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}$
Set $P_{\mathrm{R}}^{(0)} = \delta > 0$
for i = 1: L
$\boldsymbol{w} = \boldsymbol{w}^{(i)} = \boldsymbol{U}\sqrt{\boldsymbol{\Sigma}}\boldsymbol{x}$, $\boldsymbol{x} \in \mathbb{C}^{M_{\mathrm{R}}^2}$ is a randomly generated ZMCSCG vector
$\boldsymbol{W}^{(i)} = \boldsymbol{w}^{(i)}\boldsymbol{w}^{(i)\mathrm{H}}$
$\alpha_k^{(i)} = \dfrac{\gamma_k \sigma_n^2}{\mathrm{tr}\left\{\left(P_{\mathrm{T}}\boldsymbol{a}^*\boldsymbol{a}^{\mathrm{T}} - \gamma_k P_{\mathrm{T}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} - \gamma_k\sigma_n^2\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\right)\boldsymbol{W}\right\}} \rightarrow \alpha^{(i)} = \max(\alpha_k^{(i)})$
$P_{\mathrm{R}}^{(i)} = \alpha^{(i)}\mathrm{tr}\left\{P_{\mathrm{T}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{W}\right\}$
if $P_{\mathrm{R}}^{(i)} < P_{\mathrm{R}}^{(i-1)}$
$\boldsymbol{w} = \sqrt{\alpha^{(i)}}\boldsymbol{w}^{(i)}$
else
$P_{\mathrm{R}}^{(i)} = P_{\mathrm{R}}^{(i-1)}$, $\boldsymbol{w} = \boldsymbol{w}^{(i-1)}$
end
end

In order to simplify the design of $\boldsymbol{F}_{\mathrm{R}}$, the following structure of $\boldsymbol{F}_{\mathrm{R}}$ is used,

$$\boldsymbol{F}_{\mathrm{R}} = \boldsymbol{F}_{\mathrm{BC}}\boldsymbol{\Phi}\boldsymbol{F}_{\mathrm{MAC}}^{\mathrm{H}}. \tag{5.72}$$

The matrices $\boldsymbol{F}_{\mathrm{MAC}}$ and $\boldsymbol{F}_{\mathrm{BC}}$ suppress the interferences generated in the multiple access (MAC) from the BSs to the relay and the broadcasting (BC) phase from the relay to UTs, respectively. The diagonal matrix $\boldsymbol{\Phi} = \mathrm{diag}\{\sqrt{\Phi_1}, \cdots, \sqrt{\Phi_K}\} \in \mathbb{C}^{K\times K}$ is used to allocate the power to each UT.

As the first step, the interference is mitigated during the MAC and the BC phase using the ZF method. To design $\boldsymbol{F}_{\mathrm{MAC}}$, the pseudo inverse of the channel $\boldsymbol{H}$ is first calculated as

$$\boldsymbol{H}^{+} = (\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{H}} = \left[\bar{\boldsymbol{h}}_1, \cdots, \bar{\boldsymbol{h}}_K\right]^{\mathrm{T}} \in \mathbb{C}^{K\times M_{\mathrm{R}}}$$

with $\bar{\boldsymbol{h}}_k \in \mathbb{C}^{M_{\mathrm{R}}}$. Then $\boldsymbol{F}_{\mathrm{MAC}}$ is formed as

$$\boldsymbol{F}_{\mathrm{MAC}} = \left[\boldsymbol{f}_{\mathrm{MAC},1}, \cdots, \boldsymbol{f}_{\mathrm{MAC},K}\right] \in \mathbb{C}^{M_{\mathrm{R}}\times K}, \tag{5.73}$$

where $\boldsymbol{f}_{\mathrm{MAC},k}$ is a unit variance vector and $\boldsymbol{f}_{\mathrm{MAC},k} = \bar{\boldsymbol{h}}_k / \left\|\bar{\boldsymbol{h}}_k\right\|_2 \in \mathbb{C}^{M_{\mathrm{R}}}$.

Similarly in the BC phase, the pseudo inverse of $\boldsymbol{G}$ is

$$\boldsymbol{G}^{+} = \boldsymbol{G}^{\mathrm{H}}(\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}})^{-1} = \left[\bar{\boldsymbol{g}}_1, \cdots, \bar{\boldsymbol{g}}_K\right] \in \mathbb{C}^{M_{\mathrm{R}}\times K}$$

with $\bar{\boldsymbol{g}}_k \in \mathbb{C}^{M_{\mathrm{R}}}$. Then the matrix $\boldsymbol{F}_{\mathrm{BC}}$ is obtained as

$$\boldsymbol{F}_{\mathrm{BC}} = \left[\boldsymbol{f}_{\mathrm{BC},1}, \cdots, \boldsymbol{f}_{\mathrm{BC},K}\right] \in \mathbb{C}^{M_{\mathrm{R}}\times K}, \tag{5.74}$$

where $\boldsymbol{f}_{\mathrm{BC},k} = \bar{\boldsymbol{g}}_k / \left\|\bar{\boldsymbol{g}}_k\right\|_2 \in \mathbb{C}^{M_\mathrm{R}}$.

Inserting (5.73) and (5.74) into (5.72), $\boldsymbol{F}_\mathrm{R}$ is written as

$$\begin{aligned}\boldsymbol{F}_\mathrm{R} &= \boldsymbol{F}_\mathrm{BC}\boldsymbol{\Phi}\boldsymbol{F}_\mathrm{MAC}^\mathrm{H} \\ &= \underbrace{\boldsymbol{f}_{\mathrm{BC},1}\sqrt{\Phi_1}\boldsymbol{f}_{\mathrm{MAC},1}^\mathrm{H}}_{\boldsymbol{F}_{\mathrm{R},1}} + \ldots + \underbrace{\boldsymbol{f}_{\mathrm{BC},K}\sqrt{\Phi_K}\boldsymbol{f}_{\mathrm{MAC},K}^\mathrm{H}}_{\boldsymbol{F}_{\mathrm{R},K}}\end{aligned} \tag{5.75}$$

with

$$\boldsymbol{F}_{\mathrm{R},k} = \boldsymbol{f}_{\mathrm{BC},K}\sqrt{\Phi_K}\boldsymbol{f}_{\mathrm{MAC},K}^\mathrm{H}. \tag{5.76}$$

After applying $\boldsymbol{F}_\mathrm{MAC}$ and $\boldsymbol{F}_\mathrm{BC}$, the inter-operator interferences generated during the MAC and the BC phases are completely removed. Thus the received signal at the UT_k is obtained as

$$y_k = \boldsymbol{g}_k^\mathrm{T}\boldsymbol{f}_{\mathrm{BC},k}\sqrt{\Phi_k}\boldsymbol{f}_{\mathrm{MAC},k}^\mathrm{H}\boldsymbol{h}_k s_k + \boldsymbol{g}_k^\mathrm{T}\boldsymbol{f}_{\mathrm{BC},k}\sqrt{\Phi_k}\boldsymbol{f}_{\mathrm{MAC},k}^\mathrm{H}\boldsymbol{n}_\mathrm{R} + n_k,$$

where the first term on the RHS is the desired signal for UT_k while all the others represent the effective noise. By defining $\hat{g}_k = \boldsymbol{g}_k^\mathrm{T}\boldsymbol{f}_{\mathrm{BC},k}$ and $\hat{h}_k = \boldsymbol{f}_{\mathrm{MAC},k}^\mathrm{H}\boldsymbol{h}_k$, the SINR constraint at each UT is written as

$$\mathrm{SINR}_k = \frac{P_\mathrm{T}|\hat{g}_k|^2|\hat{h}_k|^2\Phi_k}{\sigma_n^2|\hat{g}_k|^2\Phi_k + \sigma_n^2} \geq \gamma_k, \tag{5.77}$$

where γ_k is a predefined constant.

Since the system is decoupled into K independent parallel transceiver pairs, instead of minimizing the total transmit power at the relay $P_\mathrm{R} = \left\|\boldsymbol{F}_\mathrm{R}\boldsymbol{H}\boldsymbol{s} + \boldsymbol{F}_\mathrm{R}\boldsymbol{n}_\mathrm{R}\right\|_2^2$ proposed in [61], we design Φ_k in the way to minimize the summation of the individual relay transmit power of each operator, i.e., $P_{\mathrm{R},k} = \mathbb{E}\left\{ \left\|\boldsymbol{F}_{\mathrm{R},k}\boldsymbol{h}_k s_k + \boldsymbol{F}_{\mathrm{R},k}\boldsymbol{n}_\mathrm{R}\right\|_2^2 \right\}$. Combing (5.76) and the definition of $\hat{h}_k = \boldsymbol{f}_{\mathrm{MAC},k}^\mathrm{H}\boldsymbol{h}_k$, $P_{\mathrm{R},k}$ can be further written as

$$\begin{aligned}P_{\mathrm{R},k} &= \mathbb{E}\left\{ \left\|\boldsymbol{F}_{\mathrm{R},k}\boldsymbol{h}_k s_k + \boldsymbol{F}_{\mathrm{R},k}\boldsymbol{n}_\mathrm{R}\right\|_2^2 \right\} \\ &= \mathbb{E}\left\{ \left\| \boldsymbol{f}_{\mathrm{BC},k}\sqrt{\Phi_k}\boldsymbol{f}_{\mathrm{MAC},k}^\mathrm{H}\boldsymbol{h}_k s_k + \boldsymbol{f}_{\mathrm{BC},k}\sqrt{\Phi_k}\boldsymbol{f}_{\mathrm{MAC},k}^\mathrm{H}\boldsymbol{n}_\mathrm{R}\right\|_2^2 \right\} \\ &= \mathrm{tr}\left\{ \boldsymbol{f}_{\mathrm{BC,k}}(P_\mathrm{T}|\hat{h}_k|^2\Phi_k + \sigma_n^2\Phi_k)\boldsymbol{f}_{\mathrm{BC,k}}^\mathrm{H} \right\} \\ &= \left(P_\mathrm{T}|\hat{h}_k|^2 + \sigma_n^2\right)\Phi_k.\end{aligned} \tag{5.78}$$

The last step comes from that $\boldsymbol{f}_{\mathrm{BC,k}}^\mathrm{H}\boldsymbol{f}_{\mathrm{BC,k}} = \bar{\boldsymbol{g}}_k^\mathrm{H}\bar{\boldsymbol{g}}_k / \left\|\bar{\boldsymbol{g}}_k\right\|_2^2 = 1$ due to $\boldsymbol{f}_{\mathrm{BC,k}}$ is a unit variance vector.

Our objective turns out to find the diagonal elements Φ_k for each user to minimize the individual relay transmit power $P_{\mathrm{R},k}$ in (5.78) under the individual constraint

$\mathrm{SINR}_k \geq \gamma_k$ as described in (5.77). Therefore, the problem is formulated as

$$\begin{aligned} \min_{\Phi_k \geq 0} \quad & \left(P_{\mathrm{T}}|\hat{h}_k|^2 + \sigma_n^2\right)\Phi_k, \\ \text{s.t.} \quad & \frac{P_{\mathrm{T}}|\hat{g}_k|^2|\hat{h}_k|^2\Phi_k}{\sigma_n^2|\hat{g}_k|^2\Phi_k + \sigma_n^2} \geq \gamma_k. \end{aligned} \tag{5.79}$$

It is obvious that $\left(P_{\mathrm{T}}|\hat{h}_k|^2 + 1\right)\Phi_k$ increases monotonically with Φ_k and thereby Φ_k is obtained in a closed-form,

$$\Phi_k = \frac{\gamma_k\sigma_n^2}{|\hat{g}_k|^2\left(P_{\mathrm{T}}|\hat{h}_k|^2 - \gamma_k\sigma_n^2\right)}. \tag{5.80}$$

In order to guarantee that Φ_k is non-negative, the feasibility check has to be passed, i.e.,

$$P_{\mathrm{T}}|\hat{h}_k|^2 - \gamma_k\sigma_n^2 > 0. \tag{5.81}$$

Using equations (5.72) - (5.74) and (5.80) - (5.81), the relay amplification matrix is finally constructed as $\boldsymbol{F}_{\mathrm{R}} = \boldsymbol{F}_{\mathrm{BC}}\boldsymbol{\Phi}\boldsymbol{F}_{\mathrm{MAC}}^{\mathrm{H}}$. The relay transmit power is calculated as $P_{\mathrm{R}} = \|\boldsymbol{F}_{\mathrm{R}}\boldsymbol{H}\boldsymbol{s} + \boldsymbol{F}_{\mathrm{R}}\boldsymbol{n}_{\mathrm{R}}\|_2^2 = \mathrm{tr}\left\{\boldsymbol{F}_{\mathrm{R}}(P_{\mathrm{T}}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \sigma_n^2\boldsymbol{I}_{M_{\mathrm{R}}})\boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}}\right\}$.

5.3.3 BD based solutions

In this section, two alternative solutions are presented for relay power minimization with SINR constraint, which are based on the block diagonalization (BD) [62]. Unlike EReSh-PM method introduced in Section 5.3.2 and the ZFBF beamforming [61], we apply the classical BD method to mitigate the interference generated in the MAC and BC phases.

The relay amplification matrix is still designed as $\boldsymbol{F}_{\mathrm{R}} = \boldsymbol{F}_{\mathrm{BC}}\boldsymbol{\Phi}\boldsymbol{F}_{\mathrm{MAC}} \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{R}}}$, where $\boldsymbol{F}_{\mathrm{MAC}}$ and $\boldsymbol{F}_{\mathrm{BC}}$ are the precoding matrices in the MAC and BC phases for interference cancelation while $\boldsymbol{\Phi}$ is a block diagonal matrix used for relay transmit power minimization. In the MAC phase, we define $\tilde{\boldsymbol{H}}_k = \left[\boldsymbol{h}_1, \cdots, \boldsymbol{h}_{k-1}, \boldsymbol{h}_{k+1}, \cdots, \boldsymbol{h}_K\right] \in \mathbb{C}^{M_{\mathrm{R}}\times(K-1)}$ to represent the channel between all the BSs and relay excluding the k-th BS, where $\boldsymbol{h}_i \in \mathbb{C}^{M_{\mathrm{R}}}$ with $i = 1, \cdots, K, i \neq k$ denote the channel vector from BS_i to the relay. To completely cancel the inter-operator interference generated in the MAC phase for user k, we assign the precoding matrix $\boldsymbol{F}_{\mathrm{MAC},k}$ at BS_k to lie in the null space of the interference channel $\tilde{\boldsymbol{H}}_k$ from all the other users. In order to extract this null space, the singular value decomposition (SVD) of $\tilde{\boldsymbol{H}}_k$ is first computed,

$$\tilde{\boldsymbol{H}}_k = \tilde{\boldsymbol{U}}_{\tilde{\boldsymbol{H}}_k} \cdot \tilde{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{H}}_k} \cdot \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{H}}_k}^{\mathrm{H}} = \left[\tilde{\boldsymbol{U}}_{\tilde{\boldsymbol{H}}_k}^{(1)}, \tilde{\boldsymbol{U}}_{\tilde{\boldsymbol{H}}_k}^{(0)}\right] \cdot \tilde{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{H}}_k} \cdot \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{H}}_k}^{\mathrm{H}}.$$

The matrices $\tilde{\boldsymbol{U}}^{(1)}_{\tilde{\boldsymbol{H}}_k}$ and $\tilde{\boldsymbol{U}}^{(0)}_{\tilde{\boldsymbol{H}}_k}$ are acquired from the left $K-1$ columns and the right $M_{\mathrm{R}}-K+1$ columns of $\tilde{\boldsymbol{U}}_{\tilde{\boldsymbol{H}}_k}$, which stand for the signal and null space of $\tilde{\boldsymbol{H}}_k$, respectively. To mitigate the inter-operator interference received from others, the precoding matrix for user k at the MAC phase is obtained as

$$\boldsymbol{F}_{\mathrm{MAC},k} = \tilde{\boldsymbol{U}}^{(0)}_{\tilde{\boldsymbol{H}}_k}\tilde{\boldsymbol{U}}^{(0)\mathrm{H}}_{\tilde{\boldsymbol{H}}_k} \tag{5.82}$$

To completely cancel the inter-operator interferences between all interfering pairs, the relay precoding matrix at the MAC phase is designed as

$$\boldsymbol{F}_{\mathrm{MAC}} = \begin{bmatrix} \boldsymbol{F}_{\mathrm{MAC},1} \\ \vdots \\ \boldsymbol{F}_{\mathrm{MAC},K} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{U}}^{(0)}_{\tilde{\boldsymbol{H}}_1}\tilde{\boldsymbol{U}}^{(0)\mathrm{H}}_{\tilde{\boldsymbol{H}}_1} \\ \vdots \\ \tilde{\boldsymbol{U}}^{(0)}_{\tilde{\boldsymbol{H}}_K}\tilde{\boldsymbol{U}}^{(0)\mathrm{H}}_{\tilde{\boldsymbol{H}}_K} \end{bmatrix} \in \mathbb{C}^{KM_{\mathrm{R}}\times M_{\mathrm{R}}}.$$

Similarly in the BC phase, we define $\tilde{\boldsymbol{G}}_k = \left[\boldsymbol{g}_1, \cdots, \boldsymbol{g}_{k-1}, \boldsymbol{g}_{k+1}, \cdots, \boldsymbol{g}_K\right]^{\mathrm{T}} \in \mathbb{C}^{(K-1)\times M_{\mathrm{R}}}$ with $\boldsymbol{g}_i \in \mathbb{C}^{M_{\mathrm{R}}}$ denoting the channel vector from the relay to the UT_i. Thereby, the precoding matrix for user k at the BC phase is designed as

$$\boldsymbol{F}_{\mathrm{BC},k} = \tilde{\boldsymbol{V}}^{(0)}_{\tilde{\boldsymbol{G}}_k}\tilde{\boldsymbol{V}}^{(0)\mathrm{H}}_{\tilde{\boldsymbol{G}}_k} \tag{5.83}$$

and thus the precoding matrix $\boldsymbol{F}_{\mathrm{BC}}$ at the BC phase is obtained as

$$\boldsymbol{F}_{\mathrm{BC}} = \left[\boldsymbol{F}_{\mathrm{BC},1}, \cdots, \boldsymbol{F}_{\mathrm{BC},K}\right] = \left[\tilde{\boldsymbol{V}}^{(0)}_{\tilde{\boldsymbol{G}}_1}\tilde{\boldsymbol{V}}^{(0)\mathrm{H}}_{\tilde{\boldsymbol{G}}_1}, \cdots, \tilde{\boldsymbol{V}}^{(0)}_{\tilde{\boldsymbol{G}}_K}\tilde{\boldsymbol{V}}^{(0)\mathrm{H}}_{\tilde{\boldsymbol{G}}_K}\right] \in \mathbb{C}^{M_{\mathrm{R}}\times KM_{\mathrm{R}}},$$

where the columns of $\tilde{\boldsymbol{V}}^{(0)}_{\tilde{\boldsymbol{G}}_k}$ span the null space of $\tilde{\boldsymbol{G}}_k$, which is obtained from computing the SVD of $\tilde{\boldsymbol{G}}_k$, i.e.,

$$\tilde{\boldsymbol{G}}_k = \tilde{\boldsymbol{U}}_{\tilde{\boldsymbol{G}}_k}\cdot\tilde{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{G}}_k}\cdot\tilde{\boldsymbol{V}}^{\mathrm{H}}_{\tilde{\boldsymbol{G}}_k} = \tilde{\boldsymbol{U}}_{\tilde{\boldsymbol{G}}_k}\cdot\tilde{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{G}}_k}\cdot\left[\tilde{\boldsymbol{V}}^{(1)}_{\tilde{\boldsymbol{G}}_k}, \tilde{\boldsymbol{V}}^{(0)}_{\tilde{\boldsymbol{G}}_k}\right]^{\mathrm{H}}.$$

After applying BD at both MAC and BC phases, the received signal vector excluding the noise is obtained as follows,

$$\begin{aligned}
\boldsymbol{G}\boldsymbol{F}_{\mathrm{R}}\boldsymbol{H}\boldsymbol{s} &= \begin{bmatrix} \boldsymbol{g}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{g}_K^{\mathrm{T}} \end{bmatrix} \left[\boldsymbol{F}_{\mathrm{BC},1}, \cdots, \boldsymbol{F}_{\mathrm{BC},K}\right] \underbrace{\begin{bmatrix} \boldsymbol{\Phi}_1 & & \\ & \ddots & \\ & & \boldsymbol{\Phi}_K \end{bmatrix}}_{\boldsymbol{\Phi}} \begin{bmatrix} \boldsymbol{F}_{\mathrm{MAC},1} \\ \vdots \\ \boldsymbol{F}_{\mathrm{MAC},1} \end{bmatrix} \left[\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K\right] \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC},1} & & \\ & \ddots & \\ & & \boldsymbol{g}_K^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC},K} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_1 & & \\ & \ddots & \\ & & \boldsymbol{\Phi}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{\mathrm{MAC},1}\boldsymbol{h}_1 & & \\ & \ddots & \\ & & \boldsymbol{F}_{\mathrm{MAC},K}\boldsymbol{h}_K \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC},1}\boldsymbol{\Phi}_1\boldsymbol{F}_{\mathrm{MAC},1}\boldsymbol{h}_1 & & \\ & \ddots & \\ & & \boldsymbol{g}_K^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC},K}\boldsymbol{\Phi}_K\boldsymbol{F}_{\mathrm{MAC},K}\boldsymbol{h}_K \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}.
\end{aligned}$$

All the off-diagonal elements become zero since all the interferences are removed. The matrix $\boldsymbol{\Phi}$ is block diagonal, written as $\boldsymbol{\Phi} = \text{blockdiag}\{\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_K\} \in \mathbb{C}^{KM_\text{R} \times KM_\text{R}}$. It can be seen that the system is now decoupled into K independent transceiver pairs and more degrees of freedom are achieved for the design of $\boldsymbol{\Phi} \in \mathbb{C}^{KM_\text{R} \times KM_\text{R}}$ after applying BD instead of $\boldsymbol{\Phi} \in \mathbb{C}^{K \times K}$ by utilizing ZF.

The received signal for user k is written as

$$y_k = \boldsymbol{g}_k^\text{T} \boldsymbol{F}_{\text{BC,k}} \boldsymbol{\Phi}_k \boldsymbol{F}_{\text{MAC,k}} \boldsymbol{h}_k s_k + \boldsymbol{g}_k^\text{T} \boldsymbol{F}_{\text{BC,k}} \boldsymbol{\Phi}_k \boldsymbol{F}_{\text{MAC,k}} \boldsymbol{n}_R + n_k. \tag{5.84}$$

where the first term is the desired signal whereas the left terms represent the effective noise vectors.

Since the system is decoupled into K independent transceiver pairs after interference mitigation, the original problem turns out to minimize the relay transmit power for each pair, i.e.,

$$\begin{aligned}
P_{\text{R},k} &= \mathbb{E}\left\{ \left\| \boldsymbol{F}_{\text{BC,k}} \boldsymbol{\Phi}_k \boldsymbol{F}_{\text{MAC,k}} \boldsymbol{h}_k s_k + \boldsymbol{F}_{\text{BC,k}} \boldsymbol{\Phi}_k \boldsymbol{F}_{\text{MAC,k}} \boldsymbol{n}_\text{R} \right\|^2 \right\} \\
&= P_\text{T} \text{tr}\left\{ \boldsymbol{F}_{\text{BC,k}} \boldsymbol{\Phi}_k \boldsymbol{F}_{\text{MAC,k}} \boldsymbol{h}_k \boldsymbol{h}_k^\text{H} \boldsymbol{F}_{\text{MAC,k}}^\text{H} \boldsymbol{\Phi}_k^\text{H} \boldsymbol{F}_{\text{BC,k}}^\text{H} \right\} + \sigma_n^2 \text{tr}\left\{ \boldsymbol{F}_{\text{BC,k}} \boldsymbol{\Phi}_k \boldsymbol{F}_{\text{MAC,k}} \boldsymbol{F}_{\text{MAC,k}}^\text{H} \boldsymbol{\Phi}_k^\text{H} \boldsymbol{F}_{\text{BC,k}}^\text{H} \right\} \\
&= \text{tr}\left\{ \left(P_\text{T} \boldsymbol{\Phi}_k \underbrace{\boldsymbol{F}_{\text{MAC,k}} \boldsymbol{h}_k}_{\triangleq \boldsymbol{b}} \underbrace{\boldsymbol{h}_k^\text{H} \boldsymbol{F}_{\text{MAC,k}}^\text{H}}_{\triangleq \boldsymbol{b}^\text{H}} \boldsymbol{\Phi}_k^\text{H} + \sigma_n^2 \boldsymbol{\Phi}_k \underbrace{\boldsymbol{F}_{\text{MAC,k}} \boldsymbol{F}_{\text{MAC,k}}^\text{H}}_{\boldsymbol{I}_{M_\text{R}}} \boldsymbol{\Phi}_k^\text{H} \right) \underbrace{\boldsymbol{F}_{\text{BC,k}}^\text{H} \boldsymbol{F}_{\text{BC,k}}}_{\boldsymbol{I}_{M_\text{R}}} \right\} \\
&= \text{tr}\left\{ P_\text{T} \boldsymbol{\Phi}_k \boldsymbol{b} \boldsymbol{b}^\text{H} \boldsymbol{\Phi}_k^\text{H} + \sigma_n^2 \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^\text{H} \right\} \\
&= P_\text{T} \left\| \boldsymbol{\Phi}_k \boldsymbol{b} \right\|_2^2 + \sigma_n^2 \left\| \boldsymbol{\Phi}_k \right\|_F^2 \\
&= P_\text{T} \left\| \underbrace{(\boldsymbol{b}^\text{T} \otimes \boldsymbol{I}_{M_\text{R}})}_{\boldsymbol{B}} \underbrace{\text{vec}(\boldsymbol{\Phi}_k)}_{\boldsymbol{w}} \right\|_2^2 + \sigma_n^2 \left\| \underbrace{\text{vec}(\boldsymbol{\Phi}_k)}_{\boldsymbol{w}} \right\|_2^2 \\
&= \boldsymbol{w}^\text{H} \underbrace{(P_\text{T} \boldsymbol{B}^\text{H} \boldsymbol{B} + \sigma_n^2 \boldsymbol{I}_{M_\text{R}})}_{\boldsymbol{D}} \boldsymbol{w} \\
&= \text{tr}\left\{ \boldsymbol{D} \underbrace{\boldsymbol{w} \boldsymbol{w}^\text{H}}_{\boldsymbol{W}_k} \right\} \\
&= \text{tr}\left\{ \boldsymbol{D} \boldsymbol{W}_k \right\}.
\end{aligned} \tag{5.85}$$

Based on (5.84), the SINR constraint for each user is

$$\begin{aligned}
\text{SINR}_k &= \frac{P_\text{T} |\overbrace{\boldsymbol{g}_k^\text{T} \boldsymbol{F}_{\text{BC,k}}}^{\boldsymbol{a}^\text{T}} \boldsymbol{\Phi}_k \overbrace{\boldsymbol{F}_{\text{MAC,k}} \boldsymbol{h}_k}^{\boldsymbol{b}}|^2}{\sigma_n^2 \left\| \boldsymbol{g}_k^\text{T} \boldsymbol{F}_{\text{BC,k}} \boldsymbol{\Phi}_k \boldsymbol{F}_{\text{MAC,k}} \right\|_2^2 + \sigma_n^2} \\
&= \frac{P_\text{T} |\boldsymbol{a}^\text{T} \boldsymbol{\Phi}_k \boldsymbol{b}|^2}{\sigma_n^2 \left\| \boldsymbol{a}^\text{T} \boldsymbol{\Phi}_k \right\|_2^2 + \sigma_n^2},
\end{aligned} \tag{5.86}$$

where $\boldsymbol{a}^{\mathrm{T}} \triangleq \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{BC,k}}$ and $\boldsymbol{b} \triangleq \boldsymbol{F}_{\mathrm{MAC,k}} \boldsymbol{h}_k$. Furthermore, it is obtained that

$$\boldsymbol{a}^{\mathrm{T}} \boldsymbol{\Phi}_k \boldsymbol{b} = \mathrm{vec}(\boldsymbol{a}^{\mathrm{T}} \boldsymbol{\Phi}_k \boldsymbol{b}) = \underbrace{(\boldsymbol{b}^{\mathrm{T}} \otimes \boldsymbol{a}^{\mathrm{T}})}_{\boldsymbol{c}^{\mathrm{T}}} \underbrace{\mathrm{vec}(\boldsymbol{\Phi}_k)}_{\boldsymbol{w}}, \tag{5.87}$$

and

$$\left\| \boldsymbol{a}^{\mathrm{T}} \boldsymbol{\Phi}_k \right\|_2^2 = \left\| \boldsymbol{a}^{\mathrm{T}} \boldsymbol{\Phi}_k \boldsymbol{I}_{M_{\mathrm{R}}} \right\|_2^2 = \left\| \underbrace{(\boldsymbol{I}_{M_{\mathrm{R}}} \otimes \boldsymbol{a}^{\mathrm{T}})}_{\boldsymbol{A}} \underbrace{\mathrm{vec}(\boldsymbol{\Phi}_k)}_{\boldsymbol{w}} \right\|_2^2. \tag{5.88}$$

By inserting (5.87) and (5.88) into (5.86), we have

$$\mathrm{SINR}_k = \frac{P_{\mathrm{T}} \mathrm{tr}\left\{ \boldsymbol{w}^{\mathrm{H}} \boldsymbol{c}^* \boldsymbol{c}^{\mathrm{T}} \boldsymbol{w} \right\}}{\sigma_n^2 \mathrm{tr}\left\{ \boldsymbol{w}^{\mathrm{H}} \boldsymbol{A}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{w} \right\} + \sigma_n^2} \geq \gamma_k,$$

which is finally transformed to

$$\mathrm{tr}\left\{ (P_{\mathrm{T}} \boldsymbol{c}^* \boldsymbol{c}^{\mathrm{T}} - \gamma \sigma_n^2 \boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}) \boldsymbol{W}_k \right\} \geq \gamma_k \sigma_n^2.$$

Thereby, the problem is formulated as

$$\begin{aligned} \min_{\boldsymbol{W}_k} \quad & \mathrm{tr}\left\{ \boldsymbol{D} \boldsymbol{W}_k \right\}, \\ \text{s.t.} \quad & \mathrm{tr}\left\{ (P_{\mathrm{T}} \boldsymbol{c}^* \boldsymbol{c}^{\mathrm{T}} - \gamma \sigma_n^2 \boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}) \boldsymbol{W}_k \right\} \geq \gamma_k \sigma_n^2, \\ & \mathrm{rank}(\boldsymbol{W}_k) = 1. \end{aligned}$$

Without considering the constraint $\mathrm{rank}(\boldsymbol{W}_k) = 1$, $\boldsymbol{W}_k$ can be solved by semi-definite relaxation [64, 65]. To retrieve $\boldsymbol{w}_k$ from $\boldsymbol{W}_k$, the rank-1 approximation is performed either using dominant eigenvector or randomization method, which are introduced briefly in the following.

The dominant eigenvector method is quite simple. We first find a scaling factor of $\boldsymbol{W}_k$ to exactly meet the SINR requirement,

$$\alpha_k = \frac{\gamma_k \sigma_n^2}{\mathrm{tr}\left\{ (P_{\mathrm{T}} \boldsymbol{c}^* \boldsymbol{c}^{\mathrm{T}} - \gamma_k \sigma_n^2 \boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}) \boldsymbol{W}_k \right\}}.$$

Then by computing the SVD $\boldsymbol{W}_k = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}$, the stacked vector of $\boldsymbol{\Phi}_k$ is obtained as

$$\boldsymbol{w} = \sqrt{\alpha_k} \boldsymbol{u}_1 \sqrt{\boldsymbol{\sigma}_1}, \tag{5.89}$$

where $\boldsymbol{u}_1$ is the dominant eigenvector of $\boldsymbol{U}$ and $\boldsymbol{\sigma}_1$ is the largest singular value of $\boldsymbol{W}_k$.

Consider the randomization method, the SVD is first computed $\boldsymbol{W}_k = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}$. To initialize, $\boldsymbol{w}^{(i)}$ is set to $\boldsymbol{w}^{(i)} = \boldsymbol{U} \sqrt{\boldsymbol{\Sigma}} \boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{C}^M$ is a randomly ZMCSCG vector and i is the iteration index. The corresponding $\boldsymbol{W}_k^{(i)}$ is at i-th iteration

Table 5.3: Rank-1 relaxation

Compute SVD $\boldsymbol{W}_k = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}$
Set $P_{\mathrm{R},k}^{(0)} = \delta > 0$
for i = 1: NumIt
  $\boldsymbol{w} = \boldsymbol{w}^{(i)} = \boldsymbol{U}\sqrt{\boldsymbol{\Sigma}}\boldsymbol{x}$, $\boldsymbol{x} \in \mathbb{C}^M$ is a randomly generated ZMCSCG vector
  $\boldsymbol{W}_k^{(i)} = \boldsymbol{w}^{(i)}\boldsymbol{w}^{(i)\mathrm{H}}$
  $\alpha_k^{(i)} = \frac{\gamma_k \sigma_n^2}{\mathrm{tr}\left\{(P_{\mathrm{T}}\boldsymbol{c}^*\boldsymbol{c}^{\mathrm{T}} - \gamma\sigma_n^2\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\boldsymbol{W}_k^{(i)}\right\}}$
  $P_{\mathrm{R},k}^{(i)} = \alpha_k^{(i)}\mathrm{tr}^{(i)}\left\{\boldsymbol{D}\boldsymbol{W}_k^{(i)}\right\}$
  if $P_{\mathrm{R},k}^{(i)} < P_{\mathrm{R},k}^{(i-1)}$
    $\boldsymbol{w} = \sqrt{\alpha_k^{(i)}}\boldsymbol{w}^{(i)}$
  else
    $P_{\mathrm{R},k}^{(i)} = P_{\mathrm{R},k}^{(i-1)}$, $\boldsymbol{w} = \boldsymbol{w}^{(i-1)}$
  end
end

obtained as $\boldsymbol{W}_k^{(i)} = \boldsymbol{w}^{(i)}\boldsymbol{w}^{(i)\mathrm{H}}$. Then we find the coefficients $\alpha_k^{(i)}$ for each user to exactly fulfill its SINR requirement, which is calculated as

$$\alpha_k^{(i)} = \frac{\gamma_k \sigma_n^2}{\mathrm{tr}\left\{(P_{\mathrm{T}}\boldsymbol{c}^*\boldsymbol{c}^{\mathrm{T}} - \gamma_k\sigma_n^2\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\boldsymbol{W}_k^{(i)}\right\}}.$$

The relay transmit power for user k is obtained as $P_{\mathrm{R},k}^{(i)} = \alpha_k^{(i)}\mathrm{tr}^{(i)}\left\{\boldsymbol{D}\boldsymbol{W}_k^{(i)}\right\}$ and compare to that from the last iteration $P_{\mathrm{R},k}^{(i-1)}$. If the obtained $P_{\mathrm{R},k}^{(i)}$ is smaller than $P_{\mathrm{R},k}^{(i-1)}$, $\boldsymbol{w} = \sqrt{\alpha_k^{(i)}}\boldsymbol{w}^{(i)}$ and $P_{\mathrm{R},k}^{(i)}$ is used as a threshold in the next iteration. Otherwise, $\boldsymbol{w}$ does not change and $P_{\mathrm{R},k}^{(i-1)}$ is used for comparison in the next iteration. The iteration continues until all the predefined number of iterations is complete. The Table 5.3 gives a summary on the randomization method for the rank-1 approximation.

Another BD based algorithm is called block diagonalization single channel algebraic norm maximization (BD/SC-ANOMAX), which is a closed-form solution to minimize the relay transmit power under the SINR constraint.

The basic structure of the relay amplification matrix is designed as

$$\boldsymbol{F}_{\mathrm{R}} = \boldsymbol{F}_{\mathrm{BC}}\boldsymbol{\Phi}\boldsymbol{F}_{\mathrm{MAC}} \tag{5.90}$$

As the first step, BD is applied in both MAC and BC phases for complete interference cancellation as previously described. Since the system is decoupled into K

independent transceiver pairs after applying BD for interference mitigation, (5.90) can be further written as

$$\begin{aligned}\boldsymbol{F}_{\mathrm{R}} &= \boldsymbol{F}_{\mathrm{BC}}\boldsymbol{\Phi}\boldsymbol{F}_{\mathrm{MAC}} \\ &= \sum_{k=1}^{K} \boldsymbol{F}_{\mathrm{BC},k}\boldsymbol{\Phi}_k\boldsymbol{F}_{\mathrm{MAC},k} \\ &= \sum_{k=1}^{K} \boldsymbol{F}_k \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{R}}} \end{aligned} \tag{5.91}$$

and $\boldsymbol{F}_{\mathrm{MAC},k}$ and $\boldsymbol{F}_{\mathrm{BC},k}$ are obtained from (5.82) and (5.83). Then the algorithm SC-ANOMAX is to design the diagonal matrix $\boldsymbol{\Phi}_k$.

$$\begin{aligned}\boldsymbol{\Phi}_k &= \underset{\boldsymbol{\Phi}_k,\|\boldsymbol{\Phi}_k\|_F=1}{\arg\max} \left\|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC},k}\boldsymbol{\Phi}_k\boldsymbol{F}_{\mathrm{MAC},k}\boldsymbol{h}_k\right\|_F^2 \\ &= \underset{\boldsymbol{\Phi}_k,\|\boldsymbol{\Phi}_k\|_F=1}{\arg\max} \left\| \underbrace{(\boldsymbol{F}_{\mathrm{MAC},k}\boldsymbol{h}_k)^{\mathrm{T}} \otimes (\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC},k})}_{\boldsymbol{K}} \cdot \underbrace{\mathrm{vec}\{\boldsymbol{\Phi}_k\}}_{\boldsymbol{\phi}_k} \right\|_2^2 \\ &= \underset{\boldsymbol{\phi}_k,\|\boldsymbol{\phi}_k\|_2=1}{\arg\max} \frac{\boldsymbol{\phi}_k^{\mathrm{H}}\boldsymbol{K}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\phi}_k}{\boldsymbol{\phi}_k^{\mathrm{H}}\boldsymbol{\phi}_k} \\ &= \lambda_{\max}(\boldsymbol{K}^{\mathrm{H}}\boldsymbol{K}). \end{aligned} \tag{5.92}$$

By computing the SVD of $\boldsymbol{K} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$, $\boldsymbol{\phi}_k$ is obtained as $\boldsymbol{\phi}_k = \boldsymbol{v}_1$, where $\boldsymbol{v}_1$ denotes the first column of $\boldsymbol{V}$. Then $\boldsymbol{\Phi}_k \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{R}}}$ is formed by the rearrangement of elements of $\boldsymbol{v}_1$.

In order to fulfill the SINR constraint for each user, we define $\tilde{\boldsymbol{F}}_k = \beta_k\boldsymbol{F}_k$, where β is a scalar. Then the SINR constraint is written as

$$\begin{aligned}\mathrm{SINR}_k &= \frac{P_{\mathrm{T}}|\boldsymbol{g}_k^{\mathrm{T}}\tilde{\boldsymbol{F}}_k\boldsymbol{h}_k|^2}{\sigma_n^2\left\|\boldsymbol{g}_k^{\mathrm{T}}\tilde{\boldsymbol{F}}_k\right\|_2^2 + \sigma_n^2} \\ &= \frac{P_{\mathrm{T}}\beta^2|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC,k}}\boldsymbol{\Phi}_k\boldsymbol{F}_{\mathrm{MAC,k}}\boldsymbol{h}_k|^2}{\sigma_n^2\beta^2\left\|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC,k}}\boldsymbol{\Phi}_k\boldsymbol{F}_{\mathrm{MAC,k}}\right\|_2^2 + \sigma_n^2} \geq \gamma_k. \end{aligned}$$

Thereby, the scaling factor β_k for each transceiver pair is obtained as

$$\beta_k = \sqrt{\frac{\gamma_k\sigma_n^2}{P_{\mathrm{T}}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC,k}}\boldsymbol{\Phi}_k\boldsymbol{F}_{\mathrm{MAC,k}}\boldsymbol{h}_k|^2 - \gamma_k\sigma_n^2\left\|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{F}_{\mathrm{BC,k}}\boldsymbol{\Phi}_k\boldsymbol{F}_{\mathrm{MAC,k}}\right\|_2^2}}. \tag{5.93}$$

It is obvious that the feasible set for the BD/SC-ANOMAX solution is that $\beta \geq 0$.

Combing (5.91)-(5.93) and (5.82)-(5.83), the relay amplification matrix is finally designed as

$$\boldsymbol{F}_{\mathrm{R}} = \sum_{k=1}^{K}\tilde{\boldsymbol{F}}_k = \sum_{k=1}^{K}\beta_k\boldsymbol{F}_{\mathrm{BC},k}\boldsymbol{\Phi}_k\boldsymbol{F}_{\mathrm{MAC},k} \tag{5.94}$$

with the feasibility check $\beta \geq 0$.

5.3.4 Simulation results

In this part the proposed algorithms are evaluated via Monte-Carlo simulations. Each element of all channel matrices is a zero mean circularly symmetric complex Gaussian random variable with unit variance. The results are obtained by averaging over 1000 channel realizations.

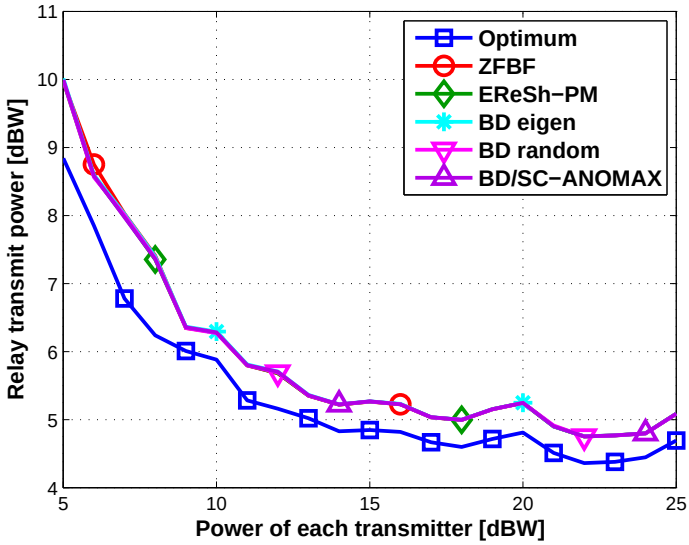


Figure 5.11: P_{R} versus P_{T} with target SINR = 5 dB

Figure 5.11 gives the consumed relay transmit power versus P_{T} using $M_{\mathrm{R}} = 4$. The SINR constraint at each UT is set to be 5 dB and unit noise variance is assumed. Compared to the ZFBF method in [61], our BD based algorithms with rank-1 relaxation using eigenvector method or randomization method obtains the same good performance. Furthermore, the proposed closed-form solutions EReSh-PM as well as BD/SC-ANOMAX algorithms provide the same relay transmit power which achieves the local optimum as ZFBF. Moreover, instead of using the convex optimization tool in [61], the EReSh-PM and the BD/SC-ANOMAX method gives the solution directly and no iteration is required. Compared to the optimum solution which is depicted by the blue curve, all the suboptimal solutions have 1 dB degradation, of which EReSh-PM and BD/SC-ANOMAX methods are proposed due to its extremely low computational complexity.

To verify that, the elapsed time of CPU for all the methods is plotted in Figure 5.12 based on 100 channel realizations. The SINR constraint at each UT is still set to 5 dB and $M_{\mathrm{R}} = 4$. It is observed that the optimum solution and the BD based algorithms with eigen or randomization rank-1 relaxation are more time consuming than other methods due to the use of the convex optimization tool as well as the rank-1 relaxation procedure. Furthermore, although ZFBF can be efficiently

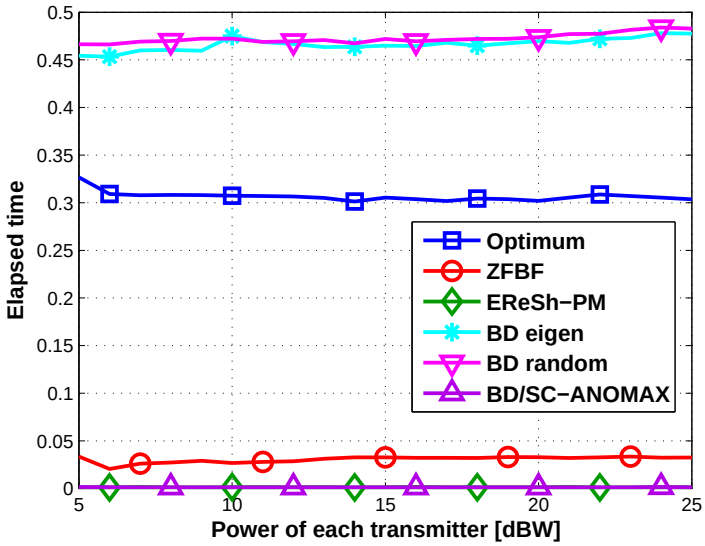


Figure 5.12: Elapsed time versus P_{T} with target SINR = 5 dB

performed with the convex optimization tool, it is apparent to see that the closed-form solutions EReSh-PM and BD/SC-ANOMAX consume even much less time, of which EReSh-PM performs slightly better than BD/SC-ANOMAX.

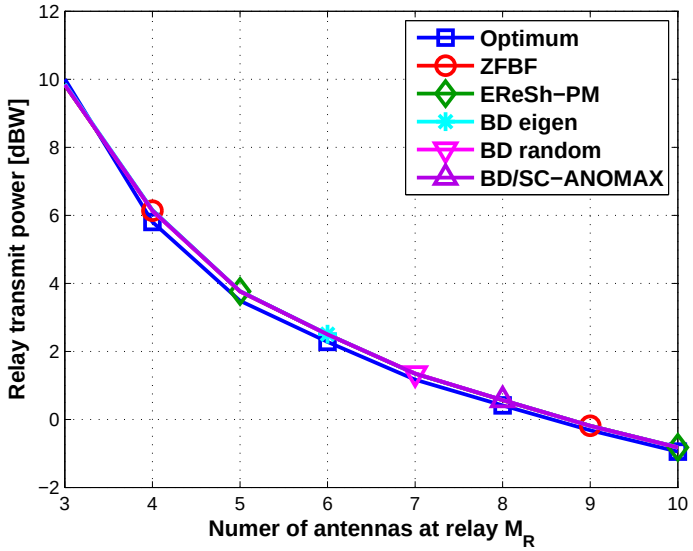


Figure 5.13: P_{R} versus M_{R} with target SINR = 10 dB, P_{T}= 20 dBW

The impact of the number of relay antennas on the relay transmit power is shown in Figure 5.13. Here P_{T} is 10 dBW and the target SINR at each UT is set to 5 dB. All the suboptimal solutions give almost the same performance, which is slightly

worse than the optimum method.

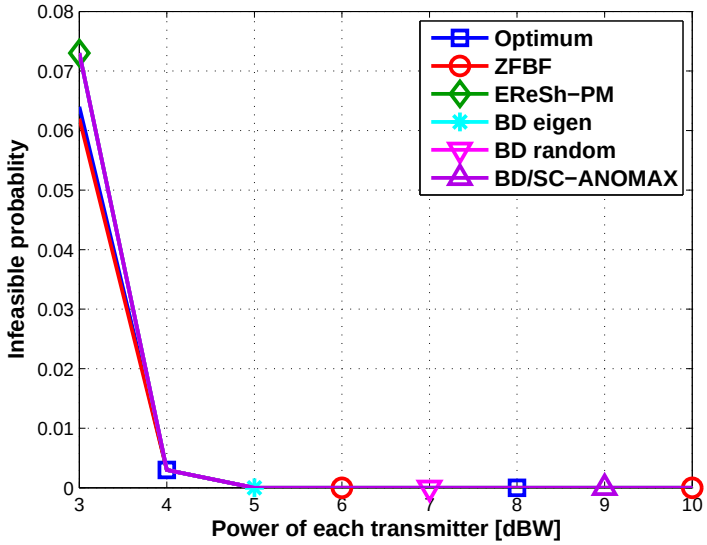


Figure 5.14: Failure probability of the feasibility check versus M_{R} with target SINR $= 5$ dB, $P_{\mathrm{T}} = 10$ dBW

The feasibility check is investigated for all the schemes. For the optimum solution, the convex feasible region is checked in (5.71). Concerning the benchmark work ZFBF [61], the condition $P_{\mathrm{T}}|\hat{h}_k|^2 - \gamma_k > 0$ and $P_{\mathrm{T}} - \gamma_k \left\|\boldsymbol{f}_{\mathrm{MAC,k}}\right\|^2 \geq 0$ must be fulfilled for the implementation of this method. Furthermore, EReSh-PM and BD/SC-ANOMAX are restricted by (5.81) (5.93), respectively. The probability that the feasibility test fails is depicted in Figure 5.14. The transmit power at each BS P_{T} and the target SINR are the same as in Figure 5.13. It is observed that EReSh-PM and BD based solutions have only slightly higher infeasible probability than the optimum and ZFBF methods when M_{R} is smaller due to the limited degree of freedom for interference cancelation. Otherwise, they exhibit the same infeasible probability. Therefore, EReSh-PM and BD/SC-ANOMAX provide a good compromise between the performance and complexity since iterations are avoidable.

In addition to the SAPHYRE gain with respect to the system rate, we can also interpret the SAPHYRE sharing gain in terms of the consumed transmit power. That is, the transmit power consumed in the sharing scenario is compared to the that by the exclusive use of the spectrum and infrastructure for a single operator (TDMA access). The fractional SAPHYRE gain in terms of relay transmit power

is defined as

$$\Xi_{\mathrm{F,power}} = \frac{K\sum_{k=1}^{K} P_k^{\mathrm{SU}}}{\sum_{k=1}^{K} P_k}, \tag{5.95}$$

where the relay transmit power of the k-th user in the sharing scenario and the time division case are denoted by P_k and P_k^{SU}. The numerator denotes the average required transmit power for achieving certain QoS metrics (e.g., minimum required SINR for each user, minimum required total data rate of the network, minimum required SNR per user, etc., in this section we use relay transmit power as the performance metric) in the non-sharing case and the factor K is due to the use of resources in subsequent K time slots. The denominator denotes the required transmit power for achieving the same performance metrics in the sharing case.

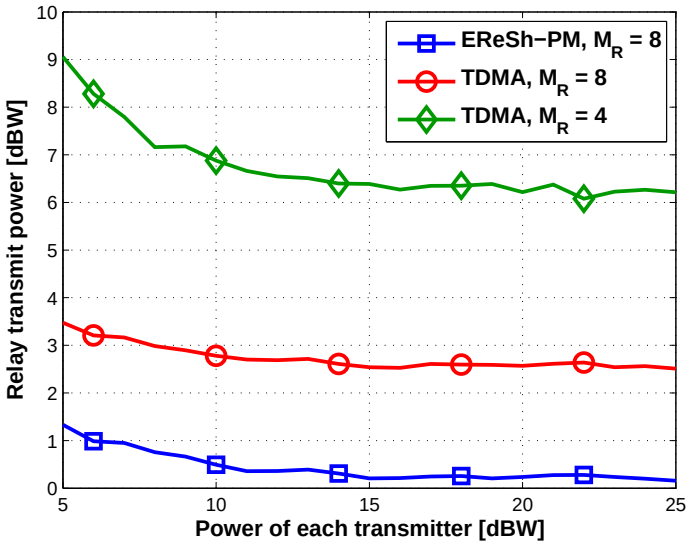


Figure 5.15: SAPHYRE gain in terms of power for multiple operator one-way relaying

For the SAPHYRE scenario, we use the EReSh-PM [68] described in Section 5.3.2 compromising between the performance and the computational complexity. The TDMA scenario is used as a benchmark. At each time slot, the relay amplification matrix is designed for each operator to minimize the relay transmit power subject to the SINR constraint at a single user. The optimum solution for the TDMA access is derived as follows.

In the case that the relay is accessed by each operator, the received signal at the UT is

$$y_k = \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{h}_k s_k + \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{n}_R + n_k.$$

The relay transmit power is calculated as

$$\begin{aligned}
&\mathbb{E}\left\{ \left\| \boldsymbol{F}_{\mathrm{R}} \boldsymbol{h}_k s_k + \boldsymbol{F}_{\mathrm{R}} \boldsymbol{n}_{\mathrm{R}} \right\|^2 \right\} \\
&\qquad = \mathrm{tr}\left\{ P_{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}} \right\} + \mathrm{tr}\left\{ \sigma_n^2 \boldsymbol{F}_{\mathrm{R}} \boldsymbol{F}_{\mathrm{R}}^{\mathrm{H}} \right\} \\
&\qquad = P_{\mathrm{T}} \left\| \boldsymbol{F}_{\mathrm{R}} \boldsymbol{h}_k \right\|_2^2 + \sigma_n^2 \left\| \boldsymbol{F}_{\mathrm{R}} \right\|_F^2 \\
&\qquad = P_{\mathrm{T}} \left\| \underbrace{(\boldsymbol{h}_k^{\mathrm{T}} \otimes \boldsymbol{I}_{M_{\mathrm{R}}})}_{\boldsymbol{B}} \underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{w}} \right\|_2^2 + \sigma_n^2 \left\| \underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{w}} \right\|_2^2 \\
&\qquad = \boldsymbol{w}^{\mathrm{H}} \underbrace{(P_{\mathrm{T}} \boldsymbol{B}^{\mathrm{H}} \boldsymbol{B} + \sigma_n^2 \boldsymbol{I}_{M_{\mathrm{R}}})}_{\boldsymbol{D}} \boldsymbol{w} \\
&\qquad = \mathrm{tr}\left\{ \boldsymbol{D} \boldsymbol{W}_k \right\}. \qquad (5.96)
\end{aligned}$$

The SINR constraint for each UT is

$$\mathrm{SINR}_k = \frac{P_{\mathrm{T}} |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{h}_k|^2}{\sigma_n^2 \left\| \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \right\|_2^2 + \sigma_n^2}. \qquad (5.97)$$

(5.97) could be further simplified using the following equations,

$$\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{h}_k = \mathrm{vec}(\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{h}_k) = \underbrace{(\boldsymbol{h}_k^{\mathrm{T}} \otimes \boldsymbol{g}_k^{\mathrm{T}})}_{\boldsymbol{c}^{\mathrm{T}}} \underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{w}}, \qquad (5.98)$$

and

$$\left\| \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \right\|_2^2 = \left\| \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{F}_{\mathrm{R}} \boldsymbol{I}_{M_{\mathrm{R}}} \right\|_2^2 = \left\| \underbrace{(\boldsymbol{I}_{M_{\mathrm{R}}} \otimes \boldsymbol{g}_k^{\mathrm{T}})}_{\boldsymbol{A}} \underbrace{\mathrm{vec}(\boldsymbol{F}_{\mathrm{R}})}_{\boldsymbol{w}} \right\|_2^2. \qquad (5.99)$$

By inserting (5.98) and (5.99) into (5.97), we have

$$\mathrm{SINR}_k = \frac{P_{\mathrm{T}} \mathrm{tr}\left\{ \boldsymbol{w}^{\mathrm{H}} \boldsymbol{c}^* \boldsymbol{c}^{\mathrm{T}} \boldsymbol{w} \right\}}{\sigma_n^2 \mathrm{tr}\left\{ \boldsymbol{w}^{\mathrm{H}} \boldsymbol{A}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{w} \right\} + \sigma_n^2} \geq \gamma_k,$$

which is finally transformed to

$$\mathrm{tr}\left\{ (P_{\mathrm{T}} \boldsymbol{c}^* \boldsymbol{c}^{\mathrm{T}} - \gamma \sigma_n^2 \boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}) \boldsymbol{W}_k \right\} \geq \gamma_k \sigma_n^2.$$

Thereby, the problem is formulated as

$$\begin{aligned}
\min_{\boldsymbol{W}_k} \quad & \mathrm{tr}\left\{ \boldsymbol{D} \boldsymbol{W}_k \right\}, \\
\text{s.t.} \quad & \mathrm{tr}\left\{ (P_{\mathrm{T}} \boldsymbol{c}^* \boldsymbol{c}^{\mathrm{T}} - \gamma \sigma_n^2 \boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}) \boldsymbol{W}_k \right\} \geq \gamma_k \sigma_n^2, \\
& \mathrm{rank}(\boldsymbol{W}_k) = 1. \qquad (5.100)
\end{aligned}$$

Without considering the constraint $\mathrm{rank}(\boldsymbol{W}_k) = 1$, $\boldsymbol{W}_k$ could be obtained through convex optimization tools. To retrieve $\boldsymbol{w}_k$ from $\boldsymbol{W}_k$, the rank-1 approximation

is performed either using dominant eigenvector or randomization method, which are similar as introduced in Section 5.3.3. The required relay transmit power using EReSh-PM for the SAPHYRE scenario and the TDMA solution for the non-sharing case is plotted in Figure 5.15. The gap between the blue curve obtained by EReSh-PM and the red one for TDMA scenario with $M_{\mathrm{R}} = 8$ denote the spectrum sharing gain. There is around 3 dB loss by making exclusive use of the spectrum with 2 operators. Moreover, the gap between the EReSh-PM method and TDMA access with $M_{\mathrm{R}} = 4$ give the spectrum and the infrastructure sharing gain in the case that half number of the relay antennas are accessed by each operator subsequently.

5.4 Relay Broadcasting Channel with a Two-Way Amplify-and-Forward MIMO Relay

In this section, we study a sharing scenario where both the BS and an amplify-and-forward relay are shared among multiple operators. To develop efficient signal processing algorithms for the sharing scenario, we first consider a multi-user two-way relaying scenario which has the same mathematical model as the sharing scenario. To be specific, we look for transmit strategies which maximize the sum rate of the system. However, finding the sum rate optimal strategy involves a non-tractable optimization problem. To avoid this complex problem, we introduce three sub-optimal algorithms for computing the transmit and receive beamforming matrices at the BS as well as the amplification matrix at the relay. They are based on conventional channel inversion (CI), BD [69] combined with ANOMAX (BD ANOMAX) and ZFDPC (OWR ZFDPC). We also compare our algorithms with the algorithm in reference [70]. It turns out that BD ANOMAX provides the best balance between complexity and performance, OWR ZFDPC can still perform well for large loaded systems, while CI yields the lowest complexity. Compared to the non-sharing case, a significant SAPHYRE gain in terms of sum rate is obtained in the high SNR regime and when there are sufficient number of antennas at the relay.

5.4.1 System model

The scenario under investigation is shown in Figure 5.16. Due to the poor quality of the direct channel between the BS and the UTs, they can only communicate with each other with the help of the relay. Assume that we have K single antenna UTs. The BS is equipped with M_{B} antennas and the relay has M_{R} antennas. For notational simplicity, in the rest of our work we assume that $M_{\mathrm{B}} = K$. The channel is flat fading. The channel between the kth user and the relay is denoted by $\boldsymbol{h}_k \in \mathbb{C}^{M_{\mathrm{R}}}$. The channel between the base station and the relay is full rank and denoted by $\boldsymbol{H}_{\mathrm{B}} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{B}}}$.

The two-way AF relaying protocol consists of two transmission phases: in the first

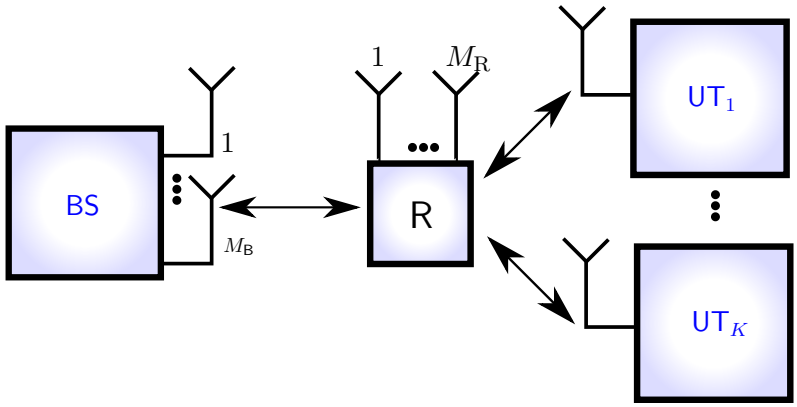


Figure 5.16: Multi-user two-way relaying with a MIMO amplify and forward relay.

phase all the users and the base station transmit their data simultaneously to the relay. Let the BS transmit the data symbol vector $\boldsymbol{d}_{\mathrm{B}} = [d_{\mathrm{B},1}, \ldots, d_{\mathrm{B},K}]^{\mathrm{T}} \in \mathbb{C}^{K}$ using the transmit beamforming matrix $\boldsymbol{F}_{\mathrm{B}} \in \mathbb{C}^{M_{\mathrm{B}} \times K}$. The data symbols in $\boldsymbol{d}_{\mathrm{B}}$ are independently distributed with zero mean and unit variance. Let us further assume that $d_{\mathrm{B},k}$ is the symbol transmitted from the BS to the kth UT and the relay knows the order of the data streams from the BS. The total power at the base station is denoted by P_{B}. The transmit power constraint can be written as

$$\mathrm{E}\{\|\boldsymbol{F}_{\mathrm{B}}\boldsymbol{d}_{\mathrm{B}}\|^2\} = \mathrm{Tr}\{\boldsymbol{F}_{\mathrm{B}}\boldsymbol{F}_{\mathrm{B}}^{\mathrm{H}}\} \leq P_{\mathrm{B}}. \tag{5.101}$$

Then, the received signal vector at the relay is given by

$$\boldsymbol{r} = \sum_{k=1}^{K} \boldsymbol{h}_k \cdot d_k + \boldsymbol{H}_{\mathrm{B}}\boldsymbol{F}_{\mathrm{B}}\boldsymbol{d}_{\mathrm{B}} + \boldsymbol{n}_{\mathrm{R}} \in \mathbb{C}^{M_{\mathrm{R}}}, \tag{5.102}$$

where d_k is the transmitted scalar from the kth user to the BS and $\boldsymbol{n}_{\mathrm{R}} \in \mathbb{C}^{M_{\mathrm{R}}}$ is the zero mean circularly symmetric complex Gaussian (ZMCSCG) noise with $\mathrm{E}\{\boldsymbol{n}_{\mathrm{R}}\boldsymbol{n}_{\mathrm{R}}^{\mathrm{H}}\} = \sigma_{\mathrm{R}}^2\boldsymbol{I}_{M_{\mathrm{R}}}$. Moreover, we assume that each user has identical transmit power P_U and the transmit power constraint is equivalent to $\mathrm{E}\{|d_k|^2\} \leq P_{\mathrm{U}}$.

In the second phase, the relay amplifies the received signal and then forwards it to all the UTs as well as the BS. The signal transmitted by the relay can be expressed as

$$\bar{\boldsymbol{r}} = \gamma_0 \cdot \boldsymbol{G} \cdot \boldsymbol{r}. \tag{5.103}$$

where $\boldsymbol{G} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{R}}}$ is the relay amplification matrix and $\gamma_0 \in \mathcal{R}e^{+}$ is chosen such that the transmit power constraint at the relay is fulfilled, i.e.,

$$\begin{aligned}\mathrm{E}\{\mathrm{Tr}\{\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}^{\mathrm{H}}\}\} &= \mathrm{Tr}\{\gamma_0^2 \cdot \boldsymbol{G}\{P_{\mathrm{U}}\boldsymbol{H}_{\mathrm{U}}\boldsymbol{H}_{\mathrm{U}}^{\mathrm{H}} \\ &+ P_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}\boldsymbol{F}_{\mathrm{B}}\boldsymbol{F}_{\mathrm{B}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{B}}^{\mathrm{H}} + \sigma_{\mathrm{R}}^2\boldsymbol{I}_{M_{\mathrm{R}}}\}\boldsymbol{G}^{\mathrm{H}}\} = P_{\mathrm{R}},\end{aligned} \tag{5.104}$$

where $\boldsymbol{H}_{\mathrm{U}} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K] \in \mathbb{C}^{M_{\mathrm{R}} \times K}$ is the concatenated channel matrix of all UTs.

For notational simplicity, we assume that the reciprocity assumption between the first- and second- phase channels is valid. This assumption is fulfilled in a TDD system if identical RF chains are applied. Then the received signal vector at the BS can be expressed as

$$\begin{aligned}\boldsymbol{y}_{\mathrm{B}} &= \boldsymbol{W}_{\mathrm{B}}(\boldsymbol{H}_{\mathrm{B}}^{\mathrm{T}}\bar{\boldsymbol{r}} + \boldsymbol{n}_{\mathrm{B}}) \\ &= \underbrace{\gamma_0 \boldsymbol{W}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{H}_{\mathrm{U}}\boldsymbol{d}_{\mathrm{U}}}_{\text{useful signal}} + \underbrace{\gamma_0 \boldsymbol{W}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{H}_{\mathrm{B}}\boldsymbol{F}_{\mathrm{B}}\boldsymbol{d}_{\mathrm{B}}}_{\text{self-interference}} \\ &\quad + \underbrace{\gamma_0 \boldsymbol{W}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{n}_{\mathrm{R}} + \boldsymbol{W}_{\mathrm{B}}\boldsymbol{n}_{\mathrm{B}}}_{\text{effective noise}} \in \mathbb{C}^{M_{\mathrm{R}}}\end{aligned} \tag{5.105}$$

where $\boldsymbol{d}_{\mathrm{U}} = [d_1, \ldots, d_K]^{\mathrm{T}} \in \mathbb{C}^K$ is the concatenated data vector of all the UTs and $\boldsymbol{n}_{\mathrm{B}} \in \mathbb{C}^{M_{\mathrm{B}}}$ is the ZMCSCG noise with $\mathrm{E}\{\boldsymbol{n}_{\mathrm{B}}\boldsymbol{n}_{\mathrm{B}}^{\mathrm{H}}\} = \sigma_{\mathrm{B}}^2\boldsymbol{I}_{M_{\mathrm{B}}}$. The receive beamforming matrix is denoted by $\boldsymbol{W}_{\mathrm{B}} \in \mathbb{C}^{K\times M_{\mathrm{B}}}$. It can be seen from (5.105) that the BS only experiences the self-interference caused by its own transmitted signal. If the BS has perfect channel knowledge, the self-interference can be subtracted.

On the other hand, the received scalar y_k at the kth UT can be written as

$$\begin{aligned}y_k = \boldsymbol{h}_k^{\mathrm{T}}\bar{\boldsymbol{r}} + n_k &= \underbrace{\gamma_0 \boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{G}\boldsymbol{H}_{\mathrm{B}}\boldsymbol{f}_{\mathrm{B},k}d_{\mathrm{B},k}}_{\text{useful signal}} \\ &\quad + \underbrace{\gamma_0 \boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{G}\boldsymbol{h}_k d_k}_{\text{self-interference}} + \underbrace{\sum_{\substack{m=1\\m\neq k}}^{K} \gamma_0 \boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{G}\boldsymbol{H}_{\mathrm{B}}\boldsymbol{f}_{\mathrm{B},m}d_{\mathrm{B},m}}_{\text{interference from other streams to other UTs}} \\ &\quad + \underbrace{\sum_{\substack{j=1\\j\neq k}}^{K} \gamma_0 \boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{G}\boldsymbol{h}_j d_j}_{\text{interference from other UTs}} + \underbrace{\gamma_0 \boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{G}\boldsymbol{n}_{\mathrm{R}} + n_k}_{\text{effective noise}}\end{aligned} \tag{5.106}$$

where $\boldsymbol{f}_{\mathrm{B},k}$ is the kth column of $\boldsymbol{F}_{\mathrm{B}}$ and n_k is ZMCSCG noise at each UT with identical variance σ_{U}^2. As can be seen from (5.106), unlike the BS, each UT experiences self-interference, interference caused by other UTs, and the interference caused by the signal which is transmitted from the BS but intended for another UT.

The overall sum rate of the system could be written as

$$R_{\mathrm{sum}} = R_{\mathrm{U}} + R_{\mathrm{B}} \tag{5.107}$$

where R_{B} and R_{U} are the achievable data rate at the BS and the accumulated achievable data rate at all UTs, respectively. The optimization problem to find the relay amplification matrix structure which maximizes (5.107) subject to transmit power constrains in (5.101) and (5.104) is non-convex. To avoid a non-tractable optimization problem, we resort to sub-optimal algorithms instead.

In [70], a linear beamforming is proposed such that

$$\begin{aligned} \boldsymbol{G} &= \gamma_1 (\boldsymbol{H}_{\mathrm{U}}^{\mathrm{T}})^{-1} \boldsymbol{H}_{\mathrm{U}}^{-1} \\ \boldsymbol{F}_{\mathrm{B}} &= \gamma_2 \boldsymbol{H}_{\mathrm{B}}^{-1} \boldsymbol{H}_{\mathrm{U}} \\ \boldsymbol{W}_{\mathrm{B}} &= \boldsymbol{H}_{\mathrm{U}}^{\mathrm{T}} (\boldsymbol{H}_{\mathrm{B}}^{\mathrm{T}})^{-1} \end{aligned} \tag{5.108}$$

where γ_1 and γ_2 are the normalizing coefficients satisfying the transmit power constraint at the relay and the BS, respectively.

However, it can be seen that the inverses of $\boldsymbol{H}_{\mathrm{U}}$ and $\boldsymbol{H}_{\mathrm{B}}$ do not always exist. Hence, this method can hardly be utilized since (5.108) requires that $M_{\mathrm{R}} = M_{\mathrm{B}} = K$. Our algorithms in Sections 5.4.2, 5.4.3, and 5.4.4 are applicable for a broader range of antenna configurations. We specify the corresponding dimensionality constraints below.

Moreover, to have a common framework for the proposed suboptimal solutions, we decompose $\boldsymbol{G}$ into

$$\boldsymbol{G} = \boldsymbol{G}_{\mathrm{T}} \boldsymbol{G}_{\mathrm{S}} \boldsymbol{G}_{\mathrm{R}} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{R}}} \tag{5.109}$$

5.4.2 Channel inversion based design

In this section, we introduce a straightforward beamforming design based on channel inversion. Using this method, orthogonal channels are created between the BS and the UTs for interference free communication. This algorithm can efficiently eliminate the self-interference as well as the co-channel interference. However, the well-known disadvantage of it is the enhancement of the noise power.

Let us define $\boldsymbol{H} = [\boldsymbol{H}_{\mathrm{B}} \ \boldsymbol{H}_{\mathrm{U}}] \in \mathbb{C}^{M_{\mathrm{R}} \times (K + M_{\mathrm{B}})}$. Then the channel inversion receive beamforming is then given by

$$\boldsymbol{G}_{\mathrm{R}} = \boldsymbol{H}^{+} = (\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H})^{-1} \boldsymbol{H}^{\mathrm{H}} \tag{5.110}$$

and the transmit beamforming is given by $\boldsymbol{G}_{\mathrm{T}} = \boldsymbol{G}_{\mathrm{R}}^{\mathrm{T}}$.

In this case, the matrix $\boldsymbol{G}_{\mathrm{S}}$ is chosen to be a block matrix of the form $\boldsymbol{G}_{\mathrm{S}} = \boldsymbol{\Pi}_2 \otimes \boldsymbol{I}_K \in \mathbb{C}^{2 \cdot K \times 2 \cdot K}$, where $\boldsymbol{\Pi}_2 = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ is the exchange matrix which ensures that the BS and the UTs will not receive their own transmitted signals. Furthermore, for simplicity, we choose $\boldsymbol{F}_{\mathrm{B}} = \boldsymbol{W}_{\mathrm{B}} = \sqrt{\frac{P_{\mathrm{B}}}{M_{\mathrm{B}}}} \boldsymbol{I}_{M_{\mathrm{B}}}$. As can be seen from (5.110), this channel inversion method requires that $M_{\mathrm{R}} \geq 2K$. Moreover, compared to the other algorithms proposed in the following sections, the complexity for calculating the Moore-Penrose pseudo inverse is much lower.

5.4.3 BD combined with ANOMAX

For simplicity, we again choose $\boldsymbol{F}_{\mathrm{B}} = \boldsymbol{W}_{\mathrm{B}} = \sqrt{\frac{P_{\mathrm{B}}}{M_{\mathrm{B}}}} \boldsymbol{I}_{M_{\mathrm{B}}}$ in this section. Let us further fix the order of the users such that the kth user communicates with the BS only via

the kth antenna at the BS. Then this system can be regarded as multiple pairs of single-antenna users which communicate with each other with the help of the relay. This scheme is mathematically analogous to multi-pair two-way relaying in [71] or multi-operator two-way AF relaying in [72].

Although all proposed schemes in the above papers can be applied, we recommend BD combined with ANOMAX since according to our work in [72] for the multi-operator two-way relaying case (2 UTs per operator) it provides the best performance. Here we briefly extend this BD ANOMAX method to the multi-user two-way relaying case.

First, partition $\boldsymbol{G}_{\mathrm{T}}$, $\boldsymbol{G}_{\mathrm{S}}$, and $\boldsymbol{G}_{\mathrm{R}}$ as

$$\begin{aligned}\boldsymbol{G}_{\mathrm{T}} &= \left[\boldsymbol{G}_{\mathrm{T}}^{(1)},\ \ldots,\ \boldsymbol{G}_{\mathrm{T}}^{(K)}\right] \in \mathbb{C}^{M_{\mathrm{R}}\times KM_{\mathrm{R}}} \\ \boldsymbol{G}_{\mathrm{S}} &= \mathrm{blkdiag}\left\{\boldsymbol{G}_{\mathrm{S}}^{(1)},\ \ldots,\ \boldsymbol{G}_{\mathrm{S}}^{(K)}\right\} \in \mathbb{C}^{KM_{\mathrm{R}}\times KM_{\mathrm{R}}} \\ \boldsymbol{G}_{\mathrm{R}} &= \left[\boldsymbol{G}_{\mathrm{R}}^{(1)^{\mathrm{T}}},\ \ldots,\ \boldsymbol{G}_{\mathrm{R}}^{(K)^{\mathrm{T}}}\right]^{\mathrm{T}} \in \mathbb{C}^{KM_{\mathrm{R}}\times M_{\mathrm{R}}}\end{aligned} \tag{5.111}$$

where $\boldsymbol{G}_{\mathrm{T}}^{(k)}$, $\boldsymbol{G}_{\mathrm{S}}^{(k)}$, and $\boldsymbol{G}_{\mathrm{R}}^{(k)} \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{R}}}$.

BD ANOMAX consists of two steps. In the first step, the system is converted into K parallel independent sub-systems via the BD design of $\boldsymbol{G}_{\mathrm{R}}$ and $\boldsymbol{G}_{\mathrm{T}}$. Then, in the second step, for each single-pair two-way relaying sub-system, we use the ANOMAX algorithm to calculate $\boldsymbol{G}_{\mathrm{S}}^{(k)}$.

Let us define the combined channel matrix $\tilde{\boldsymbol{H}}^{(k)}$ for all UTs except for the kth UT as

$$\tilde{\boldsymbol{H}}^{(k)} = \left[\boldsymbol{H}^{(1)}\ \ldots\ \boldsymbol{H}^{(k-1)}\ \boldsymbol{H}^{(k+1)}\ \ldots\ \boldsymbol{H}^{(K)}\right], \tag{5.112}$$

where $\boldsymbol{H}^{(k)} = [\boldsymbol{h}_{\mathrm{B},k}\ \boldsymbol{h}_k]$ and $\boldsymbol{h}_{\mathrm{B},k}$ is the kth column of $\boldsymbol{H}_{\mathrm{B}}$.

Let $\tilde{L}^{(k)} = \mathrm{rank}\{\tilde{\boldsymbol{H}}^{(k)}\}$ and calculate the singular value decomposition (SVD)

$$\tilde{\boldsymbol{H}}^{(k)} = [\tilde{\boldsymbol{U}}_{\mathrm{s}}^{(k)}\ \tilde{\boldsymbol{U}}_{\mathrm{n}}^{(k)}]\tilde{\boldsymbol{\Sigma}}^{(k)}\tilde{\boldsymbol{V}}^{(k)^{\mathrm{H}}}. \tag{5.113}$$

where $\tilde{\boldsymbol{U}}_{\mathrm{n}}^{(k)}$ contains the last $(M_{\mathrm{R}} - \tilde{L}^{(k)})$ left singular vectors. Thus, $\tilde{\boldsymbol{U}}_{\mathrm{n}}^{(k)}$ forms an orthogonal basis for the null space of $\tilde{\boldsymbol{H}}^{(k)}$. Therefore, we choose $\boldsymbol{G}_{\mathrm{R}}^{(k)} = \tilde{\boldsymbol{U}}_{\mathrm{n}}^{(k)}\tilde{\boldsymbol{U}}_{\mathrm{n}}^{(k)^{\mathrm{H}}} \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{R}}}$ which is a projection matrix that projects any matrix into the null space of $\tilde{\boldsymbol{H}}^{(k)}$. Due to the channel reciprocity, we can simply set $\boldsymbol{G}_{\mathrm{T}}^{(k)} = \boldsymbol{G}_{\mathrm{R}}^{(k)^{\mathrm{T}}}$.

Next, we define the matrix

$$\begin{aligned}\boldsymbol{K}_{\beta}^{(k)} = \big[&\beta((\boldsymbol{G}_{\mathrm{R}}^{(k)}\boldsymbol{h}_k) \otimes (\boldsymbol{G}_{\mathrm{T}}^{(k)^{\mathrm{T}}}\boldsymbol{h}_{\mathrm{B},k})), \\ &(1-\beta)((\boldsymbol{G}_{\mathrm{R}}^{(k)}\boldsymbol{h}_{\mathrm{B},k}) \otimes (\boldsymbol{G}_{\mathrm{T}}^{(k)^{\mathrm{T}}}\boldsymbol{h}_k)))\big].\end{aligned} \tag{5.114}$$

which is needed to calculate the ANOMAX solution of $\boldsymbol{G}_{\mathrm{S}}^{(k)}$ [73]. The parameter $\beta \in [0, 1]$ is a weighting factor.

Then we compute the SVD of $\boldsymbol{K}_{\beta}^{(k)}$ as $\boldsymbol{K}_{\beta}^{(k)} = \boldsymbol{U}_{\beta}^{(k)}\boldsymbol{\Sigma}_{\beta}^{(k)}\boldsymbol{V}_{\beta}^{(k)^{\mathrm{H}}}$. Let the first column of $\boldsymbol{U}_{\beta}^{(k)}$, i.e., the dominant left singular vector of $\boldsymbol{K}_{\beta}^{(k)}$ be denoted by $\boldsymbol{u}_{\beta,1}^{(k)}$. According to the ANOMAX concept, the matrix $\boldsymbol{G}_{\mathrm{S}}^{(k)}$ is then obtained via

$$\boldsymbol{G}_{\mathrm{S}}^{(k)} = \mathrm{unvec}_{M_{\mathrm{R}}\times M_{\mathrm{R}}}\{\boldsymbol{u}_{\beta,1}^{(k)^*}\}. \tag{5.115}$$

where the operator $\mathrm{unvec}_{M_{\mathrm{R}}\times M_{\mathrm{R}}}\{\cdot\}$ inverts the $\mathrm{vec}\{\cdot\}$ operation by forming a M_{R}-by-M_{R} matrix $\boldsymbol{G}_{\mathrm{S}}^{(k)}$.

In this work we use equal weighting and therefore β is set to 0.5. This algorithm has the dimensionality constraint that $M_{\mathrm{R}} > (2K-2)$.

5.4.4 ZFDPC based design

The multi-antenna BS has the ability of jointly encoding its transmitted data streams or jointly decoding of its received data streams. To further make use of this capability, we introduce the ZFDPC based beamforming design.

Let us partition $\boldsymbol{G}_{\mathrm{R}} = [\boldsymbol{G}_{\mathrm{B}}^{\mathrm{T}}\ \boldsymbol{G}_{\mathrm{U}}^{\mathrm{T}}]^{\mathrm{T}}$ and assume that $\boldsymbol{G}_{\mathrm{T}} = \boldsymbol{G}_{\mathrm{R}}^{\mathrm{T}}$. Moreover, let $L_{\mathrm{U}} = \mathrm{rank}(\boldsymbol{H}_{\mathrm{U}})$ and define the SVD of $\boldsymbol{H}_{\mathrm{U}}$ as

$$\boldsymbol{H}_{\mathrm{U}} = [\boldsymbol{U}_{\mathrm{U,s}}\ \boldsymbol{U}_{\mathrm{U,n}}]\boldsymbol{\Sigma}_{\mathrm{U}}\boldsymbol{V}_{\mathrm{U}}^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{R}}\times K}. \tag{5.116}$$

where $\boldsymbol{U}_{\mathrm{U,n}}$ contains the last $\bar{L}_{\mathrm{U}} = M_{\mathrm{R}} - L_{\mathrm{U}}$ left singular vectors. Thus, with the same reasoning as in Section 5.4.3, we choose $\boldsymbol{G}_{\mathrm{B}} = \boldsymbol{U}_{\mathrm{U,n}}\boldsymbol{U}_{\mathrm{U,n}}^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{R}}}$.

Furthermore, let us define $\boldsymbol{G}_{\mathrm{S}} = \boldsymbol{\Pi}_2 \otimes \boldsymbol{I}_{M_{\mathrm{R}}} \in \mathbb{C}^{2\cdot M_{\mathrm{R}}\times 2\cdot M_{\mathrm{R}}}$ and $\boldsymbol{0}_{K\times K}$ to be the K-by-K matrix with all zero elements. Then the concatenated received signal at the BS and all UTs can be written as

$$\begin{bmatrix}\boldsymbol{y}_{\mathrm{B}}\\ \boldsymbol{y}_{\mathrm{U}}\end{bmatrix} = \underbrace{\begin{bmatrix}\gamma_0\boldsymbol{W}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{H}_{\mathrm{B}}\boldsymbol{F}_{\mathrm{B}} & \boldsymbol{W}_B\boldsymbol{H}_B^{\mathrm{T}}\boldsymbol{G}_{\mathrm{B}}^{\mathrm{T}}\boldsymbol{G}_{\mathrm{U}}\boldsymbol{H}_{\mathrm{U}}\\ \boldsymbol{H}_U^{\mathrm{T}}\boldsymbol{G}_{\mathrm{U}}^{\mathrm{T}}\boldsymbol{G}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}\boldsymbol{F}_{\mathrm{B}} & \boldsymbol{0}_{K\times K}\end{bmatrix}}_{\boldsymbol{H}_{\mathrm{eff}}} \cdot \begin{bmatrix}\boldsymbol{d}_{\mathrm{B}}\\ \boldsymbol{d}_{\mathrm{U}}\end{bmatrix} + \tilde{\boldsymbol{n}} \in \mathbb{C}^{(M_{\mathrm{B}}+K)}. \tag{5.117}$$

In equation (5.117), the first M_{B} rows represent the received signal at the BS ($\boldsymbol{y}_{\mathrm{B}}$). We further assume that the BS has perfect channel knowledge, and thus, the self-interference term which corresponds to the upper left block of $\boldsymbol{H}_{\mathrm{eff}}$ can be subtracted from $\boldsymbol{y}_{\mathrm{B}}$. Then, the system is further decomposed into two-sub systems where the upper right part is equivalent to the uplink of a one-way relay broadcast channel and the lower left part is equivalent to the downlink of a one-way relay multiple access channel. In the next step, we show how to design $\boldsymbol{G}_{\mathrm{U}}$, $\boldsymbol{F}_{\mathrm{B}}$, and $\boldsymbol{W}_{\mathrm{B}}$ using ZFDPC.

ZFDPC is a sub-optimal beamforming solution which has been used in several multi-user MIMO relaying references ([74], [75], [76]). Thus, we will also modify the ZFDPC design for our scenario.

First, we apply the QR decomposition and the SVD to the channel matrices $\boldsymbol{H}_{\mathrm{U}}^{\mathrm{T}}$ and $\boldsymbol{G}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}$ respectively,

$$\boldsymbol{H}_{\mathrm{U}}^{\mathrm{T}} = \boldsymbol{M}_{\mathrm{U}}\boldsymbol{Q}_{\mathrm{U}} \in \mathbb{C}^{K\times M_{\mathrm{R}}}, \tag{5.118}$$

where $\boldsymbol{M}_{\mathrm{U}}$ is a lower triangular matrix and $\boldsymbol{Q}_{\mathrm{U}}$ is a unitary matrix. The singular value decomposition of $\boldsymbol{G}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}}$ is denoted by

$$\boldsymbol{G}_{\mathrm{B}}\boldsymbol{H}_{\mathrm{B}} = \boldsymbol{U}_{\mathrm{B}}\boldsymbol{\Sigma}_{\mathrm{B}}\boldsymbol{V}_{\mathrm{B}}^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{B}}}. \tag{5.119}$$

Then the linear processing matrix $\boldsymbol{G}_{\mathrm{U}}$ can be expressed as:

$$\boldsymbol{G}_{\mathrm{U}} = \boldsymbol{U}_{\mathrm{B}}^{*}\boldsymbol{Q}_{\mathrm{U}}^{*} \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{R}}}. \tag{5.120}$$

Moreover, the precoding matrix $\boldsymbol{F}_{\mathrm{B}}$ is chosen as $\boldsymbol{F}_{\mathrm{B}} = \sqrt{\frac{P_{\mathrm{B}}}{M_{\mathrm{B}}}}\boldsymbol{V}_{\mathrm{B}}$ and the decoding matrix $\boldsymbol{W}_{\mathrm{B}}$ is constructed as $\boldsymbol{W}_{\mathrm{B}} = \boldsymbol{F}_{\mathrm{B}}^{\mathrm{T}} \in \mathbb{C}^{M_{\mathrm{B}}\times M_{\mathrm{B}}}$.

Inserting $\boldsymbol{G}_{\mathrm{U}}$, $\boldsymbol{F}_{\mathrm{B}}$ and $\boldsymbol{W}_{\mathrm{B}}$ into (5.117), the upper right matrix in $\boldsymbol{H}_{\mathrm{eff}}$ is converted into an upper-triangular matrix while the lower left part of it is converted into a lower-triangular matrix. Thus, for each UT, the interference can be canceled by applying a successive interference cancellation (SIC) receiver with perfect knowledge of the interfering signals. Assuming that the BS has also perfect knowledge of the interference signals, it can also utilize a SIC receiver to decode each data stream. Unfortunately, the ZFDPC design has also a dimensionality constraint, which means $M_{\mathrm{R}} \geq 2K$. Furthermore, since this is a non-linear algorithm, it has the highest computational complexity among the three proposed algorithms.

5.4.5 Numerical results

In this section, the performance of the proposed algorithms is evaluated via Monte-Carlo simulations. The simulated MIMO flat fading channels $\boldsymbol{h}_k$ and $\boldsymbol{H}_{\mathrm{B}}$ are spatially uncorrelated Rayleigh fading channels. The SNRs at all nodes are defined as $\mathrm{SNR} = 1/\sigma_{\mathrm{B}}^2 = 1/\sigma_{\mathrm{R}}^2 = 1/\sigma_{\mathrm{U}}^2$. All the simulation results are obtained by averaging over 1000 channel realizations. "CI", "BD ANOMAX", "OWR ZFDPC", and "Toh09" denote the algorithms in Section 5.4.2, 5.4.3, 5.4.4 and [70], respectively. Note that the curves labeled "Toh09" in our results are obtained by using the pseudo-inverse of $\boldsymbol{H}_{\mathrm{B}}$ and $\boldsymbol{H}_{\mathrm{U}}$ in (5.108).

As can be seen from Figure 5.17, "BD ANOMAX" provides the best performance and is 8 dB away form "Toh09" in the high SNR regime. The "OWR ZFDPC" curve is as good as "CI" and is close to "BD ANOMAX". However, it should be noted

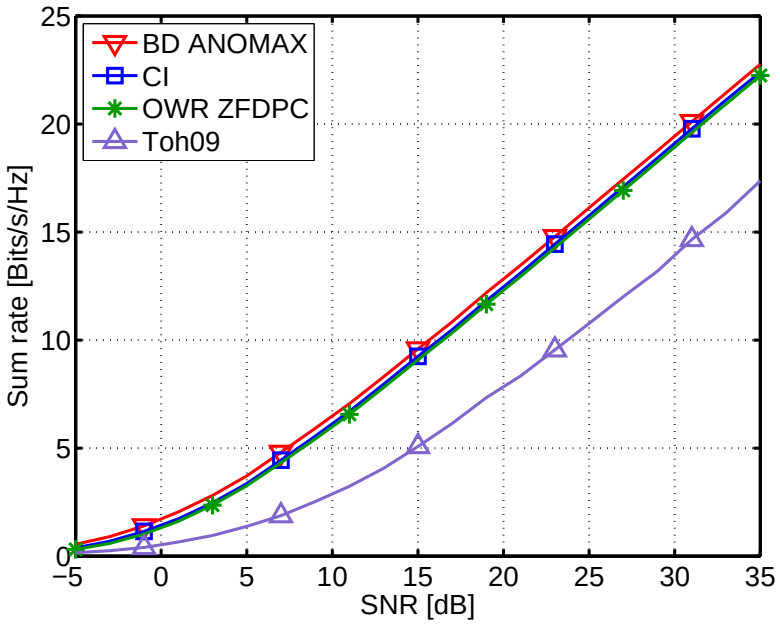


Figure 5.17: Sum rate comparison for $M_{\mathrm{R}} = 8$ and $K = 2$.

that "OWR ZFDPC" has the highest complexity. Moreover, all the curves have the same slope at high SNRs which implies that they possess the same multiplexing gain.

Figure 5.18 show the system loading when $M_{\mathrm{R}} = 20$ and the SNRs at all nodes are 25 dB. It can be seen that due to the dimensionality constraint for "BD ANOMAX" and "CI", there is an inflexion point after which increasing the number of UTs will decrease the system sum rate. For "OWR ZFDPC", although there seems to be also an inflexion point when the system is heavily loaded (at $K = 9$), the sum rate does not drop as quickly as in the case of the other two algorithms.

To demonstrate the SAPHYRE gain in a two-operator case, we compared to a baseline scenario where each operator has only half number of antennas at its BS (e.g., $K = M_{\mathrm{B}} = 3$) and its relay (e.g., $M_{\mathrm{R}} = 8$). In the sharing scenario, both the BS and the relay are shared, i.e., there is only one BS and one relay in the network but they have twice number of antennas than the non-sharing scenario, e.g., the shared BS has $K = M_{\mathrm{B}} = 6$ antennas and the shared relay has $M_{\mathrm{R}} = 16$ antennas. Simulation results in Figure 5.19 have demonstrated an approximate two-fold gain in terms of sum rate is achieved in the high SNR regime and when there are sufficient number of antennas at the relay.

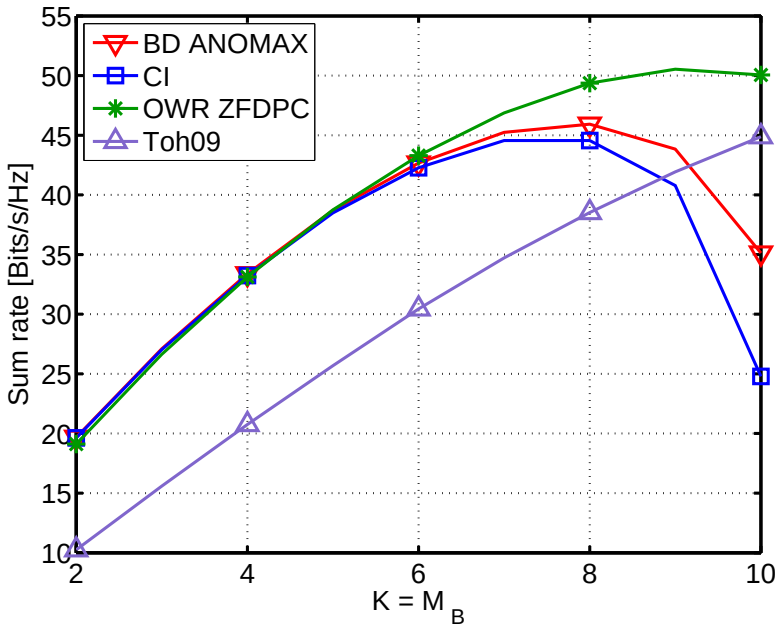
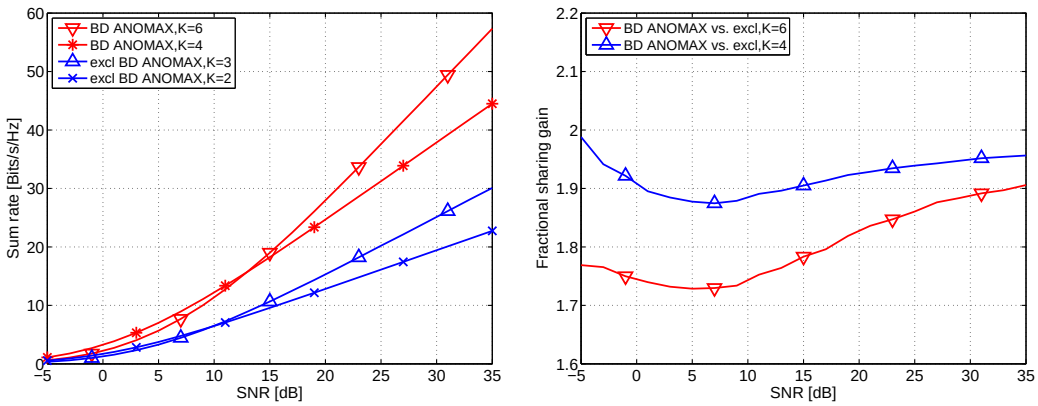


Figure 5.18: Sum rate comparison for $M_{\mathrm{R}} = 20$ and SNR = 25 dB.



(a) Sharing vs. Non-sharing (b) Fractional SAPHYRE gain

Figure 5.19: Sum rate comparison of the SAPHYRE scenario and non-sharing scenario in a two-operator case. BD ANOMAX is used as the precoding technique. In the non-sharing case, the relay has $M_{\mathrm{R}} = 8$ antennas. In the sharing case, the relay has $M_{\mathrm{R}} = 16$ antennas.

5.5 Resource Sharing in Two-Way Relaying Networks with Amplify-and-Forward Relays

In this section, we investigate a relay sharing scenario where the spectrum and multiple relays are shared among different operators. We will first develop signal

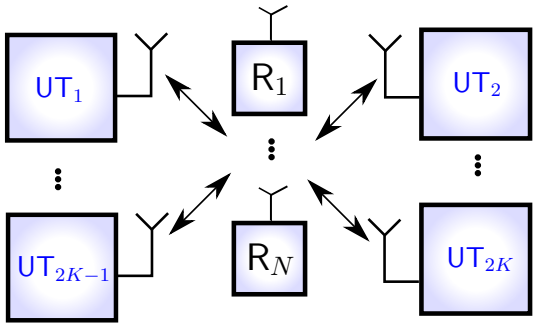


Figure 5.20: Multi-pair two-way relaying with multiple single-antenna amplify and forward relays.

processing algorithms and then demonstrate the achievable SAPHYRE gain. To start, we consider the sum rate maximization problem in a multi-pair two-way AF relaying network where each relay has its individual power constraint, which has the same mathematical formulation as the aforementioned relay sharing scenario. This optimization problem is non-convex and in general NP-hard. Since the maximization problem fulfills the monotonic optimization framework, we first propose an (ϵ, η)-global optimal solution which is obtained using the polyblock approximation algorithm. However, due to its high computational complexity, this algorithm is only suitable for benchmarking. Afterwards, inspired by the polynomial time difference of convex functions (POTDC) method in [77], we develop a sub-optimal algorithm which converges faster than the polyblock algorithm and has a comparable performance. To further reduce the computational complexity, we propose the total SINR eigen-beamformer and the interference neutralization based design which are the low SNR and the high SNR approximation of the original optimization problem, respectively. Simulation results have demonstrated that especially when there are enough relays in the network all the proposed suboptimal algorithms have close to optimum performance. Moreover, the interference neutralization based design yields the lowest computational complexity. Compared to the non-sharing scenario, a significant sharing gain in terms of sum rate is achievable only in the high SNR regime and when there are many relays in the network.

5.5.1 System model

The scenario under investigation is shown in Figure 5.20. K pairs of single-antenna users would like to communicate with each other via the help of N single-antenna relays. We assume perfect synchronization and the channel is frequency flat and quasi-static block fading. The channel vector from the $(2k-1)$th user (on the left-hand side of Figure 5.20) to the relays is denoted as $\boldsymbol{f}_{2k-1} = [f_{2k-1,1}, f_{2k-1,2}, \ldots, f_{2k-1,N}]^{\mathrm{T}} \in \mathbb{C}^N$, while the channel from the $2k$th user (on the right-hand side of Figure 5.20) to the relay is denoted as $\boldsymbol{g}_{2k} = [g_{2k,1}, g_{2k,2}, \ldots, g_{2k,N}]^{\mathrm{T}} \in \mathbb{C}^N$, for $k \in \{1, 2, \cdots, K\}$.

For notational simplicity, we assume an ideal time-division duplex (TDD) system, i.e., the channels are *reciprocal*. The transmission takes two time slots. In the first time slot, the signal received at all relays can be combined in a vector as

$$\boldsymbol{r} = \sum_{k=1}^{K} (\boldsymbol{f}_{2k-1} s_{2k-1} + \boldsymbol{g}_{2k} s_{2k}) + \boldsymbol{n}_{\mathrm{R}} \in \mathbb{C}^N \tag{5.121}$$

where s_{2k-1} and s_{2k} are i.i.d. symbols with zero mean and unit power. The vector $\boldsymbol{n}_{\mathrm{R}}$ contains the zero-mean circularly symmetric complex Gaussian (ZMCSCG) noise and $\mathbb{E}\{\boldsymbol{n}_{\mathrm{R}} \boldsymbol{n}_{\mathrm{R}}^{\mathrm{H}}\} = \sigma_{\mathrm{r}}^2 \boldsymbol{I}_N$.

Afterwards, the AF relays broadcast the weighted signal as

$$\bar{\boldsymbol{r}} = \boldsymbol{W} \cdot \boldsymbol{r} \tag{5.122}$$

where $\boldsymbol{W} = \mathrm{diag}\{\boldsymbol{w}^*\}$ and $\boldsymbol{w} = [w_1, w_2, \ldots, w_N]^{\mathrm{T}}$ is the vector which consists of the N complex weights of all the relays.

In the second time slot, the received signal at the $(2k-1)$th user (on the left-hand side of Figure 5.20) is expressed as [78]

$$\begin{aligned} y_{2k-1} = & \underbrace{\boldsymbol{w}^{\mathrm{H}} \boldsymbol{F}_{2k-1} \boldsymbol{g}_{2k} s_{2k}}_{\text{desired signal}} + \underbrace{\boldsymbol{w}^{\mathrm{H}} \boldsymbol{F}_{2k-1} \boldsymbol{f}_{2k-1} s_{2k-1}}_{\text{self-interfernce}} \\ & + \underbrace{\boldsymbol{w}^{\mathrm{H}} \boldsymbol{F}_{2k-1} \sum_{\substack{\ell \neq k \\ \ell=1}}^{K} (\boldsymbol{f}_{2\ell-1} s_{2\ell-1} + \boldsymbol{g}_{2\ell} s_{2\ell})}_{\text{inter-pair interference}} \\ & + \underbrace{\boldsymbol{w}^{\mathrm{H}} \boldsymbol{F}_{2k-1} \boldsymbol{n}_{\mathrm{R}} + n_{2k-1}}_{\text{effective noise}} \end{aligned} \tag{5.123}$$

where $\boldsymbol{F}_{2k-1} = \mathrm{diag}\{\boldsymbol{f}_{2k-1}\}$ and n_{2k-1} is the ZMCSCG noise with variance $\sigma_{\mathrm{u}}^2, \forall k$. Similar expressions can be obtained for the $2k$th user.

We assume that perfect channel knowledge can be obtained such that the self-interference terms can be canceled and perfect synchronization is available. Let $P_{\mathrm{R},i}$ be the transmit power constraint of the ith relay in the network. Our goal is to find the optimal weighting vector $\boldsymbol{w}_{\mathrm{opt}}$ such that the sum rate of the system is maximized subject to these individual power constraint.

5.5.2 Sum rate maximization

The optimization problem can be formulated as

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \frac{1}{2} \sum_{m=1}^{2K} \log_2(1 + \mathrm{SINR}_m) \\ \text{subject to} \quad & \mathbb{E}\{\|\bar{r}_i\|^2\} \leq P_{\mathrm{R},i}, \forall i \in \{1, 2, \cdots, N\} \end{aligned} \tag{5.124}$$

where the factor 1/2 is due to the two channel uses (half duplex). When $m = 2k-1$, from the expression (5.123), the SINR of the mth UT is described as [79]

$$\mathrm{SINR}_{2k-1} = \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}_{2k-1}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}(\boldsymbol{D}_{2k-1}+\boldsymbol{E}_{2k-1})\boldsymbol{w}+\sigma_{2k-1}^2} \tag{5.125}$$

where $\boldsymbol{D}_{2k-1} = \sum_{\substack{\ell=1\\ \ell\neq k}}(\tilde{\boldsymbol{h}}_{2k-1,\ell}^{(\mathrm{o})}\tilde{\boldsymbol{h}}_{2k-1,\ell}^{(\mathrm{o})^{\mathrm{H}}} + \tilde{\boldsymbol{h}}_{2k-1,\ell}^{(\mathrm{e})}\tilde{\boldsymbol{h}}_{2k-1,\ell}^{(\mathrm{e})^{\mathrm{H}}})$ and $\boldsymbol{B}_{2k-1} = \boldsymbol{h}_{2k-1}\boldsymbol{h}_{2k-1}^{\mathrm{H}}$ are $N \times N$ positive semidefinite Hermitian matrices. The matrices $\boldsymbol{D}_{2k-1}$ and $\boldsymbol{B}_{2k-1}$ are related to the interference power and the desired signal power, respectively, ($\boldsymbol{h}_{2k-1} = \boldsymbol{f}_{2k-1} \odot \boldsymbol{g}_{2k}$, $\tilde{\boldsymbol{h}}_{2k-1,\ell}^{(\mathrm{o})} = \boldsymbol{f}_{2k-1} \odot \boldsymbol{f}_{2\ell-1}$ and $\tilde{\boldsymbol{h}}_{2k-1,\ell}^{(\mathrm{e})} = \boldsymbol{f}_{2k-1} \odot \boldsymbol{g}_{2\ell}$). The term which is related to the forwarded noise from the relay is denoted by an $N \times N$ full rank diagonal matrix $\boldsymbol{E}_{2k-1} = \sigma_{\mathrm{R}}^2\boldsymbol{F}_{2k-1}\boldsymbol{F}_{2k-1}^{\mathrm{H}}$. Similar SINR expressions can be obtained when $m = 2k$. Furthermore, the ith relay's transmit power is given by $\mathbb{E}\{\|\bar{r}_i\|^2\} = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Upsilon}_i\boldsymbol{w}$ with $\boldsymbol{\Upsilon}_i = \Gamma_{i,i}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{H}}$. The vector $\boldsymbol{e}_i$ is the ith column of an identity matrix. The scalar $\Gamma_{i,i}$ is the (i,i)th element of the following diagonal matrix

$$\boldsymbol{\Gamma} = \sum_{k=1}^{K}(\boldsymbol{F}_{2k-1}\boldsymbol{F}_{2k-1}^{\mathrm{H}} + \boldsymbol{G}_{2k}\boldsymbol{G}_{2k}^{\mathrm{H}}) + \sigma_{\mathrm{R}}^2\boldsymbol{I}_N. \tag{5.126}$$

Given the above definitions, problem (5.124) can be rewritten as

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \prod_{m=1}^{2K}\frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{C}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2} \\ \text{subject to} \quad & \boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Upsilon}_i\boldsymbol{w} \leq P_{\mathrm{R},i}, \forall i \end{aligned} \tag{5.127}$$

or equivalently

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \sum_{m=1}^{2K}\left(\log(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2) - \log(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{C}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2)\right) \\ \text{subject to} \quad & \boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Upsilon}_i\boldsymbol{w} \leq P_{\mathrm{R},i}, \forall i \end{aligned} \tag{5.128}$$

where $\boldsymbol{C}_m = \boldsymbol{D}_m + \boldsymbol{E}_m$ and $\boldsymbol{A}_m = \boldsymbol{B}_m + \boldsymbol{C}_m$ are positive definite. Note that for simplicity the scalar $\frac{1}{2}$ is dropped and the natural logarithm is used instead. The formulations (5.127) or (5.128) are still non-convex.

Generalized polyblock algorithm

In [79] we have proven that the sum rate maximization problem in such a relay network with a total power constraint satisfies the monotonic optimization framework. Similarly, problem (5.127) is also a monotonic optimization problem which can be solved using a unified algorithm, which is called the polyblock approximation approach [80]. In the following we prove that the problem (5.127) is a monotonic optimization problem and then adapt the polyblock algorithm to solve it.

Problem (5.127) is equivalent to the following problem

$$\max_{\boldsymbol{y}} \{\Phi(\boldsymbol{y})|\boldsymbol{y} \in \mathbb{D}\} \tag{5.129}$$

where $\Phi(\boldsymbol{y}) = \prod_{m=1}^{2K} y_m$ and $\mathbb{D} = \mathbb{G} \cap \mathbb{L}$. The sets $\mathbb{G} = \{\boldsymbol{y} \in \mathcal{R}e_+^{2K}|y_m \leq \max_{\boldsymbol{w}} \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{C}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2}\}$ and $\mathbb{L} = \{\boldsymbol{y} \in \mathcal{R}e_+^{2K}|y_m \geq \min_{\boldsymbol{w}} \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{C}_m\boldsymbol{w}+\sigma_{\mathrm{u}}^2}\}$ are a normal set and a reverse normal set, respectively [80]. The domain of $\boldsymbol{w}$ is defined as $\{\boldsymbol{w} \in \mathbb{C}^N|\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Upsilon}_i\boldsymbol{w} \leq P_{\mathrm{R},i}, \forall i\}$. Moreover, the function $\Phi(\boldsymbol{y})$ is an increasing function since $\Phi(\bar{\boldsymbol{y}}) \geq \Phi(\tilde{\boldsymbol{y}})$, $\forall \bar{\boldsymbol{y}} \succeq \tilde{\boldsymbol{y}}$. Thereby, problem (5.127) is the maximization of an increasing function over an intersection of normal and reverse normal sets. As shown in [80], such a formulation is a monotonic optimization problem. The definitions of the increasing function, the normal set, and the reverse normal set are the same as in [80].

Now let us briefly review the polyblock algorithm in [79]. A polyblock $\mathbb{P}$ with vertex set $\mathbb{T} \subset \mathcal{R}e_+^{2K}$ is defined as the finite union of all the boxes $[\boldsymbol{0}, \boldsymbol{z}]$, $\boldsymbol{z} \in \mathbb{T}$. It is dominated by its proper vertices. A vertex $\boldsymbol{z}$ is proper if there is no $\bar{\boldsymbol{z}} \neq \boldsymbol{z}$ and $\bar{\boldsymbol{z}} \succeq \boldsymbol{z}$ for $\bar{\boldsymbol{z}} \in \mathbb{T}$. The global maximum of a monotonic optimization problem, if it exists, is obtained on $\partial^+\mathbb{D}$, i.e., the upper boundary of $\mathbb{D}$. The main idea of the polyblock approximation algorithm for solving (5.129) is to approximate $\partial^+\mathbb{D}$ by polyblocks, i.e., to construct a nested sequence of polyblocks which approximate $\mathbb{D}$ from above, that is,

$$\mathbb{P}_1 \supset \mathbb{P}_2 \supset \cdots \supset \mathbb{D} \text{ s.t. } \max_{\boldsymbol{y} \in \mathbb{P}_k} \Phi(\boldsymbol{y}) \to \max_{\boldsymbol{y} \in \mathbb{D}} \Phi(\boldsymbol{y}) \tag{5.130}$$

when $k \to \infty$ and $\boldsymbol{y}_k \succeq \boldsymbol{y}_\ell$ for all $\ell \geq k$ [80].

Following the same procedure as in [79], the (ϵ, η)-optimal solution of problem (5.127) is obtained. Note that the major difference between the problem in [79] and our problem is the calculation of $\alpha_k \in (0, 1]$ at the kth step. The scalar α_k determines the unique intersection between the ray through $\boldsymbol{0}$ and $\bar{\boldsymbol{y}}_k$ and the upper boundary $\partial^+\mathbb{D}$ where $\bar{\boldsymbol{y}}_k$ is the vertex in $\mathbb{T}_k$ which maximizes the function $\Phi(\boldsymbol{y})$. Instead of solving an unconstrained max-min problem as in [79], we need to solve the following constrained problem

$$\begin{aligned} \alpha_k = \max_{\boldsymbol{w}} \min_{m} \quad & \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{A}_m\boldsymbol{w} + \sigma_{\mathrm{u}}^2}{\bar{y}_{k,m}\boldsymbol{w}^{\mathrm{H}}\boldsymbol{C}_m\boldsymbol{w} + \sigma_{\mathrm{u}}^2} \\ \text{subject to} \quad & \boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Upsilon}_i\boldsymbol{w} \leq P_{\mathrm{R},i}, \forall i \end{aligned} \tag{5.131}$$

Similar as in [79], problem (5.131) is solved using semidefinite relaxation together with the bisection search (the concept of this method is elaborated in Section 5.5.2).

POTDC inspired approach

As illustrated in [79], the computational complexity of the polyblock algorithm can be non-polynomial time in the worst case. Thus, it is worth to look for a

polynomial time solution. In this section, we develop a suboptimal but polynomial time solution.

Let us first define $\boldsymbol{X} = \boldsymbol{w}\boldsymbol{w}^{\mathrm{H}}$. Then problem (5.128) can be reformulated as

$$\begin{aligned} \min_{\boldsymbol{X},\alpha_m,\beta_m,\forall m} \quad & -\sum_{m=1}^{2K}\log(\alpha_m) - \left(-\sum_{m=1}^{2K}\log(\beta_m)\right) \\ \text{subject to} \quad & \mathrm{Tr}\{\boldsymbol{\Upsilon}_i\boldsymbol{X}\} \leq P_{\mathrm{R},i}, \forall i, \\ & \mathrm{Tr}\{\boldsymbol{A}_m\boldsymbol{X}\} + \sigma_{\mathrm{u}}^2 = \alpha_m, \\ & \mathrm{Tr}\{\boldsymbol{C}_m\boldsymbol{X}\} + \sigma_{\mathrm{u}}^2 = \beta_m, \forall m, \\ & \boldsymbol{X} \succeq 0, \quad \mathrm{rank}\{\boldsymbol{X}\} = 1. \end{aligned} \tag{5.132}$$

If we further drop the non-convex rank-1 constraint, such a method is called semidefinite relaxation (SDR) [81].

The objective function of problem (5.132) is the difference of convex functions (D.C.) and therefore is non-convex and in general NP-hard. Inspired by the POTDC algorithm in [77], we replace the concave part of the objective function in (5.132) by its linear approximation, i.e., $\log(\beta_m)$ is replaced by its first order Taylor polynomial $\log(\beta_{m,0}) + \frac{\beta_m - \beta_{m,0}}{\beta_{m,0}}, \forall m$. After the substitutions, the cost function in (5.132) becomes convex. Finally, we obtain the following problem:

$$\begin{aligned} \min_{\boldsymbol{X},\alpha_m,\beta_m,t_m\forall m} \quad & -\sum_{m=1}^{2K}\log(\alpha_m) + \sum_{m=1}^{2K} t_m \\ \text{subject to} \quad & \mathrm{Tr}\{\boldsymbol{\Upsilon}_i\boldsymbol{X}\} \leq P_{\mathrm{R},i}, \forall i, \boldsymbol{X} \succeq 0 \\ & \mathrm{Tr}\{\boldsymbol{A}_m\boldsymbol{X}\} + \sigma_{\mathrm{u}}^2 = \alpha_m, \\ & \mathrm{Tr}\{\boldsymbol{C}_m\boldsymbol{X}\} + \sigma_{\mathrm{u}}^2 = \beta_m \\ & \log(\beta_{m,0}) + \frac{1}{\beta_{m,0}}(\beta_m - \beta_{m,0}) \leq t_m. \end{aligned} \tag{5.133}$$

Problem (5.133) is a convex semidefinite programming (SDP) problem and can be solved using the standard interior-point algorithm [82].

Clearly, the first order Taylor polynomial approximation in problem (5.133) is the true Taylor expansion of $\log(\beta_m)$ in (5.132) only if $\beta_{m,0}$ equals to the optimal $\beta_{m,\mathrm{opt}}$. Thus, similarly as in [77], we propose an iterative algorithm as described in Table 5.4 for obtaining the optimal $\boldsymbol{X}_{\mathrm{opt}}$ of problem (5.133). The proposed algorithm has preserved the convergence properties from the original POTDC. That is, the optimal values obtained over the iterations are non-decreasing. Furthermore, the proposed algorithm provides a polynomial-time solution since it solves a sequence of convex problems.

In the end, to obtain $\boldsymbol{w}_{\mathrm{opt}}$ we need to extract a rank-1 solution from $\boldsymbol{X}_{\mathrm{opt}}$. In our work, the randomization technique described in [81] is applied.

Table 5.4: Iterative algorithm for obtaining $\boldsymbol{X}_{\text{opt}}$ inspired by POTDC

Initialization step: set initial values $\beta_{m,0}, \forall m$
maximum iteration number $N_{\max}$ and the tolerance factor ϵ.

Main step:

1: **for** $p = 1$ to $N_{\max}$ **do**
2: Solve (5.133) finding optimal value $f^{\star(p)}$ and $\beta_m^{(p)}$.
3: $\beta_{m,0}^{(p+1)} = \beta_m^{(p)}, m = 1, \cdots, 2K$
4: **if** $\left| f^{\star(p)} - f^{\star(p-1)} \right| \leq \epsilon$ **then**
5: **break**
6: **end if**
7: **end for**

Total SINR eigen-beamformer

Although the POTDC inspired algorithm has a comparable performance and guaranteed polynomial time solution compared to the polyblock algorithm, it requires iterations and therefore is still computationally inefficient. To further reduce the computational complexity, we propose a low SNR approximation of problem (5.124), i.e., the total SINR eigen-beamformer (denoted as ToT in the simulation results). As stated in [79], the total SINR eigen-beamformer aims at maximizing the ratio between the sum of the received signal powers of all the UTs and the sum of interference plus noise power of all the UTs. This beamformer design can be applied to our problem but a closed-form solution as in [79] cannot be obtained due to the individual relay power constraints. In the following we apply the concept of the total SINR eigen-beamformer and develop the solution to it.

Let us define $\boldsymbol{S}_{\text{tot}} = \sum_{m=1}^{2K} \boldsymbol{B}_m$ and $\boldsymbol{U}_{\text{tot}} = \sum_{m=1}^{2K} \boldsymbol{C}_m$. Thus, $\boldsymbol{w}^{\text{H}} \boldsymbol{S}_{\text{tot}} \boldsymbol{w}$ and $\boldsymbol{w}^{\text{H}} \boldsymbol{U}_{\text{tot}} \boldsymbol{w}$ are the sum of the signal power and the sum of the interference power plus the forwarded noise power from all the relays, respectively. Then our proposed total SINR eigen-beamformer solves the following problem

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \frac{\boldsymbol{w}^{\text{H}} \boldsymbol{S}_{\text{tot}} \boldsymbol{w}}{\boldsymbol{w}^{\text{H}} \boldsymbol{U}_{\text{tot}} \boldsymbol{w} + 2K\sigma_{\text{u}}^2} \\ \text{subject to} \quad & \boldsymbol{w}^{\text{H}} \boldsymbol{\Upsilon}_i \boldsymbol{w} \leq P_{\text{R},i}, \forall i \end{aligned} \tag{5.134}$$

Although problem (5.134) is in general non-convex and NP-hard, it is well studied in the literature, e.g., [83], [84]. In our work, we use the SDR together with a bisection search which is similar to [83]. In the following we briefly introduce this

algorithm. Applying the SDR method, problem (5.134) is reformulated as

$$\begin{aligned} \max_{\boldsymbol{X},t} \quad & -t \\ \text{subject to} \quad & \mathrm{Tr}\{\boldsymbol{\Upsilon}_i \boldsymbol{X}\} \leq P_{\mathrm{R},i}, \forall i, \boldsymbol{X} \succeq 0 \\ & \mathrm{Tr}\{(t\boldsymbol{U}_{\mathrm{tot}} - \boldsymbol{S}_{\mathrm{tot}})\boldsymbol{X}\} \leq -2Kt\sigma_{\mathrm{u}}^2, \forall i \end{aligned} \tag{5.135}$$

For a fixed t, problem (5.135) is a feasibility check problem. Thereby, the optimal $\boldsymbol{X}_{\mathrm{opt}}$ can be obtained via a bisection search over an interval $[t_{\min}, t_{\max}]$. In our case, we select $t_{\min} = 0$ and $t_{\max} = \mathcal{P}((\boldsymbol{U}_{\mathrm{tot}} + 2K\sigma_{\mathrm{u}}^2\boldsymbol{\Gamma}/(\sum_i^N P_{\mathrm{R},i}))^{-1}\boldsymbol{S}_{\mathrm{tot}})$ where $\mathcal{P}(\cdot)$ is the dominant eigenvalue of a square matrix. After obtaining $\boldsymbol{X}_{\mathrm{opt}}$, the optimal beamforming vector $\boldsymbol{w}_{\mathrm{opt}}$ is found using the randomization techniques described in [81].

Next we prove that problem (5.134) is the low SNR approximation of the original problem (5.124). Applying the the Taylor expansion of the logarithmic function $\log(1+x)$, we have $\forall m$

$$\log\left(1 + \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}_m\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}(\boldsymbol{D}_m + \boldsymbol{E}_m)\boldsymbol{w} + \sigma_{\mathrm{u}}^2}\right) \approx \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}_m\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}(\boldsymbol{D}_m + \boldsymbol{E}_m)\boldsymbol{w} + \sigma_{\mathrm{u}}^2}.$$

Using the fact that in the low SNR regime ($\sigma_{\mathrm{r}}^2 \to +\infty$) we have $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{D}_m\boldsymbol{w} \ll \boldsymbol{w}^{\mathrm{H}}\boldsymbol{E}_m\boldsymbol{w} \approx \sigma_{\mathrm{r}}^2, \forall m$, we can rewrite the objective function in (5.124) as

$$\begin{aligned} & \frac{1}{2K}\sum_{m=1}^{2K}\log\left(1 + \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}_m\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}(\boldsymbol{D}_m + \boldsymbol{E}_m)\boldsymbol{w} + \sigma_{\mathrm{u}}^2}\right) \\ \approx & \frac{1}{2K}\sum_{m=1}^{2K}\frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}_m\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}(\boldsymbol{D}_m + \boldsymbol{E}_m)\boldsymbol{w} + \sigma_{\mathrm{u}}^2} \approx \frac{\sum_{m=1}^{2K}\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}_m\boldsymbol{w}}{2K(\sigma_{\mathrm{r}}^2 + \sigma_{\mathrm{u}}^2)} \\ \approx & \frac{\sum_{m=1}^{2K}\boldsymbol{w}^{\mathrm{H}}\boldsymbol{B}_m\boldsymbol{w}}{\sum_{m=1}^{2K}(\boldsymbol{w}^{\mathrm{H}}(\boldsymbol{D}_m + \boldsymbol{E}_m)\boldsymbol{w} + \sigma_{\mathrm{u}}^2)} = \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{S}_{\mathrm{tot}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{U}_{\mathrm{tot}}\boldsymbol{w} + 2K\sigma_{\mathrm{u}}^2}. \end{aligned}$$

Interference neutralization based design

In this section, we propose a high SNR approximation of the original problem (5.124). The proposed algorithm is based on the interference neutralization which is a technique that tunes the interfering signals such that they neutralize each other at the receiver [85]. Mathematically, interference neutralization for our scenario

requires that

$$\begin{cases} (\boldsymbol{f}_{2k-1} \odot \boldsymbol{f}_{2\ell-1})^{\mathrm{H}} \boldsymbol{w} \cdot s_{2\ell-1} = 0 & \forall \ell, k, \ell \neq k \\ (\boldsymbol{f}_{2k-1} \odot \boldsymbol{g}_{2\ell})^{\mathrm{H}} \boldsymbol{w} \cdot s_{2\ell} = 0 & \forall \ell, k, \ell \neq k \\ (\boldsymbol{g}_{2k} \odot \boldsymbol{f}_{2\ell-1})^{\mathrm{H}} \boldsymbol{w} \cdot s_{2\ell-1} = 0 & \forall \ell, k, \ell \neq k \\ (\boldsymbol{g}_{2k} \odot \boldsymbol{g}_{2\ell})^{\mathrm{H}} \boldsymbol{w} \cdot s_{2\ell} = 0 & \forall \ell, k, \ell \neq k \end{cases} \tag{5.136}$$

or equivalently

$$\underbrace{\begin{bmatrix} (\boldsymbol{f}_1 \odot \boldsymbol{f}_3)^{\mathrm{H}} \\ (\boldsymbol{f}_1 \odot \boldsymbol{g}_4)^{\mathrm{H}} \\ (\boldsymbol{g}_2 \odot \boldsymbol{f}_3)^{\mathrm{H}} \\ (\boldsymbol{g}_2 \odot \boldsymbol{g}_4)^{\mathrm{H}} \\ \vdots \\ (\boldsymbol{f}_{2K-3} \odot \boldsymbol{f}_{2K-1})^{\mathrm{H}} \\ (\boldsymbol{f}_{2K-3} \odot \boldsymbol{g}_{2K})^{\mathrm{H}} \\ (\boldsymbol{g}_{2K-2} \odot \boldsymbol{f}_{2K-1})^{\mathrm{H}} \\ (\boldsymbol{g}_{2K-2} \odot \boldsymbol{g}_{2K})^{\mathrm{H}} \end{bmatrix}}_{\boldsymbol{H}^{(\mathrm{e})}} \cdot \boldsymbol{w} = \boldsymbol{0} \tag{5.137}$$

Utilizing the commutative property of the Hadamard product, duplicat rows in $\boldsymbol{H}^{(\mathrm{e})}$ are removed. The matrix $\boldsymbol{H}^{(\mathrm{e})}$ is found to have a dimension of $2K(K-1) \times N$. Equation (5.137) is solvable only if the null space of $\boldsymbol{H}^{(\mathrm{e})}$ is not empty, i.e., $N > 2K(K-1)$. Define the SVD of $\boldsymbol{H}^{(\mathrm{e})} = \boldsymbol{U}\boldsymbol{\Sigma}[\boldsymbol{V}_{\mathrm{s}} \quad \boldsymbol{V}_{\mathrm{n}}]^{\mathrm{H}}$, where $\boldsymbol{V}_{\mathrm{n}}$ contains the last $(N - 2K(K-1))$ right singular vectors and thus forms an orthonormal basis for the null subspace of $\boldsymbol{H}^{(\mathrm{e})}$. Without loss of generality, we define the interference neutralization based beamformer (denoted as IntNeu in the simulation results) as $\boldsymbol{w} = \boldsymbol{V}_{\mathrm{n}}\bar{\boldsymbol{w}}$, where $\bar{\boldsymbol{w}} \in \mathbb{C}^{N-2K^2+2K}$ has a smaller dimension than $\boldsymbol{w} \in \mathbb{C}^N$. In other words, searching over $\bar{\boldsymbol{w}}$ yields a lower computational complexity. Furthermore, observing that we have $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{E}_m\boldsymbol{w} \to \sigma_{\mathrm{r}}^2, \forall m$ also in the high SNR regime ($\sigma_{\mathrm{r}}^2 \to 0$), the cost function in (5.128) is then reformulated as

$$\sum_{m=1}^{2K} \log(\bar{\boldsymbol{w}}^{\mathrm{H}} \boldsymbol{V}_{\mathrm{n}}^{\mathrm{H}} \boldsymbol{A}_m \boldsymbol{V}_{\mathrm{n}} \bar{\boldsymbol{w}} + \sigma_{\mathrm{u}}^2) - \sum_{m=1}^{2K} \log(\sigma_{\mathrm{r}}^2 + \sigma_{\mathrm{u}}^2) \tag{5.138}$$

Replacing the cost function in (5.128) by (5.138) and dropping the constant terms, we obtain the following problem

$$\begin{aligned} \max_{\bar{\boldsymbol{w}}} \quad & \sum_{m=1}^{2K} \log(\bar{\boldsymbol{w}}^{\mathrm{H}} \bar{\boldsymbol{A}}_m \bar{\boldsymbol{w}} + \sigma_{\mathrm{u}}^2) \\ \text{subject to} \quad & \bar{\boldsymbol{w}}^{\mathrm{H}} \bar{\boldsymbol{\Upsilon}}_i \bar{\boldsymbol{w}} \leq P_{\mathrm{R},i}, \forall i \end{aligned} \tag{5.139}$$

where $\bar{\boldsymbol{A}}_m = \boldsymbol{V}_{\mathrm{n}}^{\mathrm{H}} \boldsymbol{A}_m \boldsymbol{V}_{\mathrm{n}}, \forall m$ and $\bar{\boldsymbol{\Upsilon}}_i = \boldsymbol{V}_{\mathrm{n}}^{\mathrm{H}} \Gamma_{ii} \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{H}} \boldsymbol{V}_{\mathrm{n}}$. Again applying the SDR, we have the following convex SDP problem

$$\begin{aligned} \min_{\bar{\boldsymbol{X}}, \bar{\alpha}_m, \forall m} \quad & -\sum_{m=1}^{2K} \log(\bar{\alpha}_m) \\ \text{subject to} \quad & \mathrm{Tr}\{\bar{\boldsymbol{\Upsilon}}_i \bar{\boldsymbol{X}}\} \leq P_{\mathrm{R},i}, \forall i, \boldsymbol{X} \succeq 0 \\ & \mathrm{Tr}\{\bar{\boldsymbol{A}}_m \bar{\boldsymbol{X}}\} + \sigma_{\mathrm{u}}^2 = \bar{\alpha}_m. \end{aligned} \tag{5.140}$$

where $\bar{\boldsymbol{X}} = \bar{\boldsymbol{w}} \bar{\boldsymbol{w}}^{\mathrm{H}}$. After obtaining the optimal $\bar{\boldsymbol{X}}_{\mathrm{opt}}$, the rank-1 extraction of $\boldsymbol{V}_{\mathrm{n}} \bar{\boldsymbol{X}} \boldsymbol{V}_{\mathrm{n}}^{\mathrm{H}}$, which is computed using the randomization technique, yields the final $\boldsymbol{w}_{\mathrm{opt}}$.

5.5.3 Simulation results

In this section, the performance of the proposed algorithms is evaluated via Monte-Carlo simulations. The simulated flat fading channels are spatially uncorrelated Rayleigh fading channels. Each relay has an identical transmit power constraint of $1/N$. The noise variances at all nodes are the same, i.e., $\sigma_{\mathrm{r}}^2 = \sigma_{\mathrm{u}}^2$ and the SNR is defined as SNR $= 1/\sigma_m^2$. There are $K = 2$ pairs of users in the network. All the simulation results are obtained by averaging over 1000 channel realizations. "Polyblock", "POTDC", "ToT", and "IntNeu" denote the algorithms in Sections 5.5.2, 5.5.2, 5.5.2, and 5.5.2, respectively. For the polyblock algorithm, the POTDC algorithm, and the ToT algorithm, the stopping criterion is set to be a tolerance factor of 10^{-4}.

Figure 5.21 shows the comparison of different algorithms with $N = 6$ relays and $N = 12$ relays in the network. Clearly, the POTDC algorithm has close to optimal performance especially in the low SNR regime and when there is a sufficient number of relays in the network (e.g., $N = 12$). Thus, the POTDC algorithm can also be used as a benchmark for the other sub-optimal algorithms since it has a lower computational complexity but a comparable performance when compared to the global optimal solution.

Figure 5.22 demonstrates the comparison of different suboptimal algorithms. As depicted in the figure, the total SINR eigen-beamformer (ToT) and the interference neutralization based design (IntNeu) show a low SNR performance and a high SNR performance of the global optimum solution, respectively. Moreover, when there are enough relays in the network, both the total SINR eigen-beamformer and the interference neutralization based design are very close to the optimum solution but have much lower computational complexity.

To demonstrate the SAPHYRE gain, we compared to a baseline scenario where the spectrum and half number of relays (i.e., $N = 3$) are accessed by two operators in a TDMA fashion. Simulation results in Figure 5.23 have demonstrated a 1.8 times

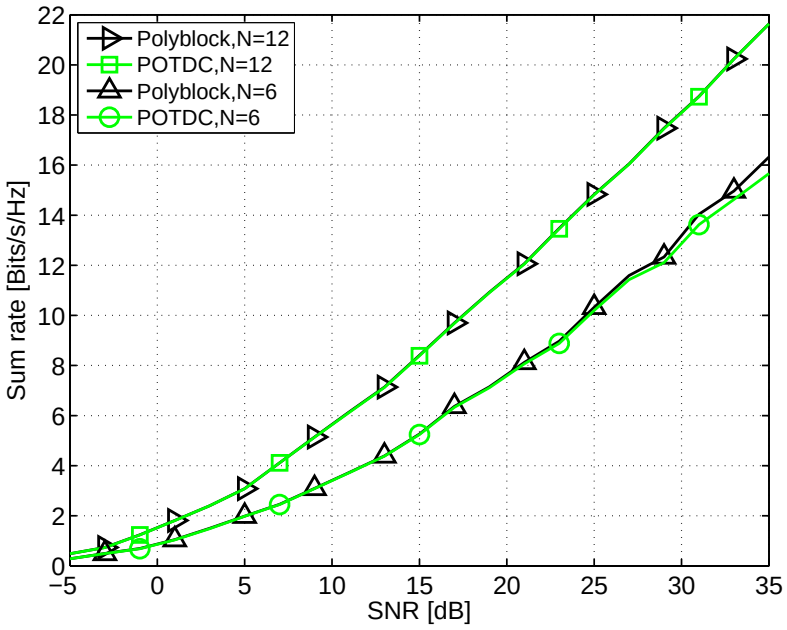


Figure 5.21: Sum rate comparison of the polyblock algorithm and the POTDC algorithm.

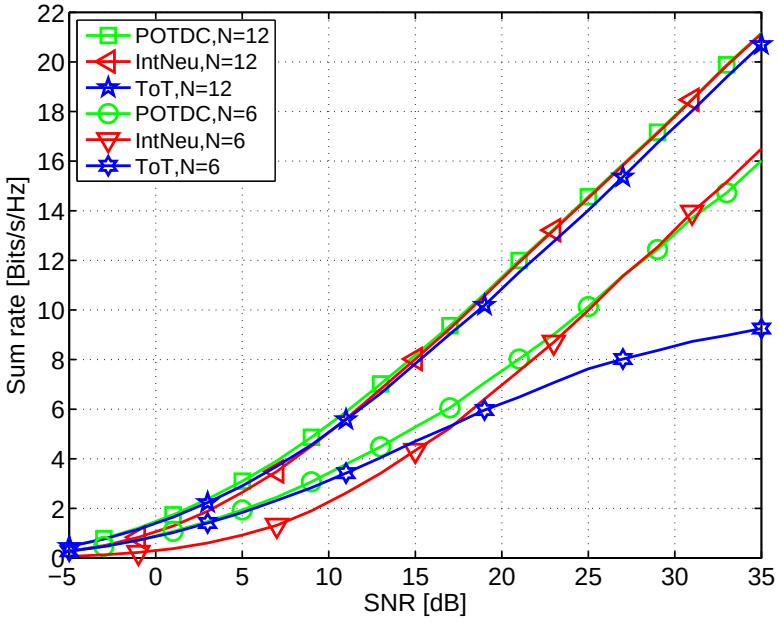
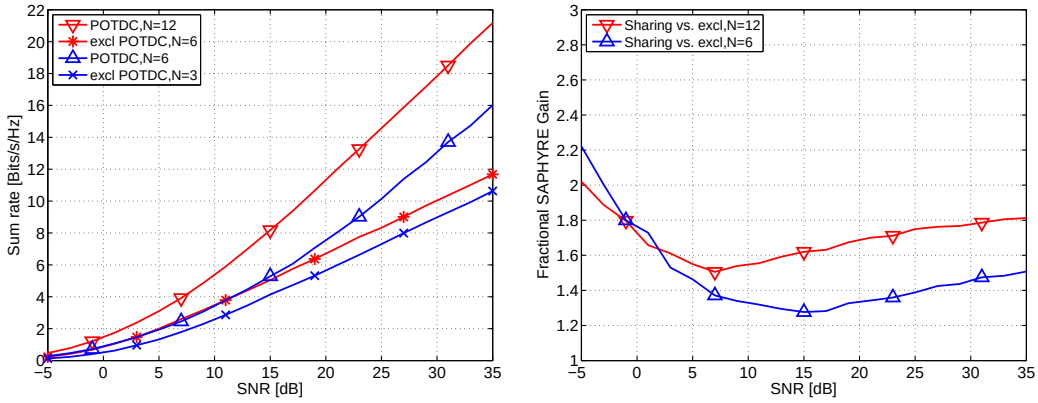


Figure 5.22: Sum rate comparison of the POTDC algorithm and the total SINR eigen-beamfomer (low SNR approximation) and the interference neutralization based design (high SNR approximation).



(a) Sharing vs. Non-sharing (b) Fractional SAPHYRE gain

Figure 5.23: Sum rate comparison of the SAPHYRE scenario and non-sharing scenario. The POTDC algorithm is used as the beamforming scheme.

SAPHYRE gain in the high SNR regime (e.g., 35 dB) and when there are enough number of relays (e.g., $N = 12$) in the network.

5.6 Network Coded Modulation in 2-Source 2-Relay Network

5.6.1 Introduction

Background and related work

Wireless Network Coding (WNC) (a.k.a. Physical Layer Network Coding (PLNC)) is the network structure aware PHY layer coding technique that currently attracts strong attention in the research community. The WNC technique utilizes the knowledge of the network structure directly at the PHY layer and in fact it overcomes the need of separate Medium Access Control or Network Routing Layer. The information is "flooded" through the whole network directly at PHY layer coding and processing. The WNC can be also seen as the step from the Network Coding (NC) principles [86] designed for discrete finite field alphabets over lossless links to the signal space representation typical for wireless communications.

Any particular WNC technique (WNC is the umbrella term) for multi-terminal (sources, destinations) and multiple-node (relays) consists of essentially three elements.

1. The coding/modulation strategy at sources which essentially defines joint signal space multi-source network aware codebook (Network Coded Modulation (NCM)),

2. the relaying processing strategy which defines the operations on relays,
3. the final destination decoding strategy which must correctly respect all available observables, some of them form the hierarchical information (contains some form of desired data), some of them form the Complementary Side Information (C-SI) (does not contain the desired data but *helps* decoding, i.e. friendly interference) and some form the classical interference.

See [87], [88] for the WNC technique overview. Selected results are available investigating various particular forms of the source, relay and destination strategies: lattice based [89], [90], [91] and other [92] network aware multi-source encoding, adaptive relay mapping/decision strategies [93], [94], [95], [96]. This work focuses on the Hierarchical Decode & Forward (HDF) relaying strategy [97], [98], [99], [100], [101]. The diversity and multiplexing performance of *discrete* NC in multi-source and multi-node scenario is treated in [102], [103]. Although this work deals with WNC at the physical layer, the results for discrete NC can serve as asymptotic bounds.

Goals and contributions

An outline of the work stands on following points.

1. We consider the NCM/HDF technique in the 2-source 2-relay scenario with SIMO (Single-Input Multiple-Output) fading channels.
2. The NCM/HDF technique is vulnerable to the *mutual* phase rotations of the signals from the sources [98], [99].
3. We take the advantage of additional degrees of freedom in SIMO channels and create *a specific* beam-forming tailored for given NCM/HDF strategy.
4. The resulting beam-forming is formed on a very different utility optimization than it is in the traditional beam-forming which is optimized for Gaussian alphabet interference channels.
5. The resulting Rx relay based beam-forming strategy is equalizing the signals at the relay to optimize processing of the *hierarchical* information to take the full advantage of HDF (confront that with classical interference nulling beam-forming).
6. The other relay which processes purely the C-SI uses zero-forcing interference nulling strategy.

We provide the following results and contributions.

1. We derive a close-form beam-forming solution for NCM/HDF scenario with asymmetric relay roles.
2. We quantify its performance by evaluating the hierarchical (mean and out-

age) achievable rates. These prove to benefit from the beam-forming. The resulting rates are approaching the single user cut-set bound and significantly outperform the classical sharing schemes (sum-rate cut-set bound).

5.6.2 System model and definitions

2-Source 2-relay network

We consider 2-Source 2-Relay (2S2R) communication network with sources SA and SB, relays R1 and R2, and a common destination node D (Figure 5.24). The relays operate under the half-duplex constraint. Our investigation goal focuses on the MAC (Multiple Access Channel) stage, i.e. SA/SB→R1/R2 communication. As a consequence, we assume that the second stage, the R1/R2→D links are lossless and orthogonal. The impact of the second stage will be addressed in the future work.

Sources SA/SB have a single antenna each. Relays R1/R2 are equipped with multiple N_R antennas and have available Channel State Information at Rx (CSIRx). Relays R1/R2 apply beam-forming to optimize the achievable rates. Data information rate of sources SA/SB are R_A, R_B. Throughout the document, $i \in \{1, 2\}$ denotes the relay index and $j \in \{A, B\}$ denotes the source index. $\mathbf{I}$ is the identity matrix.

The SA/SB transmit symbols are $s_A, s_B \in \mathcal{A}_s \subset \mathbb{C}$ with mean symbol energy $\mathrm{E}[|s_A|^2] = \mathrm{E}[|s_B|^2] = \mathrm{E}_s$. We generally omit the sequence numbers to keep the notation simple. Signal space channel symbols are mapped $s_A(c_A), s_B(c_B)$ on the (outer layer) code finite field symbols $c_A, c_B \in \mathbb{F}_M$. The signal space alphabet is indexed by symbols c_A, c_B.

The SIMO flat IID fading block-constant channel transfer vectors for all SA/SB→R1/R2 links are $\mathbf{h}_{1A}, \mathbf{h}_{1B}, \mathbf{h}_{2A}, \mathbf{h}_{2B} \in \mathbb{C}^{N_R \times 1}$ and have the unity variance per dimension $\mathrm{E}[\mathbf{h}_{ij}\mathbf{h}_{ij}^H] = \mathbf{I}$. We also denote $\mathbf{h} = \{\mathbf{h}_{1A}, \mathbf{h}_{1B}, \mathbf{h}_{2A}, \mathbf{h}_{2B}\}$. Relay R1/R2 received signals are

$$\mathbf{x}_1 = \mathbf{h}_{1A}s_A + \mathbf{h}_{1B}s_B + \mathbf{w}_1, \tag{5.141}$$

$$\mathbf{x}_2 = \mathbf{h}_{2A}s_A + \mathbf{h}_{2B}s_B + \mathbf{w}_2 \tag{5.142}$$

where $\mathbf{w}_1, \mathbf{w}_2$ is Gaussian complex rotationally invariant spatially white noise.

The relays apply beam-forming

$$y_1 = \mathbf{a}_1^H\mathbf{x}_1 = \mathbf{a}_1^H\mathbf{h}_{1A}s_A + \mathbf{a}_1^H\mathbf{h}_{1B}s_B + w_1, \tag{5.143}$$

$$y_2 = \mathbf{a}_2^H\mathbf{x}_2 = \mathbf{a}_2^H\mathbf{h}_{2A}s_A + \mathbf{a}_2^H\mathbf{h}_{2B}s_B + w_2 \tag{5.144}$$

where the beam-forming vectors $\mathbf{a}_i \in \mathbb{C}^{N_R \times 1}$ are unitary $\mathbf{a}_i^H\mathbf{a}_i = 1$ and Gaussian noise w_i has variance σ_w^2. The Signal-to-Noise Ratio (SNR) is defined

$$\gamma = \frac{\mathrm{E}[|\mathbf{a}_i^H\mathbf{h}_{ij}s_j|^2]}{\sigma_w^2} = \frac{\mathrm{E}_s}{\sigma_w^2} \tag{5.145}$$

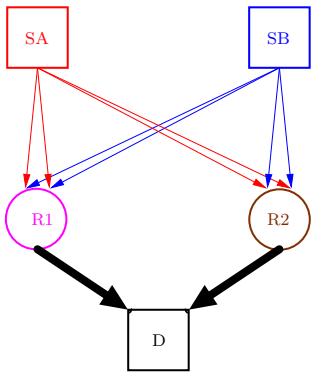


Figure 5.24: 2S2R network system model.

and it is independent of the channel dimensionality N_R including the special case of SISO system without BF (i.e. $\mathbf{a}_i = 1$).

NCM with HDF strategy for 2S2R network

Network Coded Modulation (NCM) is a particular way of implementing the WNC paradigm. It is a particular signal-space coding technique aware of the particular network structure and capable of delivering the information by "flooding" the source information through multiple wireless paths and relays to the final destination. All processing is purely done in PHY layer (i.e. there is no Medium Access nor Network Routing).

A particular flavor of NCM can be designed for the Hierarchical Decode & Forward (HDF) relaying strategy. Under HDF, each relay makes the decision on the *hierarchical* symbol. The hierarchical symbol must be defined at each relay in such a way that the final destination can decode all required sources from the available hierarchical and complementary side-information. This hierarchical symbol mapping is referred as a Hierarchical Network Code (HNC) $c_{AB} = \mathcal{X}(c_A, c_B)$. In this case, the HNC is applied *symbol-wise*. The NCM connects (or is designed for) the HNC with particular *signal-space* channel symbols $s_j \in \mathcal{A}_s = \{s^{(m)}\}_m$. The particular index mapping of the alphabet plays an important role [98]. It is important to stress that unlike for the traditional NC the inputs and the output of the HNC does *not* need to be from the same finite field $\mathbb{F}$ [101]. The hierarchical demodulator at the relay provides the hierarchical metric (see [97]) of the HNC code symbol

$$\mu(c_{AB}) = \mu\left(\cup_{c_A,c_B:\mathcal{X}(c_A,c_B)=c_{AB}}\{c_A, c_B\}\right). \tag{5.146}$$

Apart of the capability of reconstructing the source data, the NCM must also have channel coding capabilities. One of the possible solutions is the layered NCM design [98]. We will assume that the layered design is used and we concentrate only on the inner layer fulfilling the exclusive law (the destination reconstruction capability).

Throughout this work we will also restrict to the *minimal* cardinality mapping. The minimal mapping has *no* redundancy and the overall throughput is given by the weakest part and the rate region is *rectangular* [98]. Both c_{AB1}, c_{AB2} must be received at D without error. The impact of the relay to destination links is not considered here. The MAC stage is usually the performance bottleneck.

We now define two example NCM schemes that will be used later. These schemes are defined by a *particular specific choice* of the HNC matrix (see below). Generally, the HNC matrix (mapping) has many other possible options.

NCM#1 with binary symbols NCM with HDF strategy and binary channel alphabet can be constructed by layered approach with inner signal space layer having *linear* HNC c_{AB1}, c_{AB2} map at relays R1/R2

$$\mathbf{C}_{AB} = \mathbf{X}\mathbf{C} \tag{5.147}$$

where

$$\mathbf{C}_{AB} = \begin{bmatrix} c_{AB1} \\ c_{AB2} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} c_A \\ c_B \end{bmatrix}. \tag{5.148}$$

All operations with c code symbols are on $\mathbb{F}_2$ field. The HNC matrix $\mathbf{X}$ is invertible

$$\mathbf{X}^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{5.149}$$

and thus the destination D can reconstruct $\mathbf{C} = [c_A, c_B]^T$ from $\mathbf{C}_{AB} = [c_{AB1}, c_{AB2}]^T$. The signal-space alphabet $\mathcal{A}_s$ is BPSK.

Notice that the HNC matrix is asymmetric. The symbol c_{AB2} is a pure C-SI unlike the c_{AB1} which is the minimal cardinality hierarchical symbol. The asymmetry also opens an *additional* adaptation option of selecting a dual HNC matrix $\mathbf{X}' = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ where the pure C-SI is processed by R1. The C-SI can be also the symbol c_A as an option.

NCM#2 with $\mathbb{F}_{2^2}$ symbols The second example case uses HNC $\mathbb{F}_{2^2}$ matrix mapping

$$\mathbf{C}_{AB} = \mathbf{X}\mathbf{C} \tag{5.150}$$

where

$$\mathbf{C}_{AB} = \begin{bmatrix} c_{AB1,1} \\ c_{AB1,2} \\ c_{AB2,1} \\ c_{AB2,2} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} c_{A,1} \\ c_{A,2} \\ c_{B,1} \\ c_{B,2} \end{bmatrix}. \tag{5.151}$$

All operations with c code symbols are on $\mathbb{F}_{2^2}$ *extended* field, i.e. individual symbols are from the base field $c_{ABi,m}, X_{mn}, c_{j,n} \in \mathbb{F}_2$ and the vector $\mathbb{F}_{2^2}$ representations are $[c_{ABi,1}, c_{ABi,2}]^T$, $[c_{j,1}, c_{j,2}]^T$.

The HNC matrix $\mathbf{X}$ is invertible

$$\mathbf{X}^{-1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.152}$$

The signal-space alphabet $\mathcal{A}_s$ is QPSK with natural mapping. Relay R1 processes a bit-wise XOR hierarchical map and Relay R2 processes *pure* C-SI.

Reference case As a reference case, we evaluate the same NCM scheme but with SISO only channels and no beam-forming. This reference case allows us to evaluate the gains coming from additional degrees of freedom utilized by SIMO with beam-forming.

5.6.3 Beamforming optimization parametrization

Problem statement

The binary NCM case simply reveals that the HNC is asymmetric. Relay R1 processes the minimal cardinality hierarchical symbol and the results of [98] show that the performance is given by the worse of the links SA-R1 and SB-R1. The optimal is the equal received Signal to Noise Ratio (SNR). On the other side Relay R2 processes the pure C-SI from SB. Any signal from SA is the interference. Our goal is to optimize beam-forming vectors to maximize the throughput and compare it to the reference case (SISO no beam-forming). However, it should be stressed that our problem setup does not lead to the classical MMSE receive beam-former which minimizes the MSE and maximizes the SINR.

A formal goal statement is

$$\max_{\mathbf{a}=[\mathbf{a}_A,\mathbf{a}_B]:\, \|\mathbf{a}_1\|=1, \|\mathbf{a}_2\|=1} \min\left(I(c_{AB1}; y_1|\mathbf{a}), I(c_{AB2}; y_2|\mathbf{a})\right) \tag{5.153}$$

for each particular channel realization $\mathbf{h}$. A direct BF optimization for (5.153) is difficult since the mutual *finite alphabet* mutual information $I(c_{ABi}; y_i|\mathbf{a})$ does not have a closed-form solution and was obtained only numerically [98]. It also strongly depends on the alphabet $\{s^{(m)}\}_m$ and channel *parameterization*.

Optimization subspace

The utility functions (5.153) contain the beam-forming vectors only through the inner-product projections (these are the only observables at the relays)

$$\mathbf{a}_1^H\mathbf{h}_{1A},\ \mathbf{a}_1^H\mathbf{h}_{1B},\ \mathbf{a}_2^H\mathbf{h}_{2A},\ \mathbf{a}_2^H\mathbf{h}_{2B}. \tag{5.154}$$

The optimal BF vector $\hat{\mathbf{a}}_1$ must lie in the range space $\mathcal{R}(\mathbf{h}_{1A}, \mathbf{h}_{1B})$ generated by $\mathbf{h}_{1A}, \mathbf{h}_{1B}$. For any vector from the null space $\bar{\mathbf{a}}_1 \in \mathcal{N}(\mathbf{h}_{1A}, \mathbf{h}_{1B})$, it must be $\bar{\mathbf{a}}_1^H\mathbf{h}_{1A} = 0$, $\bar{\mathbf{a}}_1^H\mathbf{h}_{1B} = 0$. We denote the orthonormal basis of this range space $\mathbf{g}_{1A}, \mathbf{g}_{1B}$. It can be obtained from particular $\mathbf{h}_{1A}, \mathbf{h}_{1B}$ e.g. by Gram-Schmidt orthonormalization procedure or alternatively [104] by using the projection matrix $\mathbf{P}_{1A}$ on the space spanned by $\mathbf{h}_{1A}$ and the corresponding null space projection $\bar{\mathbf{P}}_{1A}$ matrix

$$\mathbf{P}_{1A} = \frac{\mathbf{h}_{1A}\mathbf{h}_{1A}^H}{\|\mathbf{h}_{1A}\|^2},\ \ \bar{\mathbf{P}}_{1A} = \mathbf{I} - \mathbf{P}_{1A} \tag{5.155}$$

and similarly for $\mathbf{P}_{2A}, \bar{\mathbf{P}}_{2A}$.

The basis is

$$\mathbf{g}_{1A} = \frac{\mathbf{P}_{1A}\mathbf{h}_{1B}}{\|\mathbf{P}_{1A}\mathbf{h}_{1B}\|},\ \mathbf{g}_{1B} = \frac{\bar{\mathbf{P}}_{1A}\mathbf{h}_{1B}}{\|\bar{\mathbf{P}}_{1A}\mathbf{h}_{1B}\|}. \tag{5.156}$$

The situation for relay R2 is analogous

$$\mathbf{g}_{2A} = \frac{\mathbf{P}_{2A}\mathbf{h}_{2B}}{\|\mathbf{P}_{2A}\mathbf{h}_{2B}\|},\ \mathbf{g}_{2B} = \frac{\bar{\mathbf{P}}_{2A}\mathbf{h}_{2B}}{\|\bar{\mathbf{P}}_{2A}\mathbf{h}_{2B}\|}. \tag{5.157}$$

Then $\hat{\mathbf{a}}_1 \in \mathcal{R}(\mathbf{g}_{1A}, \mathbf{g}_{1B})$, $\hat{\mathbf{a}}_2 \in \mathcal{R}(\mathbf{g}_{2A}, \mathbf{g}_{2B})$ and the parametric representation of all possible solution candidates is

$$\hat{\mathbf{a}}_1(\alpha_{1A}, \alpha_{1B}) = \alpha_{1A}\mathbf{g}_{1A} + \alpha_{1B}\mathbf{g}_{1B} \tag{5.158}$$

$$\hat{\mathbf{a}}_2(\alpha_{2A}, \alpha_{2B}) = \alpha_{2A}\mathbf{g}_{2A} + \alpha_{2B}\mathbf{g}_{2B} \tag{5.159}$$

where $\alpha_{1A}, \alpha_{1B}, \alpha_{2A}, \alpha_{2B} \in \mathbb{C}$. The normalization conditions $\|\hat{\mathbf{a}}_1\| = \|\hat{\mathbf{a}}_2\| = 1$ imply $|\alpha_{iA}|^2 + |\alpha_{iB}|^2 = 1$.

Equivalent channel gains

The inner products $\mathbf{a}_1^H\mathbf{h}_{1A}$, $\mathbf{a}_1^H\mathbf{h}_{1B}$, $\mathbf{a}_2^H\mathbf{h}_{2A}$, $\mathbf{a}_2^H\mathbf{h}_{2B}$ represent equivalent channel gains. We now examine their properties when the BF vectors are parametrized according to (5.158) and (5.159). The individual component inner products that form the channel gains for Ri are

$$\eta_{iAA} = \mathbf{g}_{iA}^H\mathbf{h}_{iA} \in \mathbb{C}, \tag{5.160}$$

$$\eta_{iBA} = \mathbf{g}_{iB}^H \mathbf{h}_{iA} = \frac{\mathbf{h}_{iB}^H \bar{\mathbf{P}}_{iA}}{\|\bar{\mathbf{P}}_{iA}\mathbf{h}_{iB}\|}\mathbf{h}_{iA} = 0, \tag{5.161}$$

$$\eta_{iAB} = \mathbf{g}_{iA}^H \mathbf{h}_{iB} = \frac{\mathbf{h}_{iB}^H \mathbf{P}_{iA}\mathbf{h}_{iB}}{\|\mathbf{P}_{iA}\mathbf{h}_{iB}\|} = \mathbf{h}_{iB}^H \mathbf{g}_{iA} \in \mathbb{R}, \tag{5.162}$$

$$\eta_{iBB} = \mathbf{g}_{iB}^H \mathbf{h}_{iB} = \frac{\mathbf{h}_{iB}^H \bar{\mathbf{P}}_{iA}\mathbf{h}_{iB}}{\|\bar{\mathbf{P}}_{iA}\mathbf{h}_{iB}\|} = \mathbf{h}_{iB}^H \mathbf{g}_{iB} \in \mathbb{R}. \tag{5.163}$$

All values $\eta_{iAA}, \eta_{iBA}, \eta_{iAB}, \eta_{iBB}$ are *exclusively* given by the channel vector realizations $\mathbf{h}_{iA}, \mathbf{h}_{iB}$.

The equivalent channel gains are

$$\tilde{h}_{iA} = \mathbf{a}_i^H \mathbf{h}_{iA} = \alpha_{iA}^* \eta_{iAA} \tag{5.164}$$

$$\tilde{h}_{iB} = \mathbf{a}_i^H \mathbf{h}_{iB} = \alpha_{iA}^* \eta_{iAB} + \alpha_{iB}^* \eta_{iBB}. \tag{5.165}$$

The relay observables expressed in terms of the BF parametrization are

$$y_i = \tilde{h}_{iA} s_A + \tilde{h}_{iB} s_B + w_i. \tag{5.166}$$

BF parametrization for NCM

As shown in [98], the hierarchical mutual information $I(c_{ABi}; y_i|\mathbf{a})$ depends, apart of the overall power levels, on the *equivalent relative channel gain*

$$\tilde{h}_i = \frac{\tilde{h}_{iB}}{\tilde{h}_{iA}} = \frac{\eta_{iAB}}{\eta_{iAA}} + \frac{\alpha_{iB}^*}{\alpha_{iA}^*}\frac{\eta_{iBB}}{\eta_{iAA}}, \tag{5.167}$$

in the equivalent channel

$$y_i = \tilde{h}_{iA}\left(s_A + \tilde{h}_i s_B\right) + w_i. \tag{5.168}$$

As we see in (5.167), the relative gain has *only two degrees of freedom* constrained by $|\alpha_{iA}|^2 + |\alpha_{iB}|^2 = 1$.

5.6.4 NCM with optimized Beamforming

NCM BF optimization

Both example schemes NCM#1/#2 have hierarchical information at R1 and pure C-SI at R2. The *BF optimization strategy* is the following.

1. The *relative* channel gain $\tilde{h}_1 = \tilde{h}_{1B}/\tilde{h}_{1A}$ has the major impact on the hierarchical rate [98]. The relay R1 will try to equalize the level of signals and balance their *relative* phase to maximize the hierarchical mutual information. The hierarchical rate is limited by the weaker link. The hierarchical constellation is typically (see [98]) constructed in such a way that the optimal relative angle is zero and the received signals have equal power, i.e. $\tilde{h}_1 = 1$. The optimization goal at R1 is

$$[\alpha_{1A}, \alpha_{1B}] = \arg\max |\tilde{h}_{1A}|^2 \tag{5.169}$$
$$\text{s.t.} |\alpha_{1A}|^2 + |\alpha_{1B}|^2 = 1 \tag{5.170}$$
$$\tilde{h}_1 = 1. \tag{5.171}$$

2. The relay R2 processes pure C-SI from SB and the signal from SA is a pure interference. The strategy at R2 is the zero-forcing (interference nulling) with maximization of the SB signal level

$$[\alpha_{2A}, \alpha_{2B}] = \arg\max |\tilde{h}_{2B}|^2 \tag{5.172}$$
$$\text{s.t.} |\alpha_{2A}|^2 + |\alpha_{2B}|^2 = 1 \tag{5.173}$$
$$\tilde{h}_{2A} = 0. \tag{5.174}$$

The strategy for R1 is solved by finding α_{1A} with maximal $|\alpha_{1A}|$ s.t. $|\alpha_{1A}|^2 + |\alpha_{1B}|^2 = 1$ and

$$\frac{\alpha_{1B}^*}{\alpha_{1A}^*} = \frac{\eta_{1AA} - \eta_{1AB}}{\eta_{1BB}}. \tag{5.175}$$

This gives

$$|\alpha_{1A}|^2 \left(1 + \left|\frac{\eta_{1AA} - \eta_{1AB}}{\eta_{1BB}}\right|^2\right) = 1. \tag{5.176}$$

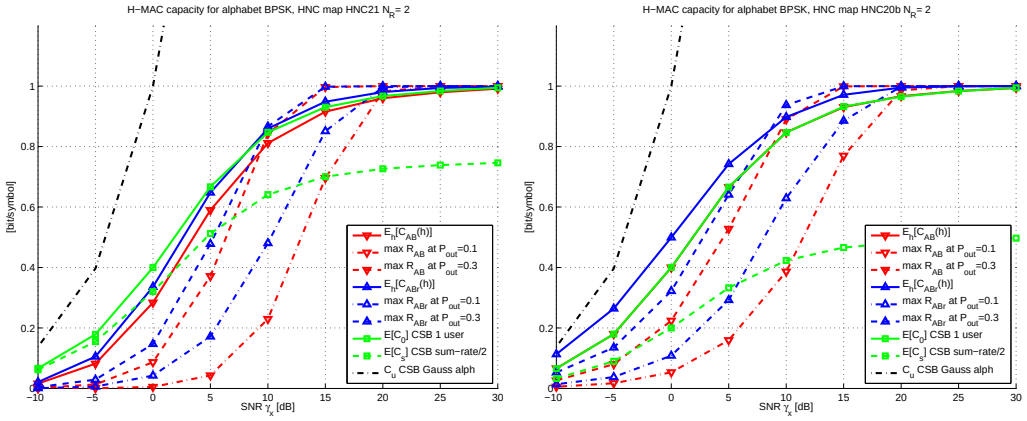
The angle of α_{1A} cannot affect the solution. Therefore we can constrain ourselves to $\alpha_{1A} \in \mathbb{R}$ and hence the solution is

$$\alpha_{1A} = \frac{1}{\sqrt{1 + \left|\frac{\eta_{1AA} - \eta_{1AB}}{\eta_{1BB}}\right|^2}} \tag{5.177}$$

and

$$\alpha_{1B} = \alpha_{1A} \left(\frac{\eta_{1AA} - \eta_{1AB}}{\eta_{1BB}}\right)^*. \tag{5.178}$$

The strategy for R2 is solved by setting $\alpha_{2A} = 0$ and then finding α_{2B} maximizing $|\alpha_{2B}^* \eta_{2BB}|$ such that $|\alpha_{2B}|^2 = 1$. An obvious solution is $\alpha_{2B} = 1$.



(a) (b)

Figure 5.25: Hierarchical rates C_{AB1}, C_{AB2} for NCM#1 with BPSK alphabet, $N_R = 2$ and reference case C_{AB1r}, C_{AB2r} at relay (a) R1 and (b) R2.

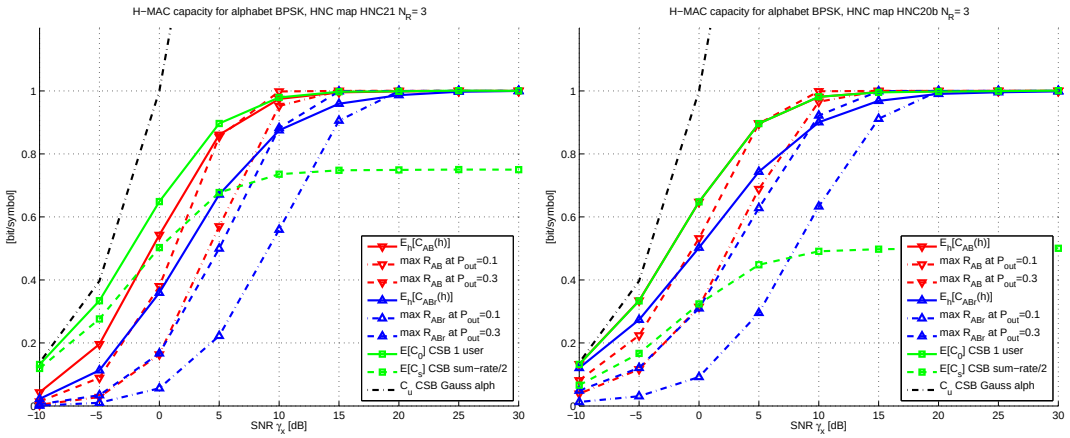
5.6.5 Numerical results

We numerically evaluate the hierarchical achievable finite uniform alphabet rates for block-fading Rayleigh channel with unity channel gain variance per dimension, Figs. 5.25, 5.26, 5.27, 5.28. The numerical results were achieved by the Monte-Carlo evaluation of the mutual information (see details in [98]). We evaluate the rates $C_{AB1} = I(c_{AB1}; y_1)$ and $C_{AB2} = I(c_{AB2}; y_2)$ with beam-forming applied for each individual channel realization. These rates depend on the channel state. We evaluate their mean values over all channel states $\mathrm{E}[C_{ABi}(\mathbf{h})]$. The outage capacity $P_{\text{out}} = \Pr\{C_{ABi}(\mathbf{h}) < R_{ABi}\}$ is evaluated by plotting the maximum achievable rate R_{ABi} at the given probability of the outage P_{out}. For a comparison, we evaluate the reference SISO system with no beam-forming and show again the mean rates and maximum rate for the given outage. Additionally we also show mean Cut-Set Bounds (CSBs) for finite alphabet and also the unconstrained Gaussian single user capacity evaluated for the mean SNR. Both CSBs are plotted only for the channel with the beam-forming applied to keep the figures simple.

5.6.6 Discussion and conclusions

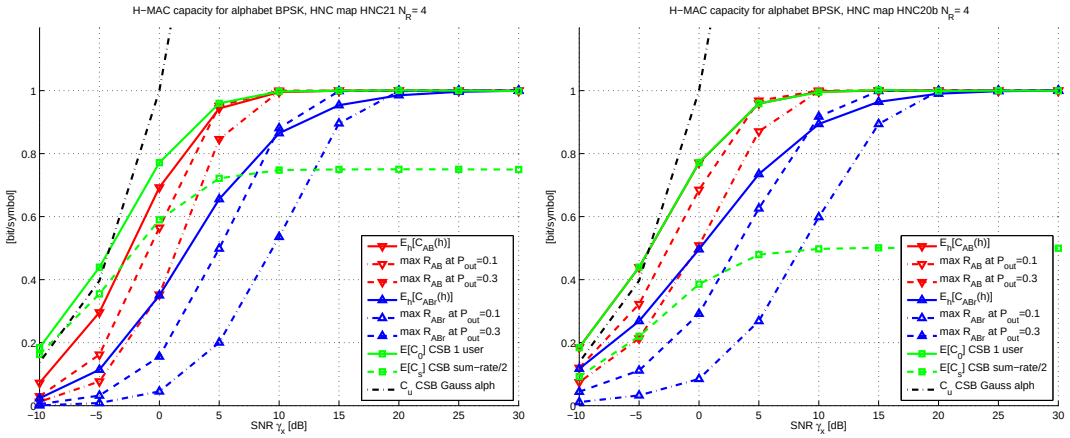
We have developed a closed form solution of the NCM/HDF specific beam-forming and applied it on example HNC mappings with BPSK and QPSK signal space alphabets. The algorithm is based on a proper beam-forming subspace parametrization and subsequent solution for the parameters.

The numerical results show that the NCM/HDF specific beam-former generally helps the performance in terms of the hierarchical rates, both, the mean values and the outage rates. This behavior does not fully apply only in the case of binary



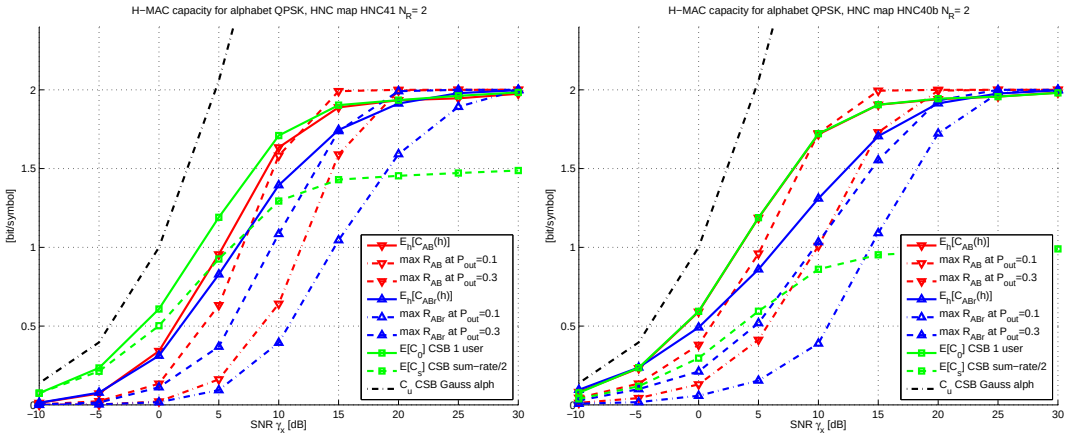
(a) (b)

Figure 5.26: Hierarchical rates C_{AB1}, C_{AB2} for NCM#1 with BPSK alphabet, $N_R = 3$ and reference case C_{AB1r}, C_{AB2r} at relay (a) R1 and (b) R2.



(a) (b)

Figure 5.27: Hierarchical rates C_{AB1}, C_{AB2} for NCM#1 with BPSK alphabet, $N_R = 4$ and reference case C_{AB1r}, C_{AB2r} at relay (a) R1 and (b) R2.



(a) (b)

Figure 5.28: Hierarchical rates C_{AB1}, C_{AB2} for NCM#2 with QPSK alphabet, $N_R = 2$ and reference case C_{AB1r}, C_{AB2r} at relay (a) R1 and (b) R2.

alphabet and $N_R = 2$ where the beam-forming apparently does not have enough degrees of freedom to match the relative channel w.r.t. hierarchical symbols. This phenomenon requires further investigation. In all other cases, including QPSK alphabet with $N_R = 2$, we observe a significant improvement in the throughput. For medium to high SNR, the hierarchical rate is much higher than the sum-rate per user CSB which corresponds to the classical (non WNC) channel sharing techniques.

6 Conclusions

This report describes novel signal processing techniques to handle the new interference that is created in the sharing scenarios of SAPHYRE. We have presented the sharing scenarios that are investigated in SAPHYRE and have introduced the new approaches that have been developed in the SAPHYRE project. Moreover, we have demonstrated the SAPHYRE gain, which is defined as the comparison between the simultaneous sharing case and the time-shared case of the spectrum and infrastructure by the operators. The main results are summarized as follows:

- In Chapter 3 interference mitigation techniques in spectrum sharing multi-cell wireless networks are studied.
  - In Section 3.1 the performance of joint precoding across cooperating operators under a varying backhaul data symbol routing overhead is studied. Simulation results show that the proposed routing algorithm outperforms the conventional non-sharing schemes for practical backhaul overhead values. For example, the SAPHYRE gain is 140% when the backhaul overhead is 50% of the full cooperation scenario. As the allowed backhaul overhead increases, the SAPHYRE gain also increases.
  - In Section 3.2 a channel estimation method which provides a substantial improvement in performance is proposed. As compared to the non-sharing case, a considerable SAPHYRE gain is obtained when a coordinated pilot assignment method is applied within the channel estimation phase (for instance, the gain is approximately 14% when the number of antennas at each BS is equal to 10). Moreover, the SAPHYRE gain increases together with the number of the antennas at each operator.
  - Section 3.3 considers the downlink of a multicell network where neighboring multi-antenna base stations share the spectrum and coordinate their frequency and spatial resource allocation strategies to improve the overall network performance. A branch & bound algorithm is proposed to maximize the number of users that can be scheduled, meeting their quality-of-service requirements with the minimum total transmit power. Compared to the non-sharing scenario, a significant SAPHYRE gain in terms of sum rate is obtained in the non-cositing network. In the cositing network, the SAPHYRE gain is relatively small.
- In Chapter 4 distributed MIMO signal processing and resource allocation algorithms for spectrum sharing multi-cell wireless networks are developed.
  - In Section 4.1 a distributed beamforming algorithm is proposed for the

two-user multiple-input single-output (MISO) interference channel (IC). The proposed algorithm is iterative and in every iteration each transmitter computes its beamforming vector distributedly. The outcome of the proposed algorithm is approximately Pareto-optimal and it is in all cases better than the Nash equilibrium (NE), which would be the outcome if there was no cooperation. The novel element of the proposed algorithm is the use of the generated interference level as bargaining value. The algorithm is equally applicable to the case of instantaneous and statistical CSI. The SAPHYRE gain increases with increasing SNR for the scenarios of instantaneous CSI and statistical CSI with low-rank matrices.

  - In Section 4.2 the problem of adaptive allocation of spectrum, power, and rate is considered for the downlink of multicell orthogonal frequency-division multiple-access (OFDMA) networks. A low-complexity distributed iterative algorithm is proposed and is compared to the solution of the individual mixed integer programs (MIP) and the iterative waterfilling (IWF) algorithm. It has been demonstrated that the proposed algorithm outperforms the sequential solution of the individual MIPs and the IWF algorithm. The performance gap between the proposed algorithm and IWF becomes larger as the number of subcarriers increases.

- In Chapter 5 we investigate the relay-aided spectrum and infrastructure sharing scenarios.
  - The work in Section 5.1 focuses on a multi-user wireless network which consists of a multiple-antenna smart instantaneous relay (i.e., the signals of the source-destination link arrive at the same time as the source-relay-destination link) and dumb relays, where the smart relay is able to gather channel state information, perform linear processing, and forward the signals whereas the dumb relays are only able to serve as amplifiers. The smart relay utilizes the gathered channel state information for interference management and signal enhancement purposes. The achievable rate region is studied and two algorithms based on interference neutralization are developed. A significant SAPHYRE gain is obtained when (i) the number of antennas at the relay is larger than the number of users in the system and (ii) the relay has enough transmit power to neutralize interference.
  - In Section 5.2 the users' privacy in heterogeneous dense networks, where spectrum is shared, is investigated. On a multi-antenna relay-assisted multi-carrier interference channel, each user shares the frequency and spatial resources with all other users. When the receivers are not only interested in their own signals but also in eavesdropping other users' signals, the cross talk on the frequency and spatial channels becomes information leakage. Thereby, a novel secrecy rate enhancing relay strategy is proposed, which utilizes both frequency and spatial resources, termed

as *information leakage neutralization*. The proposed relay strategy shows a significant improvement in terms of sum secrecy rate compared to a non-sharing baseline scenario especially in the high SNR regime. A significant SAPHYRE gain can be obtained when (i) the number of transmit antennas at the relay is larger than the number of users in the system if every user transmits with full multiplexing gains (less number of antennas at the relay are required if some users transmit less streams); (ii) the amount of transmit power at the relay is enough for information leakage neutralization.

- In Section 5.3 an energy efficient transmission is developed for the multiple operator AF MIMO relay sharing system. A significant sharing gain in terms of power savings is demonstrated in the simulation results. The sharing gain as a function of the allowable transmit power at the BSs is almost a constant as the transmit power increases.
- In Section 5.4 we study a sharing scenario where both the BS and a two-way AF MIMO relay are shared among multiple operators. An efficient signal processing algorithm which is based on the projection based separation of multiple operators (ProBaSeMO) method is proposed to accomplish this form of resource sharing. Compared to the non-sharing case, a significant SAPHYRE gain in terms of sum rate is obtained in the high SNR regime and when there are sufficient number of antennas at the relay.
- In Section 5.5 we investigate a relay sharing scenario where the spectrum and multiple two-way AF relays are shared among different operators. Compared to the non-sharing scenario, a significant sharing gain in terms of sum rate is achievable only in the high SNR regime and when there are many relays in the network.
- In Section 5.6 we have developed a closed form solution of the network coded modulation with hierarchical decode and forward relaying strategy (NCM/HDF) specific beam-forming and applied it on example hierarchical network code (HNC) mappings with BPSK and QPSK signal space alphabets. The algorithm is based on a proper beam-forming subspace parametrization and subsequent solution for the parameters. The numerical results show that for medium to high SNR the hierarchical rate is much higher than the sum-rate per user cut-set bound (CSB) which corresponds to the classical (non wireless network coding (WNC)) channel sharing techniques.

This deliverable has summarized various practical approaches to deal with the new interference in the sharing scenarios. Some of these approaches are further evaluated via the system level simulation in WP4 and via a demonstrator implementation in WP6.

Bibliography

[1] T. S. Rappaport, *Wireless Communications*. Prentice Hall, 1996.

[2] K. H. Teo, Z. Tao, and J. Zhang, "The mobile broadband WiMAX standard [standards in a nutshell]," *IEEE Signal Processing Magazine*, Sep. 2007.

[3] M. Bennis and J. Lilleberg, "Inter-Operator Resource Sharing for 3G Systems and Beyond," in *IEEE Ninth International Symposium on Spread Spectrum Techniques and Applications*, 2006.

[4] V. Heinonen, P. Pirinen, and J. Iinatti, "Capacity gains through inter-operator resource sharing in a cellular network," in *Proc. IEEE WPMC*, 2008.

[5] G. technical specification, "Evolved universal terrestrial radio access (e-utra): Physical layer procedures," 3GPP, Tech. Rep. TS 36.213, 2011.

[6] G. Document, "Reply ls to r3-070527/r1-071242 on backhaul (x2 interface) delay," 3GPP, Tech. Rep. R1-071804, 2011.

[7] A. Magee, "Synchronization in next-generation mobile backhaul networks," *IEEE Communications Magazine*, vol. 48, no. 10, pp. 110–116, Oct. 2012.

[8] K. Hann, S. Jobert, and S. Rodrigues, "Synchronous ethernet to transport frequency and phase/time," *IEEE Communications Magazine*, vol. 50, no. 8, pp. 152–160, Aug. 2012.

[9] V. Jungnickel, T. Wirth, M. Schellmann, T. Haustein, and W. Zirwas, "Synchronization of cooperative base stations," in *Proc. IEEE International Symposium on Wireless Communication Systems 2008 (ISWCS 08)*, Oct. 2008.

[10] "IEEE standard for a precision clock synchronization protocol for networked measurement and control systems," IEEE, Tech. Rep. IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002), 2008.

[11] G. T. Report, "Evolved universal terrestrial radio access (E-UTRA): Further advancements for E-UTRA physical layer aspects," 3GPP, Tech. Rep. TR 36.814 v9.0.0, 2010.

[12] NGMN, http://www.ngmn.org/workprogramme/centralisedran.html?type=98.

[13] R. Gangula, P. de Kerret, D. Gesbert, and M. Al-odeh, "Optimized Data Symbol Allocation in Multicell MIMO Channels," in *proc Asilomar*, 2011.

[14] A. Gjendemsj, D. Gesbert, G. Oien, and S. Kiani, "Binary Power Control for Sum Rate Maximization over Multiple Interfering Links," *IEEE Trans. on*

Wireless Comm., vol. 7, no. 8, pp. 3164 –3173, August 2008.

[15] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10 –22, Feb. 1998.

[16] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[17] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE International Symposium on Information Theory (ISITo09)*, Seoul, Korea, Jun. 2009, pp. 2184–2188.

[18] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSPo11)*, Prague, Czech Republic, May 2011, pp. 3464–3467.

[19] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2011, pp. 381–385.

[20] A. Scherb and K. Kammeyer, "Bayesian channel estimation for doubly correlated MIMO systems," in *Proc. IEEE Workshop Smart Antennas*, 2007.

[21] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[22] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *to appear in IEEE Journal on Selected Areas in Communications, Special Issue on Large Scale Antenna Systems*, 2013.

[23] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *Proc. 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2011, pp. 545 –550.

[24] J. A. Tsai, R. M. Buehrer, and B. D. Woerner, "The impact of AOA energy distribution on the spatial fading correlation of linear antenna array," in *Proc. IEEE Vehicular Technology Conference, (VTCo 02)*, vol. 2, May 2002, pp. 933–937.

[25] J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information," in *46th Annual Conference on Information Sciences and Systems, (CISS 2012)*, Princeton University, NJ, USA, Mar. 2012.

[26] E. Matskani, N. D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Convex approximation techniques for joint multiuser downlink beamforming and admission control," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2682–2693, Jul. 2008.

[27] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communication*, L. C. Godara, Ed. Boca Raton, FL: CRC Press, ch. 18, 2001.

[28] M. Butussi and M. Bengtsson, "Low complexity admission in downlink beamforming," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, Toulouse, France, May 2006, pp. 261–264.

[29] K. Börner, J. Dommel, S. Jaeckel, and L. Thiele, "On the requirements for quasi-deterministic radio channel models for heterogeneous networks," in *Proc. Int. Symp. on Signals, Systems and Electronics (ISSSE)*, Potsdam, Germany, Oct. 2012.

[30] J. Meinilä, P. Kyösti, L. Hentilä, T. Jämsä, E. Suikkanen, E. Kunnari, and M. Narandzic, "D5.3: WINNER+ final channel models," CELTIC CP5-026 WINNER+, Jun. 2010.

[31] J. Lindblom, E. G. Larsson, and E. A. Jorswieck, "Parametrization of the MISO IFC rate region: The case of partial channel state information," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 500–504, Feb. 2010.

[32] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, pp. 5292–5296, Oct. 2008.

[33] J. Lindblom, E. Karipidis, and E. G. Larsson, "Selfishness and altruism on the MISO interference channel: The case of partial transmitter CSI," *IEEE Commun. Lett.*, vol. 13, pp. 667–669, Sep. 2009.

[34] E. G. Larsson, E. A. Jorswieck, J. Lindblom, and R. Mochaourab, "Game theory and the flat fading gaussian interference channel," *IEEE Signal Process. Mag.*, vol. 26, pp. 18–27, Sep. 2009.

[35] E. Karipidis, A. Gründinger, J. Lindblom, and E. G. Larsson, "Pareto-optimal beamforming for the MISO interference channel with partial CSI," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Proces. (CAMSAP)*, Aruba, Dec. 2009, pp. 13–16.

[36] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, pp. 664–678, Feb. 2010.

[37] H. M. Salkin and K. Mathur, *Foundation of Integer Programming.* New York: North-Holland, 1989.

[38] I. Kim, I.-S. Park, and Y. Lee, "Use of linear programming for dynamic sub-

carrier and bit allocation in multiuser OFDM," *IEEE Trans. Veh. Technol.*, vol. 55, no. 4, pp. 1195–1207, Jul. 2006.

[39] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[40] D. Berteskas, *Nonlinear Programming*. Boston: Athena Scientific, 1999.

[41] "GNU linear programming kit," http://www.gnu.org/software/glpk, version 4.45.

[42] G. Scutari, D. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory–part II: Algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1250–1267, Mar. 2008.

[43] W. Yu, G. Ginis, and J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.

[44] M. Fathi and H. Taheri, "Utility-based resource allocation in orthogonal frequency division multiple access networks," *IET Commun.*, vol. 4, no. 12, pp. 1463–1470, Aug. 2010.

[45] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks–part II: Algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[46] A. El Gamal and N. Hassanpour, "Relay without Delay," in *Proceedings of International Symposium on Information Theory*, vol. 1, 2005, pp. 1078–1080. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1523505

[47] N. Lee and S. A. Jafar, "Aligned Interference Neutralization and the Degrees of Freedom of the 2 User Interference Channel with Instantaneous Relay," *submitted to IEEE Transaction of Information Theory, available at http://arxiv.org/abs/1102.3833*, pp. 1–17, 2011. [Online]. Available: http://arxiv.org/abs/1102.3833

[48] H. Chang and S.-Y. Chung, "Capacity of Strong and Very Strong Gaussian Interference Relay-without-delay Channels," *preprint, available at http://arXiv:1108.2846v1*, pp. 1–19, 2011.

[49] V. R. Cadambe, S. A. Jafar, and C. Wang, "Interference Alignment with Asymmetric Complex Signaling - Settling the Host-Madsen-Nosratinia Conjecture," *IEEE Transaction on Information Theory*, vol. 56, no. 9, pp. 4552–4565, 2010.

[50] Z. K. M. Ho and E. Jorswieck, "Improper Gaussian Signaling On The Two-

User SISO Interference Channel," *IEEE Transaction of Wireless Communications, accepted*, pp. 1–22, 2011.

[51] H. Lutkepohl, *Handbook of Matrices*. John Wiley & Sons, Ltd, 1996.

[52] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete Characterization of the Pareto Boundary for the MISO Interference Channel," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5292–5296, 2008.

[53] Z. K. M. Ho, D. Gesbert, E. Jorswieck, and R. Mochaourab, "Beamforming on the MISO interference channel with multi-user decoding capability," *preprint, available at http://arxiv.org/abs/1107.0416*, pp. 1–6, 2010. [Online]. Available: http://arxiv.org/abs/1107.0416

[54] M. Charafeddine, A. Sezgin, and A. Paulraj, "Rate Region Frontiers For n-User Interference Channel With Interference As Noise," in *Proceedings of IEEE 45th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2007.

[55] R. Zhang and S. Cui, "Cooperative Interference Management with MISO Beamforming," *IEEE Transaction on Signal Processing*, vol. 58, no. 10, pp. 5450 – 5458, 2010.

[56] M. Grant and S. Boyd, "{CVX}: Matlab Software for Disciplined Convex Programming, version 1.21," 2011. [Online]. Available: http://cvxr.com/cvx

[57] Y. Huang and D. P. Palomar, "Rank-Constrained Separable Semidefinite Programming With Applications to Optimal Beamforming," *IEEE Transaction on Signal Processing*, vol. 58, no. 2, pp. 664–678, 2010.

[58] W. Ai, Y. Huang, and S. Zhang, "New Results on Hermitian Matrix Rank-one Decomposition," *Mathematical Programming*, vol. 128, no. 1-2, pp. 253–283, Aug. 2009. [Online]. Available: http://www.springerlink.com/index/10.1007/s10107-009-0304-7

[59] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, pp. 20–34, 2010.

[60] Z. K. M. Ho and E. Jorswieck, "Instantaneous Relaying : Optimal Strategies and Interference Neutralization," *accepted, IEEE Transaction on Signal Processing, available at http://arxiv.org/abs/1204.5046*, pp. 1–29, 2011.

[61] Y. Liu and A. Petropulu, "Cooperative Beamforming in Multi-Source Multi-Destination Relay Systems with SINR Constraints," in *Proc. IEEE Internatinoal Conference on Acoustics, Speech, and Signal processing (ICASSP)*, Dallas, TX, Mar. 2010, pp. 2870–2873.

[62] Q. H. Spencer, A. L. Swindleherst, and M. Haardt, "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels," *IEEE Trans.*

on Signal Processing, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[63] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[64] Y. Huang and D. P. Palomar, "Rank-Constrained Separable Semidefinite Programming with Applications to Optimal Beamforming," *IEEE Trans. Signal Processing*, vol. 58, no. 2, pp. 664 – 678, Feb. 2010.

[65] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, vol. 20, May 2010.

[66] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[67] M. Grant, S. Boyd, and Y. Ye, "Matlab software for disciplined convex programming," Jan. 2009. [Online]. Available: www.stanford.edu/~boyd/papers/pdf/disc_cvx_prog.pdf

[68] J. Li, F. Roemer, and M. Haardt, "Efficient Relay Sharing (EReSh) between Multiple Operators in Amplify-and-Forward Relaying Systems," in *Proc. IEEE 4th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2011)*, San Juan, Puerto Rico, Dec. 2011, pp. 249 – 252.

[69] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multi-user MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 461–471, Feb. 2004.

[70] S. Toh and D. T. M. Slock, "A linear beamforming scheme for multi-user MIMO af two-phase two-way relaying," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2009.

[71] J. Joung and A. H. Sayed, "Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1833–1846, Mar. 2010.

[72] F. Roemer, J. Zhang, M. Haardt, and E. Jorswieck, "Spectrum and infrastructure sharing in wireless networks: A case study with Relay-Assisted communications," in *Proc. Future Network and Mobile Summit 2010*, Florence, Italy, Jun. 2010.

[73] F. Roemer and M. Haardt, "Algebraic Norm-Maximizing (ANOMAX) transmit strategy for Two-Way relaying with MIMO amplify and forward relays," *IEEE Sig. Proc. Lett.*, vol. 16, Oct. 2009.

[74] Y. Yu and Y. Hua, "Power allocation for a MIMO relay system with multiple-antenna users," *IEEE Transactions on Signal Processing*, vol. 58, pp. 2823–2835, May 2010.

[75] T. Tang, C.-B. Chae, R. W. Heath, and S. Cho, "On achievable sum rates of a multiuser MIMO relay channel," in *Proc. IEEE Int. Symp. on Information Theory*, Seattle, USA, Jul. 2006.

[76] C. Esli and A. Wittneben, "Multiuser MIMO two-way relaying for cellular communications," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, New Orleans, LA, Sep. 2008.

[77] A. Khabbazibasmenj, F. Roemer, S. A. Vorobyov, and M. Haardt, "Sum-rate maximization in two-way AF MIMO relaying: Polynomial time solutions to a class of dc programming problems," *IEEE Transactions on Signal Processing*, vol. 60, Oct. 2012.

[78] C. Wang, H. Chen, Q. Yin, A. Feng, and A. F. Molisch, "Multi-user two-way relay networks with distributed beamforming," *IEEE Transactions on Wireless Communications*, 2011, accepted for publication.

[79] J. Zhang, F. Roemer, M. Haardt, A. Khabbazibasmenj, and S. A. Vorobyov, "Sum rate maximization for multi-pair two-way relaying with single-antenna amplify and forward relays," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.

[80] N. T. H. Phuong and H. Tuy, "A unified monotonic approach to generalized linear fractional programming," *Journal of Global Optimization*, vol. 26, 2003.

[81] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *Signal Processing Magazine*, May 2010.

[82] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K., 2004.

[83] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization based beamforming," *Signal Processing Magazine*, May 2010.

[84] J. Li, A. P. Petropulu, and H. V. Poor, "Cooperative transmission for relay networks based on second-order statistics of channel state information," *IEEE Transactions on Signal Processing*, vol. 59, Mar. 2011.

[85] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. Tse, "Transmission techniques for relay-interference networks," in *Proc. 46th Annual Allerton Conf.*, UIUC, IL, Sep. 2008.

[86] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang, *Network Coding Theory*. now Publishers, 2006.

[87] J. Sykora and A. Burr, "Wireless network coding: The network aware PHY layer," in *Proc. Int. Symp. of Wireless Communication Systems (ISWCS)*, York, United Kingdom, Sep. 2010, tutorial.

[88] ——, "Wireless network coding: Network coded modulation in the net-

work aware PHY layer," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Cancun, Mexico, Mar. 2011, pp. 1–143, tutorial.

[89] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity bounds for two-way relay channels," in *Proc. Int. Zurich Seminar on Communications*, 2008.

[90] I.-J. Baik and S.-Y. Chung, "Network coding for two-way relay channels using lattices," in *Proc. IEEE Internat. Conf. on Commun. (ICC)*, 2008.

[91] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Join physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, Nov. 2010.

[92] S. Zhang and S.-C. Liew, "Channel coding and decoding in a relay system operated with physical-layer network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 788–796, Jun. 2009.

[93] P. Popovski and T. Koike-Akino, "Coded bidirectional relaying in wireless networks," in *Advances in Wireless Communications*, V. Tarokh, Ed. Springer, 2009.

[94] T. Koike-Akino, P. Popovski, and V. Tarokh, "Denoising maps and constellations for wireless network coding in two-way relaying systems," in *Proc. IEEE Global Telecommun. Conf. (GlobeCom)*, 2008.

[95] ——, "Denoising strategy for convolutionally-coded bidirectional relaying," in *Proc. IEEE Internat. Conf. on Commun. (ICC)*, 2009.

[96] ——, "Optimized constellations for two-way wireless relaying with physical network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 773–787, Jun. 2009.

[97] J. Sykora and A. Burr, "Layered design of hierarchical exclusive codebook and its capacity regions for HDF strategy in parametric wireless 2-WRC," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3241–3252, Sep. 2011. [Online]. Available: http://radio.feld.cvut.cz/~sykora/publications/papers/Sykora-Burr_2011-IEEE-T-VT.pdf

[98] ——, "Hierarchical alphabet and parametric channel constrained capacity regions for HDF strategy in parametric wireless 2-WRC," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Sydney, Australia, Apr. 2010, pp. 1–6. [Online]. Available: http://radio.feld.cvut.cz/~sykora/publications/papers/Sykora-Burr_2010-WCNC.pdf

[99] T. Uricar and J. Sykora, "Design criteria for hierarchical exclusive code with parameter-invariant decision regions for wireless 2-way relay channel," *EURASIP J. on Wireless Comm. and Netw.*, vol. 2010, no. Article ID 921427, pp. 1–14, Jul. 2010. [Online]. Available: http://radio.feld.cvut.cz/~sykora/publications/papers/Uricar-Sykora_2010-EURASIP-J-WCN.pdf

[100] J. Sykora and A. Burr, "Network coded modulation with partial side-information and hierarchical decode and forward relay sharing in multi-source wireless network," in *Proc. European Wireless Conf. (EW)*, Lucca, Italy, Apr. 2010, pp. 1–7. [Online]. Available: http://radio.feld.cvut.cz/~sykora/publications/papers/Sykora-Burr_2010-EW.pdf

[101] A. Burr and J. Sykora, "Extended mappings for wireless network coded butterfly network," in *Proc. European Wireless Conf. (EW)*, Vienna, Austria, Apr. 2011, pp. 1–7. [Online]. Available: http://radio.feld.cvut.cz/~sykora/publications/papers/Burr-Sykora_2011-EW.pdf

[102] M. Xiao and M. Skoglund, "Design of network codes for multiple-user multiple-relay wireless networks," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, 2009.

[103] C. Wang, M. Xiao, and M. Skoglund, "Diversity-multiplexing tradeoff analysis of multi-source multi-relay coded networks," in *Proc. Int. Symp. on Information Theory and its Applications (ISITA)*, 2010.

[104] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.