INCREASING THROUGHPUT IN MULTIUSER TWO-HOP RELAY NETWORKS USING TIME SCHEDULING AND SIGNAL PROCESSING

by

Fadhel Al Humaidi

Submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
October 2016

To all beloved ones who have been supporting me through my journey

Table of Contents

List of Figures

List of Tables

Abstract

This dissertation is concerned with the design of time scheduling and signal processing algorithms to mitigate interference effects in wireless networks with successive relaying where the nodes in the system are equipped with multiple antennas. These networks as a form of cooperative communications and multiple-input multiple-output (MIMO) systems have the promise of significantly higher spectral efficiency and reliability than the conventional networks. Reusing the same spectrum by simultaneous multiple transmissions in successive relaying increases spectrum efficiency however it also causes interference which if not mitigated may limit the overall wireless network performance. With these motivations, this dissertation presents three main contributions to lessen or completely cancel relay interference and improve bandwidth efficiency.

First, we consider relay-assisted downlink transmissions to support increased data rates for single antenna users. In the first stage of the communication process, the base station with multiple antennas sends to different users. In the next stage, users and their assigned relays form independent networks. Relays using an amplify-and-forward (AF) scheme aid the recovery of messages by the user which solves linear system of equations developed in this dissertation. All of these subnetworks utilize the same bandwidth concurrently producing an acceptable level of multiple access interference (MAI) where the MAI is controlled using specialized frequency reuse plan.

Secondly, we develop receiver-based inter-relay interference (IRI) cancellation in AF two-hop systems with successive transmissions. In networks with strong inter-relay channel gains, representing IRI with the recursive terms and exploiting opportunistic listening, in this dissertation, this interference is fully removed using limited channel state information (CSI) about channel gain matrices from both hops.

Finally, to complement our second contribution, we develop transmitter precoding to cancel IRI in systems with successive relaying. When designing the scheme, we benefit from the broadcast channel characteristic of the radio channel and the knowledge at the source of all the data contributing to IRI.

This dissertation advances the theory of successive relaying through its integration with time scheduling and signal processing within the framework of cooperative communications system designs. The results presented in this work are applicable to wireless networks on downlink and uplink in systems ranging from cellular to ad-hoc networks.

List of Abbreviations

The following abbreviations and acronyms are used in this dissertation.

AF	amplify-and-forward
AWGN	additive white Gaussian noise
BER	bit error rate
BPSK	binary phase shift keying
BS	base station
CSI	channel state information
DF	decode-and-forward
FDMA	frequency division multiple access
FIR	finite impulse response
IIR	infinite impulse response
i.i.d	independent, identically distributed
IRI	inter-relay interference
LOS	line of sight
MAI	multiple access interference
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MS	mobile station
MU	multiuser
OFDM	orthogonal frequency-division multiplexing
pdf	probability density function
RF	radio frequency

RV	Random Variable
SIC	successive interference canceler
SIMO	single input multiple-output
SISO	single input single-output
SIR	signal to interference ratio
SNIR	signal-to-interference-to-noise ratio
SNR	signal-to-noise ratio
SU	single user
TDMA	time division multiple access
TS	time slot
ZF	zero-forcing

Acknowledgements

I would like first to express my thanks and appreciation to my supervisor Dr. Jacek Ilow for his support, motivations, teaching and patience. I would like also to thank him for his always availability for assistance and advice over the period of my studies.
Second, I would like to thank Dr. Chen and Dr. Phillips for being members of my PhD committee and for readying my thesis and and acting as internal examiners. My thanks extended to Dr. Petersen for accepting to be the external examiner of my PhD defence.
Third, I would like to thank my colleagues Scott Melvin, Javad Hoseyni, Aasem Alyahya, Rashed Alsakarnah, Md Sabbir Hussain, Mahdi Attaran, Zichao Zhou and Hui Xiong for their friendly environment and support during my PhD studies.
Forth, I would like to thank my family with special thanks to my wife Asia Al Bashrawi for the understanding and support they have shown since the arrival to Canada. I would like also to thank my big and lovely family back in Saudi Arabia for their permanent support.
Finally, I would like to express my deep appreciation and thanks to King Salman scholarship program for the generous financial support during my studies in Canada.

Chapter 1

Introduction

With the growing demand for wireless communication services to support higher bit rates, there is an urgent need to increase the bandwidth efficiency of new communication systems. One of the most successful and efficient solutions for addressing this and other issues, such as enhancing transmission reliability, is the deployment of multiple-input multiple-output (MIMO) technology. However, MIMO systems are often designed for point-to-point or single user configurations only. Recently, several new approaches have been proposed to achieve high utilization of radio spectrum and power resources, for example, with cooperation among terminals, multiuser (MU) MIMO, and more advanced relaying strategies offering some of the more promising breakthroughs. This dissertation thus follows the trend in radio communications design of moving away from maximizing the capacity of individual links toward optimizing the capabilities of multiuser networks.

The use of multiple transmit and/or multiple receive antennas, where different data streams are transmitted at the same time and frequency, introduces new possibilities for advancing spectrum utilization [1]–[5]. In addition to time and frequency aspects, a new spatial dimension is being explored in order to develop even more effective approaches to deal with long-standing problems in the design of communication systems, such as bandwidth and power efficiency. MIMO transmission schemes represent a paradigm shift from single-input single-output (SISO) transmissions, and encompass various diversity schemes developed previously for multiple-input single-output (MISO) networks [6] and single-input multiple-output (SIMO) networks [7]. In multiuser MIMO networks which are still under development, spatial degrees of freedom offered by multiple antennas are exploited to enhance the system capacity, e.g., by scheduling radio terminals to share the spatial channel simultaneously. This dissertation works with time scheduling, frequency reuse strategies and transceiver methodologies in MU-MIMO to obtain performance gains and beneficial trade-offs.

Relays that retransmit signals between radio terminals have long been used in communication networks to extend the range [8]–[10]. However, only recently have more advanced relay network configurations been proposed to improve bandwidth efficiency as well. These configurations take advantage of more complex signal processing operations in the terminals and relays, where the radio terminals deploy pre- and post-processing strategies and the relays transmit signals that are processed versions of the received signals [11], [12]. Multihop networks, in particular the two-hop transmissions considered in this dissertation, are a special case of cooperative communications which have recently attracted considerable attention [13], [14]. Of particular interest are MIMO two-hop channels where amplify-and-forward (AF) or decode-and-forward (DF) relay terminals assist in communications between source and destination terminals [15]. When all relays operate in half-duplex mode, the relaying process consumes more time than a direct communications process, because the relays require additional time to forward signals to the destinations [16]. In this area, a promising solution to overcome the loss in spectral efficiency (the pre-log factor $\frac{1}{2}$ in capacity calculations) consists of alternate relaying, which reuses source transmission time slots (TSs) for relay transmissions. In this dissertation, system level improvements are motivated by this type of time slot reuse, with topologies that are more complex than those of conventional alternate relaying systems that have a single source and a destination pair aided by alternate retransmissions of two relays.

Because of the broadcast nature of the wireless medium, when multiple signals are transmitted at the same time and frequency to increase bandwidth utilization, multiple access interference (MAI) is inevitably introduced. When relays are deployed, they can also create inter-relay interference (IRI). Uncoordinated interference reduces wireless network throughput. In order to address this problem, this dissertation seeks to design signal processing strategies and wireless network protocols that manage interference in MU-MIMO systems in order to achieve the highest network performance.

In wireless systems with MIMO terminals and relays, multiple spatial streams should not cause interference to the data streams that are used to recover the data of interest. Therefore, a major challenge is to design the system so as to obtain interference-free signals at the destination nodes. This objective can be approached by deploying interference-cancellation techniques performed at the source, relay, or destination nodes, or a combination of these. In these scenarios, the signal processing algorithms used to cancel multiple access and IRI depend upon knowledge of the channel state information (CSI), i.e., channel gain matrices, in the system and upon the type of cooperation between the terminals and the relays. In this work, the cooperation of terminals and relays is designed via spatial reuse of time slots and specialized time scheduling that are unique to this dissertation. Time scheduling and the spatial reuse of transmission time slots are interference avoidance approaches, and the interference mitigation algorithms developed in this dissertation combine system and corresponding signal processing design aspects.

Specifically, the first chapter of this dissertation discusses transmission scheduling to improve the capacity of cellular networks aided by relays on a downlink, from a single base station (BS) to multiple mobile stations (MSs), with CSI available only at the receiving side of the network. Under such conditions the BS, which has M transmit antennas, cannot zero force its signals to communicate with M single antenna MSs simultaneously. Instead, it has to use orthogonal medium access to divide the spectrum equally among users. The capacity per MS in this scenario is bounded by the multiplexing gain of $1/N$ where N is the number of MSs sharing the spectrum. In Chapter 2, a novel scheduling method is developed that increases the capacity by increasing the number of multiplexed signals from one to $MN/(M+N-1)$. Chapters 3 and 4 address the capacity reduction caused by using half-duplex relays, when relays are employed to extend the network coverage. In this area, methods are developed to mitigate the IRI created in successive relaying networks.

1.1 Dissertation Objectives, Contributions, and Organization

1.1.1 Objectives

The general objective of this dissertation is to enhance the throughput of wireless relay networks through advanced signal processing and transmission scheduling. The focus is on networks where relays have limited involvement in processing the signals they receive. In these networks, capacity challenges are addressed according to the availability of CSI at every node in the different network topologies. The guiding principle in the present approach is to reuse transmission time slots in wireless relay networks that are under-utilized in conventional communication systems. Once opportunities to reuse time-slots or available information at different points in the network have been identified, signal processing algorithms are developed to take advantage of system level transmission scheduling. The models developed here consider

MIMO or practical variants of radio front ends utilized in current and upcoming systems. In this dissertation, because of the system and algorithm complexity, it was not feasible to use only analytical methods to evaluate the performance of the algorithms developed. Therefore, a combination of analytical and simulation results is used. All proposed schemes and algorithms have been validated and compared though a large number of computer simulations run in the MATLAB® computing environment. In this type of research, this is an acceptable research methodology for solving networking problems and arriving at new knowledge and designs.

The objectives and system models of individual chapters of this dissertation are summarized below.

The objective of the second chapter is to increase the throughput in a single-cell cellular system, with a BS located at the center of the cell and equipped with multiple antennas. This BS serves multiple single antenna MSs which are distributed randomly in the cell according to the Poisson point process. Here it is assumed that any idle MS can act as a cooperative relay node. When the CSI is not available at the BS and relays are not deployed, the maximum number of spatial streams that can be transmitted simultaneously in the downlink is one. Here the key feature of the cooperative transmission is the goal to encourage multiple single antenna users to share their antennas cooperatively. In this way, a virtual antenna array can be constructed and, as a result, the overall network capacity is improved significantly.

Chapter 3 considers networks with dedicated half-duplex relays. The objective is to increase the throughput of one-way two-hop relay networks, where signals arrive at their destinations exclusively from the relay path. This objective is achieved by using successive relaying methods for MIMO networks. The IRI created in these networks is cancelled at the destination through advanced signal processing, by taking advantage of the unique interpretation of the IRI arising in MIMO successive relaying systems.

The network topology considered in Chapter 4 is similar to that of Chapter 3, however, here the IRI is mitigated by processing at the transmitter side. The solutions provided in Chapter 3 left some problems unresolved, such as the situation where a destination can hear one relay but not the other. In this case, the destination cannot use the scheme developed in Chapter 3 to remove the IRI. Thus, the first objective of Chapter 4 is to address the IRI when the destination is close to one relay and far from the other. Another objective of Chapter 4 is to offer a solution based on schemes developed in the dissertation, for the situation where single antenna MSs are located close to the boundary of a cell.

1.1.2 Contributions

Results of the research described in this thesis have been published in the form of conference papers in [17]–[20]. In addition, one journal article has been submitted and another is in preparation. The details of these publications are outlined below.

Refereed Conference Proceeding Publications

[C-1] **F. Alhumaidi** and J. Ilow, "*Relay-assisted downlink transmissions to support increased data rates for single antenna users*," in 2014 IEEE Global Communications Conference, 8-12 Dec 2014, pp. 4150-4155.

[C-2] **F. Alhumaidi** and J. Ilow, "*Alternate AF MIMO relaying systems with full inter- relay interference cancellation*," in 2015 IEEE Vehicular Technology Conference (VTC Fall), 6-9 Sep 2015.

[C-3] **F. Alhumaidi** and J. Ilow, "*Transmitter precoding to cancel inter-relay interference in AF systems with successive transmissions,*," in 2016 IEEE 29th Canadian Conf. on Elect. and Comput. Eng. (CCECE), May 2016, pp. 60-64.

[C-4] **F. Alhumaidi** and J. Ilow, "*Increasing throughput in multi-way three-user MIMO networks using successive relaying and IRI cancellation,*," in Wireless and Mobile Computing, Networking and Communications (WiMob), 2016 IEEE 12th International Conference on, 2016.

Journal Papers (in preparation)

[IPJ-1] **F. Alhumaidi** and J. Ilow, "*Inter-relay Interference Modeling and Cancelation in MIMO Wireless Networks with Successive Transmissions,*," to be submitted to Computer Communications.

[IPJ-2] **F. Alhumaidi** and J. Ilow, "*Increasing Throughput in Relay-Assisted Cellular Networks with Single Antenna Users,*," to be submitted to the IEEE Communications Letters.

The research in each of the papers cited above was initiated and carried out by the principal author of the papers, who is also the author of this dissertation.

The research contributions of this thesis can be classified into four areas, which correspond to the four main chapters of the dissertation. The chapters and the specific papers that relate to them are listed below.

Chapter 2: **Increased Data Rate for Single Antenna Users**

Novel transmission scheduling for downlink MU-MISO cellular networks is proposed, to increase the throughput of the network. Cooperative relaying is used to recover independent messages at the desired MSs by taking advantage of independent signal replicas and channel reuse at different locations, with reduced co-channel interference.

Chapter 3: **Continuous Transmission in Two-Hop Relay Networks**

An IRI cancellation scheme at the receiver(s) in MIMO amplify-and-forward relay networks is proposed in order to increase the throughput of these networks by using successive relaying for one-way communications. The proposed scheme is generalized to networks with various topologies, by using an opportunistic listening approach. [C-2] and [C-4].

Chapter 4: **Source Precoding to Cancel IRI in Two-Hop Relay Networks with Successive Relaying**

A transmitter precoding technique to cancel the IRI in a cellular network is proposed for downlink transmissions. With an IRI filter implemented at the transmitter side, the proposed scheme solves the challenge facing IRI cancellation at the destination, when the destination can listen to one of the relays but cannot hear the others. This chapter also combines the two IRI cancellation schemes to address the case of single antenna MSs located at the boundary of a cell. [C-3] and [IPJ-1].

1.1.3 Organization

This dissertation is organized into five chapters. The first chapter outlines the objectives, contributions and thesis organization, as well as reviews the general concepts and models used throughout the dissertation, while the last chapter contains the conclusions and offers suggestions for future work that could be carried out to expand upon the results.

Chapter 2 discusses downlink transmission in wireless cellular networks from a cooperative point of view. In Section 2.1, the network topology and signal flow are introduced. This section also presents the relay retransmission strategy used in the system model. In Section 2.2, interference limiting is introduced through spatial reuse of the channel at the relay. Then the proposed signal processing scheme is generalized by introducing a clustering technique to limit the interference between subnetworks.

Section 2.3 presents the results in terms of capacity versus the signal-to-noise-plus-interference ratio (SNIR).

Chapter 3 focuses on two-hop relay networks following the conventional successive relaying approach. The first section of this chapter validates the assumption that signals leaking from the source directly to the destinations have negligible gains. In Section 3.1, successive relaying is introduced by using a $1 \times 2 \times 1$ network. The methodology of cancelling the IRI at the destinations is also proposed. In Section 3.2, the concept of opportunistic listening is introduced to generalize the IRI cancellation scheme to various types of network. The performance of the proposed IRI cancellation scheme is discussed in Section 3.3.

In Chapter 4, IRI cancellation is implemented at the transmitter side.

The network topology is considered, where there is one source and multiple receivers with reception aided by dedicated relays. The relays are also separated, so that a destination can hear one relay but not the other relays. The layout and methodology of this chapter is presented in Section 4.1. In Section 4.2, the two schemes developed in Chapters 3,4 are combined in order to address the capacity of the network partially introduced in Chapter 2. This section deals with a cellular network in which single antenna users are located at the boundary of the cell. The BS and relays are assumed to be equipped with multiple antennas. Section 4.3 presents the performance discussion for the chapter.

The remainder of this chapter is organized as follows: Sections 1.2, 1.3, 1.4 and 1.5 present the building blocks related to the research in more detail and review the corresponding literature in the areas of (i) wireless channel modelling, (ii) multiple access strategies, (iii) relay and cooperative networks, and (iv) successive relaying.

1.2 Modelling Wireless Communication Channels

This dissertation contributes in the field of signal processing algorithms for wireless relay networks. The results obtained are verified using prototype simulation models, which is an initial stage in communications system design. This methodology is widely adopted and accepted in the field of communications research. Channel models are the basis for the conceptual stage of communications system design, where they are used in order to predict and compare the performance of wireless communications systems under realistic conditions, and to devise and evaluate methods for mitigating channel impairments. This section reviews some of the wireless communications channel models used in this dissertation.

1.2.1 Additive White Gaussian Noise

A natural phenomenon in the radio frequency (RF) front end of wireless receivers is the existence of wideband noise. This noise is universally present in the RF front end, e.g., as thermal noise, resulting from a large number of random small interference effects [21]. As stated by the central limit theorem, the large number of random variables (RVs) forms a Gaussian distribution, which has the probability density function (pdf):

$$pdf(n) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp^{\frac{-(n-\mu)^2}{2\sigma_0^2}} \tag{1.1}$$

with a zero mean ($\mu = 0$) and a noise variance (σ_0^2) that represents the power spectral density ($\frac{N_0}{2}$ [W/Hz]). This dissertation deals with received signals after down-converting and matched filtering, and noise is represented as a RV rather than as a stochastic process.

1.2.2 Rayleigh Fading Channels

In travelling to receivers, signals are subject to refraction, reflection, and scattering as they pass various types of objects. Thus, the receiver obtains a combination of multiple copies of the same set of signals, yet every copy has its own attenuation and time of arrival. It is assumed that the delays are smaller that the symbol duration, so that it is not necessary to deal with time dispersive (frequency selective) channels. This phenomenon is known as fading, where the combined effects of all copies of the signals received from various multipaths are represented as the multiplicative factor affecting the received signals.

The most common and detrimental situation in wireless communications is the lack of a line of sight between two communicating devices. When passband signals arrive along two independent components, i.e., in-phase and quadrature, the fading in each case is represented as an independent, identically distributed (i.i.d.) Gaussian RV. Conventionally, the resulting complex variable representation of this effect is denoted as h. Then the gain of the fading channel is: $|h| = \sqrt{\Re(h)^2 + \Im(h)^2}$, where $\Re(h)$ and $\Im(h)$ denote the real and imaginary values of h, respectively, and h is a complex normal RV: $h \sim \mathcal{CN}(0, \sigma_h^2)$. Therefore $|h|$ has a probability density function of

$$pdf(|h|) = \frac{2|h|}{2\sigma_h^2} \exp^{\frac{|h|^2}{2\sigma_h^2}} \tag{1.2}$$

where σ_h^2 is a power scaling parameter. Hence, $|h|$ has a Rayleigh distribution. This applies only when there is no line-of-sight (LOS) path, and (1.2) represents the most adverse type of fading.

1.2.3 Large-Scale Attenuation

As is the case with any transmission medium, wireless signal power decays with distance as the signals travel through a channel. This loss in signal power is known as deterministic signal attenuation. Various mathematical models for different environments are presented in the literature in order to capture this phenomenon [22]. This work adopts a generalized formula which links the attenuation to the distance travelled $d > 1$ as follows:

$$p_r(d) = \frac{p_t}{d^{\alpha}} \tag{1.3}$$

where $p_r(d)$ and p_t represent the power of the received and transmitted signals, respectively. The α parameter corresponds to the propagation condition; this value normally ranges from 2 in free-space conditions to 6 in a dense urban area.

1.3 Wireless Network Classification

Wireless networks vary in their topology and application. The capacity of each network also differs according to the number of nodes in the network, the number of antennas at these nodes, and the available knowledge at each node. This section describes the classification of various networks that play an important role in the development of scheduling and signal processing algorithms in this dissertation.

1.3.1 One-to-One Networks

One-to-one networks consist of two nodes that communicate with one another. In such a network the available channel is used exclusively by these two nodes, and in their communications there is no interference from external nodes [23].

In relay networks, one-to-one networks can use one or more intermediate nodes to assist in communications between the two nodes. Of particular interest is the $1 \times 2 \times 1$ network, where two intermediate nodes act as relays. In one-way data flow, the two relays take turns relaying the signals from the transmitter to the receiver, in a process referred to as alternate relaying [24]. In order for this task to be performed successfully, an interference cancellation scheme should be deployed. In Chapter 3, this type of network serves as an element for developing the interference cancellation scheme.

1.3.2 Multiple Access Networks

Multiple access networks are networks where multiple transmitters transmit to a single receiver. Thus, multiple signals access the reception end of a single receiver [25]. Chapters 3 and 4 address the successive relaying problem for a multiple access network, where two or more relays are introduced to the network.

1.3.3 Broadcast Networks

In contrast to a multiple access network, in a broadcast network the transmitter sends signals to multiple receivers [26]. Chapters 3 and 4 also address the successive relaying problem for such a network, where two or more relays are available to assist communications between the transmitter and the receivers.

1.3.4 Interference Networks

In an interference network, multiple one-to-one pairs communicate by using the same available channel spectrum [27]. In Chapter 2 and part of Chapter 4, this type of network is addressed in terms of communications between the relays and the receivers.

1.3.5 Cellular Networks

In cellular networks, a geographical region is divided into multiple cells, and each cell is served by a base station located ideally at the center of the cell. Moreover, each cell is allocated a set of frequency channels that differ from those in adjacent cells.

The reuse of these channels in other cells determines the level of mutual interference between cells [28].

In order to accommodate more users within a cell, the cellular system adopts channel access techniques such as frequency division multiple access (FDMA) and time division multiple access (TDMA). These techniques arrange how users share the available channel without interfering with each other. FDMA divides the frequency band into subchannels, and at all times each user uses a unique subchannel. In TDMA, users are stacked in a queue, so that each MS uses the whole spectrum for a portion of the time [28].

The concept of a cellular system is utilized in Chapter 2 in order to accommodate more users in the proposed scheme.

1.4 Relay and Cooperative Networks

Relays offer the capability of addressing many limitations that exist in wireless communications, e.g., by extending the coverage of a network [10], [12], [29] or by improving the quality of the service [30], [31]. In order to overcome the many challenges faced by modern wireless networks, such as bandwidth efficiency, in the past decade researchers have extensively investigated various aspects of relay networks. This research has opened up many avenues for improving wireless systems, e.g., by exploring the type of relay, relaying strategies, and the impact of relaying on capacity. This section presents some aspects of relaying networks that are related to the work in this dissertation.

1.4.1 Cooperative Networks

Conventionally, a relay is a dedicated wireless network node that is a physical device which is separate from the communicating terminals. In cooperative networks, inactive nodes or terminals can act as cooperative relay nodes [32]. Cooperating nodes invest power in the relaying process, and many compensation strategies have been suggested to motivate cooperation among nodes.

In cooperative networks, relaying nodes are usually users that already exist in the network, which means that no additional hardware needs to be introduced to the network [32], [33]. Moreover, there could be more than one cooperating node near the active node, allowing some freedom in the choice of relaying node in order to improve the signal quality [33]. However, cooperative nodes are not always positioned in a fixed location, and could be mobile users [34], [35]. This results in some challenges for the transmitter, such as estimating CSI for the cooperative nodes, and relying on a certain node to act as a relay for a relatively long period of time.

In the cooperative relaying employed in Chapter 2, the network with cooperative users does not require CSI at the transmitter, and every cooperative user can cooperate for a only a short period of time.

1.4.2 Dedicated Relay Nodes

Another approach to relay networks is to use dedicated nodes that are responsible only for relaying signals from the source to the destination. The location of such relays is usually based on design factors that serve to optimize the performance of the network [36], [37].

Dedicating nodes to relay signals between the source and the destination requires additional investment in network hardware. With a source located in the center of a cell, a single relay will cover only part of the cell, and further relays are necessary to cover the entire cell. Even in networks where the source transmits in a single direction, single relays face challenges in terms of reliability, due to channels with deep fading, as well as capacity problems, which are the focus of interest of this dissertation.

In Chapters 3 and 4, dedicated relays are used to address the capacity challenge in half-duplex relay networks.

1.4.3 Relaying Strategies

Signals that arrive at the relays are attenuated and modified due to the nature of the environment surrounding the relays. Therefore, the relays must process the signals received before retransmitting them to the next node. Methods used by relays to deal with the signals include amplify-and-forward (AF) and decode-and-forward (DF) approaches, which dominate relaying schemes [38]. These two strategies differ with regard to performance, complexity, flexibility, and how the signals are handled.

The AF scheme deals with signals at the analog level. Relays receive the signals and amplify them to the desired threshold before retransmitting them to the next nodes [38], [39]. If the CSI is available, the signals may also be beamformed in a certain direction to meet network requirements [10]. This scheme is preferable for systems where processing data at the bit level requires complex manipulations, or where the relays lack the capability to decode the signals. However, as is the case with any receiver, signals received at the relays are accompanied by additive white Gaussian noise (AWGN). In AF schemes, amplification of the signals also results in the amplification of noise [40].

The DF scheme handles signals at the digital level. Relays decode the signals by removing all of the effects from the receiving side, and then encode the signals to forward them to the next nodes [38]–[40]. Although the removal of noise is an advantage of the DF scheme, this requires full data processing and decoding. In many scenarios, as described in Chapter 2, relays do not have the capability to decode signals, due to an insufficient number of antennas. As discussed in Chapter 3, DF involves complex processing to remove the transmitted noise and interference.

This work primarily adopts the AF scheme, except for part of Chapter 4, where the DF scheme is used in a downlink cellular system.

1.4.4 Capacity Challenges

Although relays can provide promising solutions in terms of reliability and range extension, there are also challenges such as a decrease in capacity. Communicating via conventional relays requires approximately twice as much time as direct communication between the source and the destination. In conventional relay networks, a time slot is used by the source to send signals to the relay, and another time slot is utilized by the relays to retransmit the signals to the destination. In the literature, this drop in the capacity of the link between the source and the destination is referred to as the prelog factor [15].

The capacity is the maximum rate that can be achieved when transmitting over a communication channel. For a direct single-input single-output Rayleigh channel, the average capacity C can be expressed as:

$$C = \mathbf{E}\left(\log_2(1+\frac{|h|^2 P}{\sigma^2})\right) \qquad \text{[bps/Hz]} \tag{1.4}$$

where h is the channel between the source and the destination, P is the signal transmit power, σ^2 is the variance of the AWGN, and $\mathbf{E}(\cdot)$ is the expectation operator [1]. When the signal passes through a relay, the capacity is scaled by $\frac{1}{2}$ due to the additional time slot used in the process. This is known as the prelog factor. Note that the exact expression of the average capacity in relays depends on the type of network and the relaying strategies adopted by the network.

In order to overcome capacity deficiency, studies have provided solutions for some types of network, however this remains an unresolved problem for other networks. In two-way communications, the prelog factor has been improved to $\frac{2}{3}$ in decode-and-forward schemes. This is accomplished using network coding by reducing the relay transmissions for two independent signals from two time periods to one time period [10]. The relay waits for two terminals to send their signals consecutively. It then encodes the signals so that each terminal can remove its own data and can decode the desired data.

The physical-layer (analog) network coding scheme using AF went further and eliminated the prelog factor by allowing two terminals to send at the same time. The relays then retransmit the received signals to two terminals. The terminals know their signals and have the channel information required to remove their own signals and to decode the desired signals [12].

These solutions do not work for one-way flow, when the signals travel in one direction. For some networks, alternative relaying solutions are already described in the literature. In this dissertation, this challenge is addressed for multiple AF MIMO networks in Chapters 3 and 4 and for multiuser MISO networks in part of Chapter 4.

Chapter 2 goes a step further, by not only removing the effects of the prelog factor but also improving the capacity of single antenna receivers beyond their one-to-one limitations.

1.4.5 Interference

The omnidirectional propagation of wireless signals imposes limitations on the utilization of the available spectrum. Multiple communication pairs that are located close to one another cannot use the same frequency band at the same time without being subject to mutual interference. In this context, interference is an undesired, deterministic signal that arrives at the receiver together with desired signals. If the spatial dimension (i.e., the number of antennas) is smaller than the number of received signals, including the interfering signals, then the interference must be treated in order to ensure reception quality.

The available spectrum is very limited, and is congested with traffic. Reuse of the spectrum is therefore necessary to meet service demands. Thus, addressing the interference phenomenon is an essential area of research for increasing the capacity of wireless networks. Approaches for dealing with interference that are described in the literature include treating interference as noise, interference avoidance, interference alignment, and interference cancellation.

Interference as Noise

The power of a signal degrades with the distance travelled. When interference signals arrive at the receiver with a power below a certain threshold, the receiver treats this interference as AWGN, and decodes the signals without trying to remove the effects of such interference [41]. This approach can solve the problem if the distance between two interfering pairs is sufficient to reduce the interference to a level where it is regarded as noise. Moreover, it is also important for users to maintain power control, in order to achieve throughput without impairing the signals of other users. If users take a selfish approach and increase signal power to improve their SNIR, this causes interference to increase for other users. Other users then have to increase their power as well to compensate, thus increasing the interference level for all users.

In cellular networks, reuse of the spectrum is implemented by dividing the area into multiple cells. The cells that uses the same frequency band are separated so that interference in the system falls below a defined threshold. In Chapter 2, this concept is used to accommodate more users in the network.

Interference Avoidance

The conservative approach of interference avoidance aims to eliminate interference in the system, especially when the power of interfering signals is comparable to that of desired signals. In cellular systems with multiple users in a cell, users can be served simultaneously over multiple frequency bands via frequency division multiple access, or served using the full available spectrum but separated in time via time division multiple access, or served with a combination these two multiple access techniques [28]. Each of these techniques ensures that users do not interfere with one another, and that the channel is not shared concurrently. This approach is also used in Chapter 2 during the transmitter phase.

Interference Alignment

Channel utilization attained by using the two approaches described above remains far below the theoretical system capacity of the model considered. Interference alignment addresses interference from a cooperative point of view. Instead of dealing with interference at a microscopic level at the receiver, interference alignment considers interference cancellation from a macroscopic perspective. With the interference alignment approach, users have to sacrifice some of their instantaneous capacity in order to achieve a gain in capacity for the system as a whole [42]. For example, a system with three pairs of one-to-one communications where all nodes are equipped with two antennas could multiplex a maximum of two signals for a single use of the spectrum, if interference avoidance is used. This represents a capacity of $\frac{2}{3}$ per user in the network [43]. With the interference alignment approach, every user should sacrifice one dimension and utilize the other dimension, permitting all users to utilize the spectrum simultaneously. Users then cooperate among themselves and control their signals to meet two conditions [42], [43]:

1. At the desired receiver, the signal occupies the desired dimension allocated to that signal.
2. At the other receivers, where the signal is considered to be interference, it is stacked with other interference in the sacrificed dimension.

Although this approach increases capacity utilization, it faces many challenges such as global CSI availability, synchronization, and the level of performance in a low signal-to-noise ratio (SNR) regime.

Interference Cancellation

Interference cancellation is a signal processing approach that allows users to utilize the spectrum concurrently. Through signal processing, users remove interfering signals at a certain point in the network. In wireless networks affected by interference, successive interference cancellation techniques can be used at the receiver to estimate and filter signals according to their power levels [44]. This approach works well if signals arrive at the receivers at different strengths.

In successive relaying networks, the transmitter sends signals to a listening relay while a transmitting relay relays a previous set of signals to a destination. If the transmitting relay uses the same frequency band as the source, this causes interference at the listening relay, referred to as IRI [45]. In this scenario, successive interference cancellation may not be feasible due to comparable signal strengths, and a lack of capability of the relay nodes to perform the required signal processing.

The interference cancellation techniques proposed in this dissertation in Chapter 3 and Chapter 4 are developed to address IRI in successive relaying for multiple relay networks.

1.5 Successive Relaying

Research to overcome drawbacks in the throughput of half-duplex wireless relay networks has taken two paths. One approach utilizes two-way communications, where the data flow is in two directions and a receiver is also a transmitter. This approach applies some kind of network coding technique to reduce the transmission time [10], [12], and is not suitable for networks where data are to be transmitted in a single direction or where different bands are used for uplink and downlink transmissions.

The second approach addresses one-way transmission. In an effort to improve diversity gain and reduce relaying time, a transmitter sends signals to all users in sequence, then the relays which overhear these signals retransmit the desired signals all at the same time [46]. This approach improves the prelog factor for the relay capacity from $\frac{1}{2}$ to $\frac{K}{K+1}$, where K represents the number of users in the network.

The work described in [46] paved the way for successive relaying techniques that can also overcome prelog factor limitations. Successive relaying, or in some cases alternate relaying as shown in Fig. 1.1, allows the transmitter to transmit continuously, while relays take turns listening to the transmitter. Using the same band as utilized by the transmitter, a relay which has completed a listening session retransmits previously received signals to the receivers. Listening and transmitting are performed by the relays successively and continuously. This improves the prelog factor to $\frac{T}{T+1}$, where T is the number of time slots used by the transmitter without interruption, before one or more relays take over and retransmit data while the original transmitter withholds its transmissions.

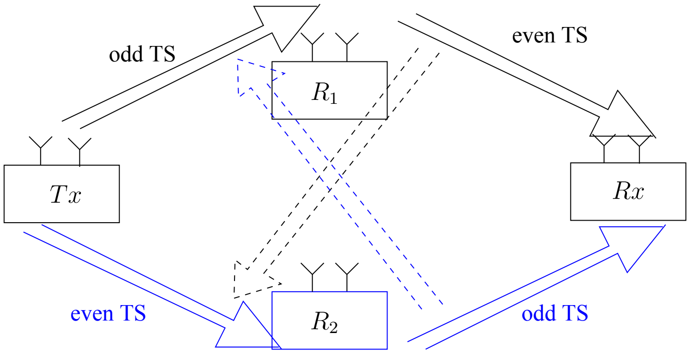


Figure 1.1: Two-path alternate relaying in a $1 \times 2 \times 1$ network.

Although the prelog factor is significantly improved, successive relaying has a cost. During the listening session, the signals transmitted by the transmitter to the listening relay are accompanied by interference (shown by the dashed path in Fig. 1.1) from the transmitting relay. If this IRI is not treated at the relays either by the relays themselves or through precoding by the transmitter, it degrades the quality of the signals received at the final destination(s).

If this IRI is not treated (either through (i) signal processing at the relays themselves or (ii) precoding by the transmitter or (iii) post-processing at the destination, or (iv) any combination of (i), (ii) and (iii)), it degrades the quality of the signals received and decoded at the final destination(s).

In successive relaying, IRI cancellation is a major challenge. One-to-one AF networks treat interference at the bit level by cancelling interference successively at the receiver [15]. This method requires extensive memory at the receiver, because it depends upon previously received signals and cancelling of the signals one by one in a decision feedback type of processing. Moreover, this method removes only deterministic interference signals, leaving the desired signals with accumulated noise. At the analog level, IRI cancellation can be achieved by using two consecutively received signals [47]. This method not only cancels the interference but also removes most of the accumulated noise. In this dissertation, this method is generalized to various MIMO networks.

In MIMO networks, alternate relaying can be performed by using interference alignment [48], however, this approach is suitable only for systems with an even number of antennas at each node. Furthermore, this method recovers only part of the lost capacity. In one-to-one alternate relay networks, IRI cancellation can also be performed at the relays, following a zero-forcing (ZF) approach [49]. However, this method requires a large number of antennas at the relays, which may not be practical.

Chapter 2

Relay-Assisted Downlink Transmissions for Single Antenna Users

The spatial spectrum reuse concept can be used to increase the capacity of cellular systems. Spectrum reuse allows multiple BS-MS pairs to use the same spectrum concurrently, provided that they are sufficiently separated in space to minimize co-channel interference while maintaining the required quality of service [28].

In this chapter, spatial spectrum reuse (also referred to as *spectrum reuse* is adopted in order to increase the downlink throughput of a single cell MISO cellular system, where inactive MSs serve as cooperative relays to assist the operation of the centrally located BS. Specifically, transmission scheduling and the corresponding signal processing are developed in order to increase the number of multiplexed signals within a channel use. For this purpose, the communication process is divided into two stages. In the first stage, a transmitter with multiple antennas using time division multiplexing transmits to different users in their allocated time slots; the number of messages sent is equal to the number of transmit antennas in the MISO frame. In the next stage, each user with its assigned relays forms an independent network, where relays that use an amplify-and-forward scheme aid the recovery of messages, with the user solving a linear system of equations. All of the subnetworks utilize the available channel concurrently, producing an acceptable level of multiple access interference (MAI), where the MAI is controlled by using a clustering technique tailored to the location of users and potential relays. This approach increases the number of downlink data streams, in order to close the gap with the maximum number of achievable simultaneous data streams determined by the number of antennas at the BS.

Here it is assumed that the active MSs and relays know their receiving instantaneous CSI, and that there is no need for CSI at the BS.

The first section of this chapter introduces the system model and transmission methodology. Section 2.2 first introduces the proposed signal recovery scheme in a simplified network, with two users aided by two relays. The scheme is then extended to a more complex network. Section 2.3 discusses the performance of the system, and the chapter is summarized in Section 2.4.

2.1 System Model and Two-Stage Relaying

This chapter considers an M antenna base station (BS) that serves a total of L single-antenna mobile stations (MSs). All antennas are omnidirectional. There are N active MSs and $L-N$ idle MSs over a communication cycle. The $L-N$ MSs are considered to be idle in the sense that they do not have their own messages to send or receive over a complete communication cycle. A complete communication cycle starts when the BS sends the first signal and ends when the active MSs receive the required signals so as to be able to decode the original messages. It is assumed that any idle MS can act as a cooperative relay node. Figure 2.1 shows an example of a BS with two antennas serving multiple MSs; however, in general the number of antennas at the BS is $M \geq 2$. Furthermore, Fig. 2.1 represents a single cell system, with one BS. The dashed hexagonal regions ("cells") within the original cell area are explained in

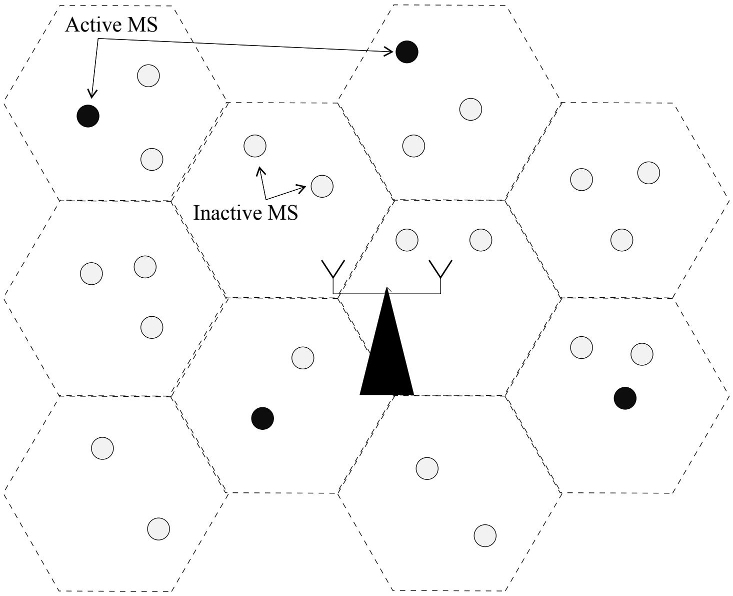


Figure 2.1: A single BS serving multiple active MSs with the assistance of idle MSs acting as relays.

Section 2.2.2 in a discussion of clustering, where subnetworks of relays support active MSs.

In an ideal case, an active MS chooses the $M - 1$ nearest idle MSs to serve as relay nodes, making the selection so as to minimize the required transmit power of the relay nodes. This also minimizes interference at undesired communication nodes, as shown in Section 2.2.2.

It is assumed that each receiving side (consisting of the relay nodes and the active MSs) knows its receiving instantaneous channel state information (CSI). Global availability of the CSI is not required, but every relay node must forward its CSI to the assisted receiver. Forwarding the CSI to the assisted receiver is an essential procedure that enables the receiver to decode the desired messages successfully. This step could be implemented within control pilots or frames between the receiver and its associated relays.

It is assumed that every communication link experiences flat fading over a complete time slot. However, the channel conditions may vary from one time slot to another without affecting the success of the communication process. Moreover, the relay nodes adopt an AF approach with controlled delay, since the receivers are interested in the raw signals rather than an altered version of the signal.

In this scheme, messages from the BS to the active MSs are delivered via two different stages. One stage is allocated for the BS to broadcast its messages, and the other stage is used by the relays to forward the signals they receive to the desired MSs. The next two subsections present the procedures followed in stage one and stage two of the proposed communication process.

2.1.1 Transmitter Stage

The first stage is comprised of N time slots. The BS follows an orthogonal division protocol and allocates one time slot per active MS, as illustrated in Table 2.1. In time slot n, the BS sends M messages (through M transmit antennas) to the nth active MS. The nth MS and its assisting $M-1$ relays (MS^n) hear the transmitted signals at approximately the same time, but every receiving node receives a different (scaled) version of these signals. The variation in the received signals is due to variations of the CSI between each receiving node and the BS antennas. The signals received at

Table 2.1: Stage one: Time slots allocated by the BS for every MS and its assisting relays

Time slot	1	2	.	.	N
Served MS	MS^1	MS^2	.	.	MS^N

all receiving nodes of group n can be expressed as follows:

$$Y_l^n(n) = \sum_{m=1}^{M} \sqrt{P}\, h_{lm}^n(n)\, x_m^n(n) + w^n(n) \tag{2.1}$$

where $Y_l^n(n)$ is the signal received at node l of receiving group n for $0 \le l \le M - 1$ and $1 \le n \le N$. Here, $l = 0$ is the index of the active MS node and $l > 0$ indicates the assisting relay nodes. Moreover, $\sqrt{P}$ is the average signal transmit power, and h_{lm}^n represents the channel gain factors between the BS m^{th} antenna and the l^{th} receiving node over the n^{th} time slot. The notation $x_m^n(n)$ represents the message intended to be transmitted from BS antenna m to the n^{th} MS, and $w^n(n)$ is the AWGN at the l^{th} receiving node during the n^{th} time slot. The AWGN is initially assumed to be relatively small, so that neglecting it would not affect the performance analysis of the system. In fact, the MAI in the second stage dominates the signal-to-interference-plus-noise ratio (SINR), and as a result limits the performance of the system. Thus, without loss of generality, the AWGN term will be omitted in the discussion below.

It should be noted that the n^{th} MS and its assisting relays are concerned only with the signal transmitted within their allocated time slot. They remain inactive over any time slot $j \neq n$ for $0 < j \le N$, and their received signals in such time slots are assumed to be zero. For brevity of notation, the time index can be omitted, since in stage one every group receives signals only in one time slot (numbered to match the index of the active node). Omitting the time index here also unifies the notation for the next stage. Thus, the equation representing the signal received at group n can be rewritten as:

$$Y_l^n = \sum_{m=1}^{M} \sqrt{P}\, h_{lm}^n\, x_m^n \tag{2.2}$$

Table 2.2: Stage two: Concurrent transmissions from relays to the desired MSs

Time slot	N+1	N+2	.	.	N+M-1
from → to	$MS_1^1 \rightarrow MS_0^1$	$MS_2^1 \rightarrow MS_0^1$	.	.	$MS_{M-1}^1 \rightarrow MS_0^1$
	$MS_1^2 \rightarrow MS_0^2$	$MS_2^2 \rightarrow MS_0^2$	.	.	$MS_{M-1}^2 \rightarrow MS_0^2$
	.	.	.	.	.
	.	.	.	.	.
	$MS_1^N \rightarrow MS_0^N$	$MS_2^N \rightarrow MS_0^N$	.	.	$MS_{M-1}^N \rightarrow MS_0^N$

2.1.2 Relay Stage

The BS turns off its transmission after sending to the last receiver in the N^{th} time slot. The second stage then begins by splitting the entire network into N subnetworks. All of these subnetworks work concurrently and independently, using the same frequency band over a period of $M-1$ time slots, as illustrated in Table 2.2. Every subgroup consists of a single receiver and $M-1$ transmitters (the relay nodes), which transmit the signals received in the first stage in sequence.

The signal received at MS_0^n over time slot $N+l$ becomes:

$$Z_l^n = d_l^n h_{0l}^n Y_l^n + \sum_j I_{jl}^n \tag{2.3}$$

where d_l^n is the amplifying gain at the relay, and h_{0l}^n is the channel gain between MS_l^n and MS_0^n. The MAI is represented by $\sum_j I_{jl}^n$ resulting from the interference signals arriving at MS_0^n from the relays of other active receivers. In order to control the impact of interference, a clustering method is used, as explained in Section 2.2.2.

Let $a_{lm}^n = d_l^n h_{0l}^n \sqrt{P} h_{lm}^n$. Then the M signals received in stage two at MS_0^n by the end of the $N + M - 1^{th}$ time slot can be represented as:

$$\begin{aligned} Z_1^n &= \sum_{m=1}^{M} a_{1m}^n \, x_m^n + \sum_j I_{j1}^n \\ &\vdots \\ Z_{M-1}^n &= \sum_{m=1}^{M} a_{M-1m}^n \, x_m^n + \sum_j I_{jM-1}^n \end{aligned} \tag{2.4}$$

In stage one, based on (2.2), MS_0^n has also received a signal directly from the BS, which can be expressed as:

$$Z_0^n = Y_0^n = \sum_{m=1}^{M} a_{0m}^n \, x_m^n \tag{2.5}$$

and $a_{0m}^n = \sqrt{P}\, h_{lm}^n$. Thus, each of the active receivers, MS_0^n, where $n \in \{1, \cdots, N\}$, has to solve a system of M equations and M unknowns as shown by (2.4) and (2.5) in order to decode the desired M messages x_m^n where $m \in \{1, \cdots, M\}$. Hence, the total number of messages communicated over $N + M - 1$ time slots to all N active MSs is equal to MN.

Although the focus here is on downlink transmissions, it is important to note that the communication procedure discussed could be made reversible by mirroring the time axis, in order to make it valid on the uplink. The first stage of reversed communications starts when the single-antenna MSs share the available resources for $M - 1$ time slots. Every MS sends a single message to an assigned relay over a time slot, and by the end of $M - 1$ time slots, each MS with its relays has simulated a single transmitter with M antennas and M independent messages (a virtual antenna array). However, this process is not optimal with regard to the degree of freedom (DoF). The maximum DoF it can achieve is $MN/(M + N - 1)$ which is smaller than M, the achievable DoF for multiple single-antenna transmitters and a single M antenna receiver. Moreover, the reverse process is less advantageous than the original process in terms of relay transmit power. In the proposed downlink scenario, the distances between the receivers and the relays are very short, resulting in a low relay transmit power requirement. In the reverse process, relays must communicate with the BS, which is not necessarily close to them, and thus greater transmit power is required.

2.2 Relay Interference Management

This section examines in more detail the operation of the transmission scheme developed in the previous section. Specifically, it considers the effects of relay interference for undesired users, and proposes solutions to manage its impact. Section 2.2.1 first presents an example of a network with two users, in order to demonstrate the scheme in limited interference conditions. This example represents a network with a minimum number of antennas at the transmitter, a minimum number of users, and a minimum number of assisting or cooperating users in the MISO network that is considered in this chapter. The purpose of this section is to provide an insight into the adopted model with regard to the distribution of nodes in the network and the transmission characteristics of each node.

Section 2.2.2 generalizes the scheme to accommodate more users, with an arbitrary number of antennas at the BS. This subsection utilizes the clustering concept and spatial reuse patterns from conventional cellular networks (without cooperating relays).

2.2.1 Two-User Broadcasting

This example considers a transmitter with two antennas, and two single-antenna receivers. If the two receivers are separated by a sufficient distance and each receiver has a single-antenna assisting relay, the achievable multiplexing gain in the proposed system is 4/3. The maximum multiplexing gain that can be achieved in a conventional system without relay assistance is equal to one, as determined by the single receiver

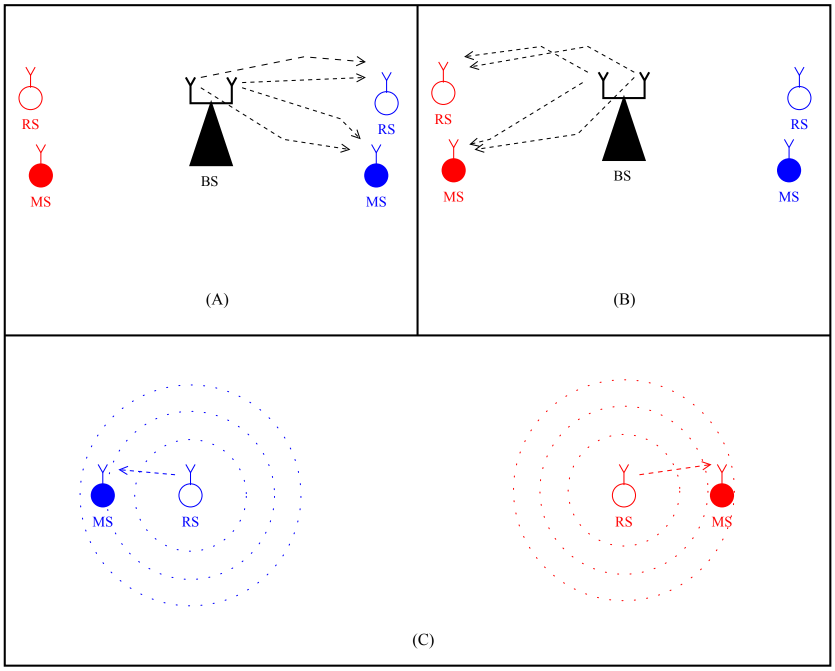


Figure 2.2: Time scheduling of transmissions in a two-user MISO network: (A) In the first time slot, the BS sends a signal received by the first MS and its assisting relay; (B) in the second time slot, the BS sends a signal received by the second MS and its assisting relay; and (C) in the third time slot, the cooperative relays retransmit their signals, destined for the corresponding MSs, potentially causing MAI.

antenna. Figure 2.2 shows the transmission process in the schemes developed for two active MSs over three time slots. The transmissions in these three time slots are represented in the corresponding subfigures (A), (B), and (C). In Fig.2.2 the first MS, indicated in blue, is shown positioned to the right of the BS; and the second MS, depicted in red, is positioned to the left of the BS. Here it is also assumed that these MSs are close to their relays (cooperating MSs), which are labeled RS and are colored to correspond to the colors used for the MSs. In Fig. 2.2(C), concentric circles show the range of transmissions from the relays, that use omnidirectional antennas. When the two MSs with their corresponding (assisting) relays are located at opposite sides of the BS, it is probable that the assisting relays will also be well separated in space. This means that the level of interference affecting non-associated MSs caused by any RS can be expected to be very low and hence negligible. It should be observed that, in general, the active MSs may be distributed anywhere in the plane, and the eventual objective is to select relays that are separated from the MSs for which they do not retransmit. This is elaborated further in Section 2.2.2.

In accordance with the assumption in Section 2.1, both MSs know their downlink CSI with regard to the BS, and the RSs forward their downlink CSI to the associated MSs. Let MS_0^1 and MS_1^1 represent the receiver and relay, respectively, located to the right of the BS; and let MS_0^2 and MS_1^2 represent the MS and relay, respectively, that are located to the left of the BS. The signals received in the first time slot at MS_0^1 and MS_1^1 can be expressed as:

$$\begin{aligned} Y_0^1 &= \sqrt{P}\, h_{01}^1\, x_1^1 + \sqrt{P}\, h_{02}^1\, x_2^1 \\ Y_1^1 &= \sqrt{P}\, h_{11}^1\, x_1^1 + \sqrt{P}\, h_{12}^1\, x_2^1 \end{aligned} \tag{2.6}$$

This group remains idle and does not receive in the next time slot, and the BS sends messages intended for MS_0^2. The signals received at MS_0^2 and MS_1^2 can be expressed as:

$$\begin{aligned} Y_0^2 &= \sqrt{P}\, h_{01}^2\, x_1^2 + \sqrt{P}\, h_{02}^2\, x_2^2 \\ Y_1^2 &= \sqrt{P}\, h_{11}^2\, x_1^2 + \sqrt{P}\, h_{12}^1\, x_2^2 \end{aligned} \tag{2.7}$$

In the third time slot, the transmitter remains silent, and the relays forward their messages simultaneously. The effects of the amplification factor and the channels between the relays and the receivers are first removed by the receivers. The signals received in the third time slot at each receiver are thus the same as the signals received at the relays in the previous two time slots (or a scaled version). The receivers can manipulate the two signals received, similarly to (2.6) with a scaled second equation for MS_0^1, and can solve for the two desired messages (x_1^1 and x_2^1 for MS_0^1) provided that the virtual channel matrix is invertible. Note that in this example it is assumed that the AWGN is small and does not affect the performance of the receivers. Moreover, it is assumed that both relays are transmitting with reduced transmit power, since the distances to the desired receivers are short. The transmit power of the relays can thus be expected to dissipate to a very low received power level by the time their signals causing MAI or relay interference arrive at undesired receivers. The impact of MAI can therefore be disregarded in this example.

2.2.2 Clustering

In the first stage of the proposed scheme, the BS allocates one time slot per active MS. This strategy does not cause interference to unintended receivers. However, in the second stage, resources are shared and each active MS forms an independent communication group with its relays. Communications within all of the independent groups occur concurrently, and controlling interference becomes a necessity. The key to ensuring that a desired received signal exceeds the power of interfering signals by a certain threshold is either to transmit with a power greater than that of interfering signals, or to position the receiver and its transmitter so that they are far from the interfering nodes and close to one another, as in the example considered in Subsection 2.2.1.

Here the option of transmitting with a higher transmit power is not productive. If a group increases its transmit power to achieve a certain signal-to-interference ratio (SIR), the SIRs of other groups will be degraded, and as a result they will also try to increase their transmit power. This selfish approach would create harmful competition, and interference would not be controlled even if all of the relays could transmit with unlimited power.

Thus, in order to accommodate larger numbers of users in the system while permitting all of them to maintain comparable, controlled SIRs, in this scheme the clustering concept of cellular systems has been adopted. Clustering provides groups of receivers that can be served in one communication cycle, and the receivers with their associated relays are well spaced so as to decrease the power of interference signals at each receiver. A single cell within a cellular system is considered, and the cell is divided into many hexagonal subcells. A reuse pattern is then utilized to space the MSs that are served in a single cycle. As in a cellular system, the reuse pattern is determined in accordance with acceptable interference levels, versus the number of users that can be served in a single cycle. However, in this scheme the subcells differ from a those of a cellular system in terms of the location of the receiver and relays. Receivers and relays can be positioned anywhere within the subcell.

The next section describes how the clustering concept is applied in a cell of fixed size with a specific distribution of MSs, and the required design parameters are discussed.

2.3 Result Analysis

A cell of 1 km^2 is considered that has approximately 400 communication nodes distributed in a deterministic, uniform fashion, as illustrated in Fig. 2.3, with the BS at the center. Any node can serve as a receiver, or as a relay, or can remain inactive. Furthermore, the service area (the main cell) is divided into C hexagonal subcells, determined according to the reuse pattern or the number of groups that are served per communication cycle. Division of the main cell serves to cluster MSs within the subcells.

Dividing the main cell into $C = 36$ subcells results in subcells that each have an area of 0.259 km^2, with sides measuring approximately 100 m. The deterministic arrangement of the communication nodes on a rectangular grid places an average of 9 MSs within each subcell. Thus, every active MS in a subcell will have 8 relays from which to choose, as a subset of MSs in a given subcell that can assist in the relaying process. The relays chosen may be those closest to the active MSs. A time

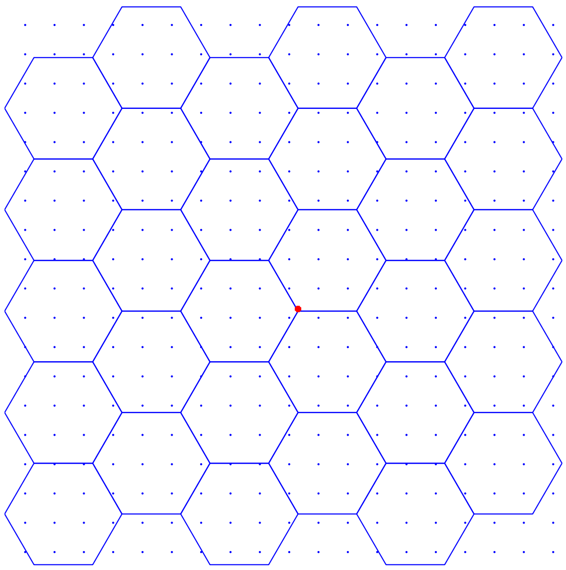


Figure 2.3: A deterministic spatial distribution of 400 MSs.

domain reuse pattern of 3 is considered, which allows the BS to serve 12 users in the service area per communication cycle. In this scheme, the reuse pattern determines the subcells where there is only one active MS, that will be served in the same communication cycle as other active MSs that are located in clusters or subcells indicated by the same fill pattern or color. This is illustrated in Fig. 2.4. Thus, in order to serve 36 MSs in the service area, 3 communication cycles are used (a time reuse pattern of 3), with MSs in subcells indicated by the same fill pattern or color receiving data from the BS in the same cycle (in two phases: directly and via relays).

Next a scenario is considered where the BS is equipped with 4 antennas, and where each active MS thus needs to select three relays. If all nodes receive signals from the transmitters according to comparable deterministic signal strength or attenuation, the active MS (or the central controller) will select the three closest nodes within

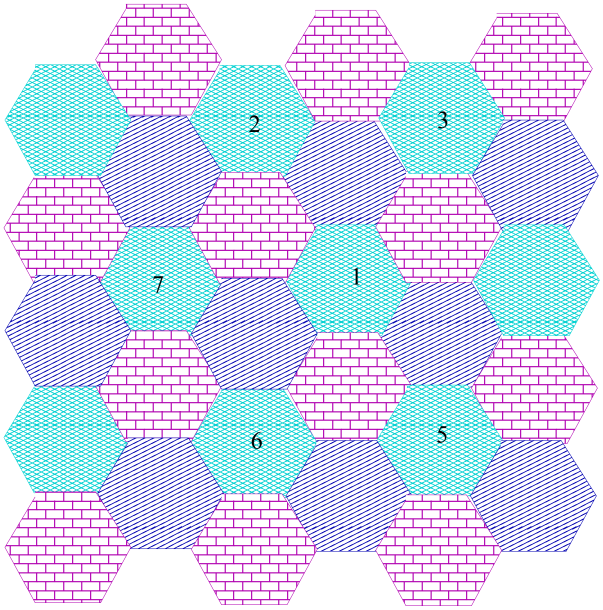


Figure 2.4: Spatial reuse contributing to MAI from relays in the second stage of a communication cycle, with a reuse pattern of 3 in the time domain.

the subcell for the relaying process. In this scenario, every active MS receives 4/15 messages per time slot and the entire system communicates $48/15 = 4 \cdot 12/(12 + 3)$ messages per time slot.

The main limitation in the proposed scheme is the approximately fixed average SIR at all receivers in the second stage. This limits the proposed scheme to a certain range of SINR. Increasing the transmit power of the relays beyond the optimal value does not improve the performance of any particular node. In fact, in this scheme a selfish approach results in higher power consumption with no further gain in signal quality. In the example analyzed, with a fixed spacing of nodes and assuming a reuse pattern of 3, when the deterministic path loss is proportional to d^{-4}, the average SIR in the second stage of the communication cycle is approximately 14 dB. This is because, in the case of the scenario illustrated in Fig. 2.4, the SIR of the transmission of interest between the relay and the MS in subcell 1 is affected primarily by the relays in the six subcells numbered 2 through 7, which are the closest subcells that

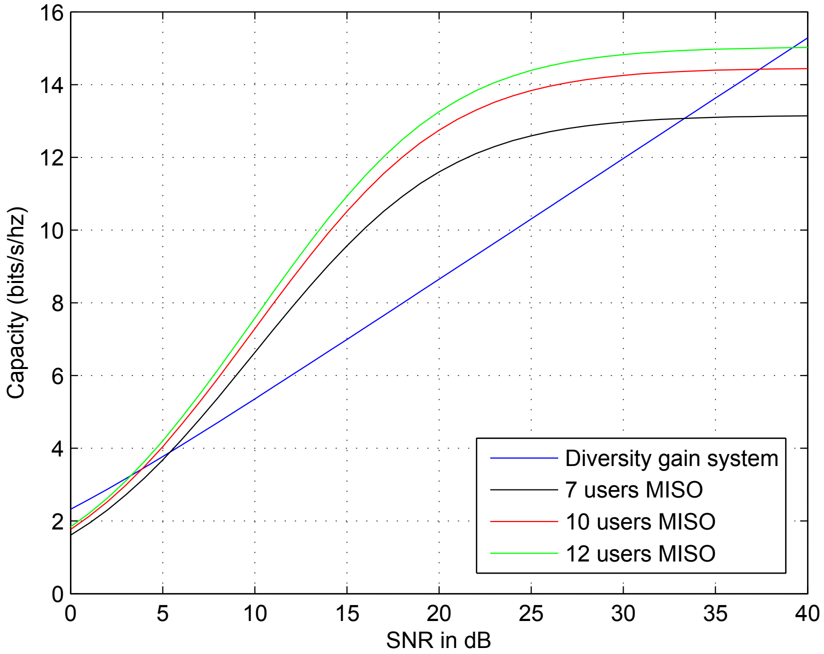


Figure 2.5: MISO network capacity comparisons, with a diversity gain system and the proposed scheme, where $M = 4$, the number of supported users N is a parameter, and the average SIR is 14 dB.

use the same time slot in the second stage of the communication cycle.

Figure 2.5 shows the impact of a fixed SIR of 14 dB on the capacity in the proposed scheme. It can be seen that with 4 antennas at the BS, the capacity improves as the number of active users increases from 7 to 12. The capacity begins to saturate at a SINR level of approximately 25 dB. The capacity for ordinary orthogonal multiple access, where the antennas at the BS are used for diversity gain, is better at a low SINR level, until the SINR is approximately equal to 5 dB. Above 5 dB, the proposed scheme outperforms the diversity gain system. For the proposed model, the capacity is saturated at SINR of about 33 dB to 38 dB, above which the orthogonal diversity gain system has a greater capacity than the proposed system. Thus, the relay-aided system has a superior performance in the range between 5 dB and 35 dB. Although the model considers deterministic positioning of the nodes, it should be recognized that in practice, one of the receivers may be located at the edge of its subcell and the relaying nodes of other receivers may be located at the edge nearest to this receiver. This would adversely affect the SINR, however, the probability of occurrence of such a situation is low.

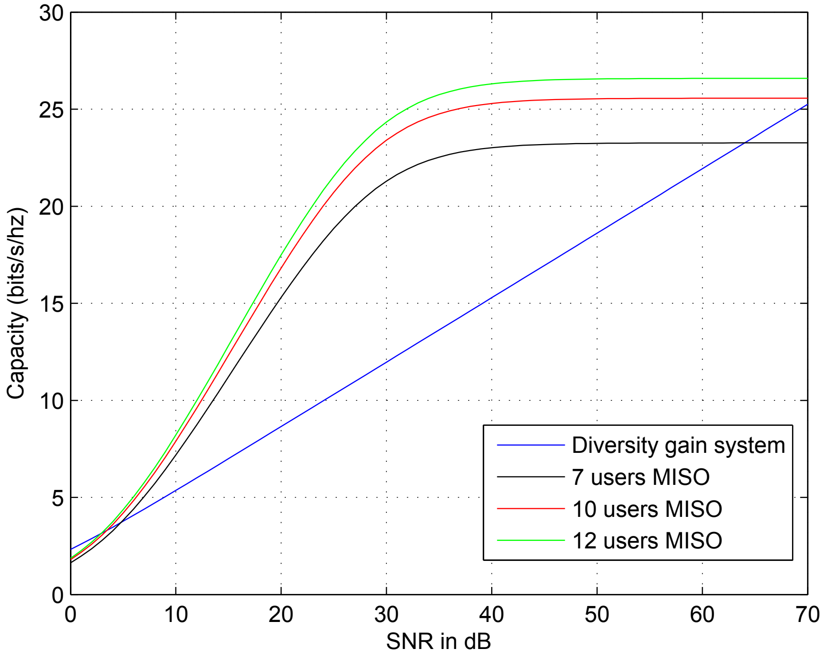


Figure 2.6: MISO network capacity comparisons, with a diversity gain system and the proposed scheme, where $M = 4$, the number of supported users N is a parameter, and the average SIR is 25 dB.

For a reuse pattern of 7 within the same 1 km^2 area, the same placement of nodes provides 5 active receivers per group for the communication cycle. Therefore, the (aggregate) data rate drops from 48/15 to $20/8 = 4 \cdot 5/(5+3)$. The advantage of employing a reuse pattern of 7 rather than 3 is a considerable enhancement in the SIR. This would be especially beneficial for a deterministic path loss with distance proportional to d^{-2}. This is because the distance between subcells that are served in the same cycle is greater, and the SIR increases to 25 dB with a path loss proportional to d^{-4}. Figure 2.6 shows that the saturation of capacity in this scenario would occur only at a very high SINR.

Here the proposed scheme outperforms the orthogonal multiple access system with diversity gain at the transmitter in the range between 5 dB and approximately 65 dB. Thus, the relay-aided single-antenna receiver in the proposed scheme provides a gain in capacity within a reasonable SINR range.

When the path loss is proportional to d^{-2}, a reuse pattern of 7 yields an average of 9 dB SIR, whereas a reuse pattern of 3 results in 3 dB SIR. This might not provide a favorable gain, and the use of schemes that avoid interference could be more desirable in this case.

2.4 Summary

This chapter presented a time scheduling scheme to increase the data rate on a downlink in a cooperative cellular network. Where zero forcing is not feasible at the transmitter due to a lack of necessary CSI at the BS, it is shown that cooperation among MSs can increase the throughput for single-antenna users from one to $MN/(M+N-1)$, where M is the number of transmit antennas at the BS and N is the number of active MSs.

The proposed transmission scheme is divided into two stages. In the first stage, the BS employs a TDMA-like mechanism to send signals to the active MSs. In this stage, the MSs acting as relays opportunistically overhear the data. In the second stage, the BS is turned off and the cooperating MSs (acting as AF relays) form independent one-to-one networks and forward their signals simultaneously over $M-1$ time slots to active MSs. Each MS receives signals from a BS with unique channel conditions. This information allows MSs to receive multiple copies of the desired signals, when cooperative MSs relay the signals to the active MSs. With a sufficient number of received signals, each receiver can construct a linear system of equations so that it can recover a total of M unique messages per communication cycle. In the proposed scheme, relay clusters communicating in the same time slot are spaced apart from each other in order to minimize the impact of MAI, an approach similar to that employed for spectral reuse in conventional cellular systems.

Chapter 3

Successive Transmission in Two-hop Relay Networks

Successive relaying enhances the capacity of wireless relay networks through simultaneous spectrum utilization by a source and relays. Nevertheless, IRI constitutes a major challenge for successive relaying schemes because the signals from the transmitting relay at a particular TS leak to the receiving relay. These IRI signals arrive at the receiving relay with a power level comparable to that of the desired signals transmitted by the source and they have to be dealt with. This is because in this chapter, we assume a network topology in which relays are separated by a relatively small distance.

In order for successive relaying to be implemented effectively, IRI needs to be controlled and eliminated. This chapter presents a finite impulse response (FIR) filtering type signal processing scheme at the destination for cancelling the effects of IRI in MIMO relay networks. In this dissertation, the proposed IRI cancellation for successive relaying is initially developed for a network with a single source, two relays and a single destination (referred to as $1 \times 2 \times 1$). Than the IRI cancellation is broadened to a more general type of networks like $1 \times 2 \times K$ where we have

K destinations receiving independent data streams via two relays from a single source.

Section 3.1 introduces the idea of successive relaying and its benefits for the network spectral efficiency as well as characterizes autoregressive (AR) type of modeling for the IRI which is unique to this thesis. Based on this modeling, we develop matching finite impulse response (FIR) filtering for the IRI removal at the destination. This section also presents the proposed scheme effects on the desired signals like noise enhancement. Section 3.2 generalizes the scheme to various networks, adding some insights while discussing the application of the original scheme to any type of two-hop networks. The performance of various schemes is then discussed in terms of bit error rate (BER) and capacity in Section 3.3. Section 3.4 provides a summary of the chapter.

3.1 Successive Relaying and Inter-Relay Interference

In this area, most studies in the literature focus on what is referred to as one-to-one communication with two relays in the middle using single antenna nodes. The system designated here as $1 \times 2 \times 1$ (Fig. 3.1) consists of a source S, two relays R_1 and R_2, and a destination D. The source sends signals to R_1 during the even time slots and to R_2 during the odd time slots. While the source transmits continuously, R_1 forwards its received signals to the destination during the odd time slot, and R_2 forwards its received signals during the even time slot, using the same spectrum that is utilised by the source. It is worth observing that because relays operate in half duplex mode, the reception by R_1 during the even time slots (TSs) from the source is affected by signals from both: the source and R_2, while in the odd TSs, R_1 can not overhear the signal from the source that is destined to R_2.

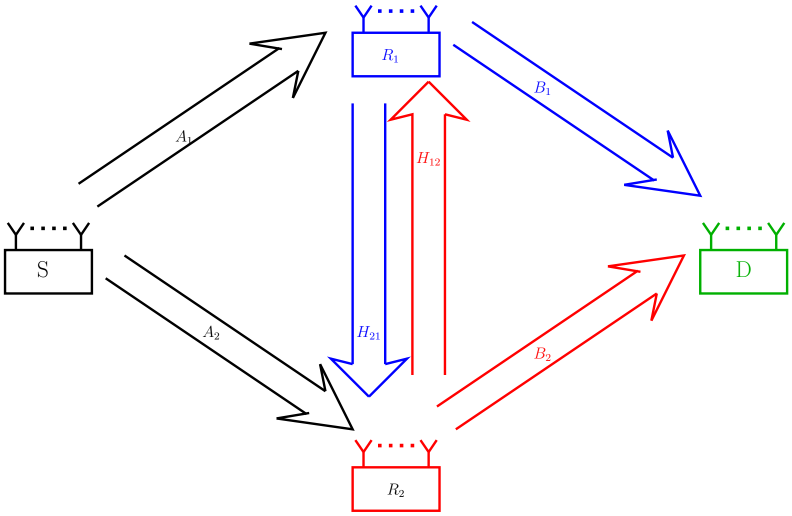


Figure 3.1: A $1 \times 2 \times 1$ relay network.

First, using the $1 \times 2 \times 1$ MIMO network model, the IRI characterization is developed in Section 3.1.1, and than in Section 3.1.2, the proposed IRI cancellation scheme is presented at the single destination continuously receiving the data. Later on in Section 3.2, the prototype scheme developed originally for the $1 \times 2 \times 1$ MIMO network model is generalized using the idea of opportunistic listening to multiple networks. This is to reflect the flexibility of the proposed scheme, which makes it suitable for a wide range of applications.

3.1.1 A $1 \times 2 \times 1$ Network

In a $1 \times 2 \times 1$ layout, the assumption is that the relays are located close to one another, since they are assisting communications from the same source to the same destination. It is also assumed that signals from the source arrive at the destination exclusively through the relays and that the direct path between the source and the destination has a negligible channel gain. The IRI in this network starts at $TS = 1$, when S (the node shown in black in Fig. 3.1) sends a new set of signals to R_2 (shown in red), while R_1 (shown in blue) forwards the set of signals it received during $TS = 0$

to D (shown in green). This means that R_2 receives a set of desired signals from S during $TS = 1$ through channel $\mathbf{A_2}$, however, this set of signals is accompanied by a set of IRI signals transmitted by R_1, which arrive at R_2 via channel $\mathbf{H_{21}}$. The problem of IRI continues, with the relays exchanging roles every time slot.

In the proposed scheme, it is assumed that every communication node is equipped with M antennas. Here the channels given by bold letters, e.g., $\mathbf{A_2}$, should be interpreted as channel gain matrices with different subscripting to differentiate between various links as shown in Fig. 3.1 and summarized in Table 3.1. In general, we use $\mathbf{A_i}, i \in \{1, 2\}$ to represent channels between the source and the relays (from S to R_i), while we use $\mathbf{B_i}$ to represent channels connecting the relays and destination (from R_i to D). Channel gain matrix $\mathbf{H_{ij}}$ stands for the inter-relay channel (from R_j to R_i). One entry in the channel gain matrix is a scalar Rayleigh (complex normal) fading gain from the transmit antenna (one of M) to the receive antenna (one of M). This is because we assume that the channels connecting the various nodes experience flat Rayleigh fading; however, this assumption does not affect the generality of the scheme, which could be used with any other type of flat fading. An AF relaying topology is adopted, and the relays amplify their signals with gains of $\mathbf{W_i}$.

Also, “a set of signals” (or signals for short) from a transmitting node should be interpreted as the column vector signal mapped on the the M transmitting antennas. The assumption of equal number of antennas (M) in all the nodes determines the number of independent data streams communicated per time slot. In addition, it is also important for both the relays and the destination to have an equal number of antennas to guarantee square matrices between the relays and the destination, since the proposed IRI cancellation scheme requires the inversion of these channel gain matrices.

During $TS = 0$, the communication process is initiated and S is the only node in the network that has signals to transmit. Therefore, the signals received at R_1

Table 3.1: Network parameter definitions

Symbol	Definition
S	The source
R_1	Relay number 1
R_2	Relay number 2
D	The destination
$\mathbf{A_1}$	The channel from S to R_1
$\mathbf{A_2}$	The channel from S to R_2
$\mathbf{H_{21}}$	The channel from R_1 to R_2
$\mathbf{H_{12}}$	The channel from R_2 to R_1
$\mathbf{B_1}$	The channel from R_1 to D
$\mathbf{B_2}$	The channel from R_2 to D
P	S signal power
$\mathbf{W_1}$	R_1 amplifying gain
$\mathbf{W_2}$	R_2 amplifying gain

during $TS = 0$ are not accompanied by any interference and are only multiplied ("attenuated") by $\mathbf{A_1}$ and corrupted by the relay AWGN noise $\mathbf{n_{R_1}}(0)$, which is assumed to have zero mean and unit variance. Hence the signals received at R_1 can be expressed as:

$$\mathbf{r_1}(0) = \sqrt{P}\mathbf{A_1}\mathbf{S}(0) + \mathbf{n_{R_1}}(0) \tag{3.1}$$

where the parameter P stands for transmit power and controls average SNR on a link in the system. During $TS = 1$, R_2 is responsible for listening to S; however, while S is transmitting signals during $TS = 1$, R_2 is also transmitting an amplified version of $\mathbf{r}(0)$. Therefore, the signals received at R_2 during $TS = 1$ are not only pure signals from S but also interference from the signals transmitted by R_1. The signals received at R_2 during $TS = 1$ can thus be expressed as:

$$\mathbf{r_2}(1) = \sqrt{P}\mathbf{A_2}\mathbf{S}(1) + \mathbf{H_{21}}\mathbf{W_1}\mathbf{r_1}(0) + \mathbf{n_{R_2}}(1) \tag{3.2}$$

The task of the relays in this chapter is only to amplify the received signals, whereas dealing with the interference is handled by the destination. Thus, the following signals are received at the destination during $TS = 1$ and $TS = 2$:

$$\mathbf{y}(1) = \sqrt{P}\mathbf{B_1 W_1 A_1 S}(0) + \mathbf{B_1 W_1 n_{R_1}}(0) + \mathbf{n_D}(1) \tag{3.3}$$

$$\mathbf{y}(2) = \sqrt{P}\mathbf{B_2 W_2 A_2 S}(1) + \mathbf{B_2 W_2 H_{21} W_1 r_1}(0) + \mathbf{B_2 W_2 n_{R_2}}(1) + \mathbf{n_D}(2) \tag{3.4}$$

where $\mathbf{n_D}(l)$ is AWGN at the destination in TS l.
Similarly, during $TS = 2$ R_1 receives signals from S as well as interference from R_2:

$$\mathbf{r_1}(2) = \sqrt{P}\mathbf{A_1 S}(2) + \mathbf{H_{12} W_2 r_2}(1) + \mathbf{n_{R_1}}(2) \tag{3.5}$$

and when R_1 forwards these signals to D during $TS = 3$, they arrive as:

$$\mathbf{y}(3) = \sqrt{P}\mathbf{B_1 W_1 A_1 S}(2) + \mathbf{B_1 W_1 H_{12} W_2 r_2}(1) + \mathbf{B_1 W_1 n_{R_1}}(2) + \mathbf{n_D}(3) \tag{3.6}$$

The process continues, and with every time slot R_1 and R_2 exchange the roles of receiving and forwarding signals. Thus (3.2) and (3.5) can be generalized as:

$$\mathbf{r_i}(t) = \sqrt{P}\mathbf{A_i S}(t) + \mathbf{H_{ij} W_j r_j}(t-1) + \mathbf{n_{R_i}}(t) \tag{3.7}$$

where $i = 1$ and $j = 2$ if the TS is even, and $i = 2$ and $j = 1$ if TS is odd. Similarly, the signals received at the destination can be expressed in a general form as:

$$\mathbf{y}(t+1) = \sqrt{P}\mathbf{B_i W_i A_i S}(t) + \mathbf{B_i W_i H_{ij} W_j r_j}(t-2) + \mathbf{B_i W_i n_{R_i}}(t) + \mathbf{n_D}(t+1) \tag{3.8}$$

The signals received at the destination are composed of desired signals, interference terms, amplified AWGN from the relays, and AWGN from the destination.

Before we proceed with the development of the IRI cancellation scheme in Section 3.1.2 we have the following observations in order which will help to intuitively understand the nature of the IRI, its modeling and the corresponding cancellation process.

Transfer Function Interpretations

First, it should be understood that our signals like $\mathbf{r_i}(t)$ are column vectors and, in this dissertation, most of the time, we work with matrix multiplications. However, for more tangible explanations, we adopt till the end of this subsection the scalar signals to capture the essence of various relations. Also, to avoid unnecessary complexity at this stage, when presenting (3.7) in the scalar form, we drop the subscripting and disregard AWGN and the irrelevant signal scaling. With these great simplifications, the equivalent "received relay signal" in (3.7) is represented as:

$$r(t) = s(t) + \alpha \cdot r(t-1) \tag{3.9}$$

or equivalently in the z-domain

$$R(z) = S(z) + \alpha \cdot R(z) \cdot z^{-1} \tag{3.10}$$

and

$$\frac{R(z)}{S(z)} = \frac{1}{1 - \alpha \cdot z^{-1}} \tag{3.11}$$

where $R(z)$ and $S(z)$ are the z-transforms of $r(t)$ and $s(t)$, respectively. With (3.9), the "received signal" at the relay is related to the "information bearing signal" s(t) through the auto-regressive or equivalently recursive model [50]. We can also say, based on (3.11), that $r(t)$ was obtained from $s(t)$ by passing $s(t)$ through the infinite impulse response (IIR) filter with the transfer function:

$$\frac{1}{1 - \alpha \cdot z^{-1}} = \sum_{i=0}^{\infty} \alpha^i z^{-i} \tag{3.12}$$

Eventually, the relay signal with some amplification and after being corrupted by the AWGN is received at the destination. For brevity, we assume that the received signal at the destination is the same as at the relay (after all, relays operating in half duplex mode do not interfere with the reception at the destination). Then from the basic signal analysis, to recover the “information bearing signal” s(t), we should use the finite impulse response (FIR) filter with the transfer function $1 - \alpha \cdot z^{-1}$ at the output of our system to recover signal of interest. Autoregressive modeling of the received signal (capturing simplified effects –in the scalar form– of IRI) in terms of the “information bearing signal” is shown in Fig. 3.2a. Figure 3.2b shows the FIR filter to recover signal of interest or equivalently removing IRI effects. If the FIR filter is placed at the output of the autoregressive model accounting for the IRI effects, we would have the IRI cancellation at the receiver. If the FIR filter is placed at the input of the autoregressive model, we would have the IRI cancellation at the transmitter. From the filter theory point of view, we can interchange the positions of different filters in a cascade of filters. However, we have not accounted here for the noise effects. In our systems, we have two type of AWGN effects: at the relays and at the destination. If we disregard the noise at the relays, by inverting the transfer function of a system at the destination, which is also referred to in the literature as the zero-forcing (ZF) post-processing, we would have to deal with noise enhancement effects, i.e., passing the destination AWGN through the recovery FIR filter. So theoretically, placing the FIR filter at the source in the anticipation of the IRI could give better results. However, we also have to deal with the relay AWGNs and the impact of this noise because of distributed character of two relays is more difficult to explain in our simplified model.

To summarize, the purpose of the significant simplifications of matrix relations in successive relaying in the MIMO $1 \times 2 \times 1$ system, was to provide an intuition for autoregressive interpretation of IRI. With this, we also indicated that the IRI cancellation can follow FIR type filtering, equivalent to deploying ZF algorithms, either at the destination as developed in this chapter or at the source as developed in Chapter 4. The intuition developed for the IRI effects and its removal based on the transfer function interpretation is one of the contributions in this dissertation.

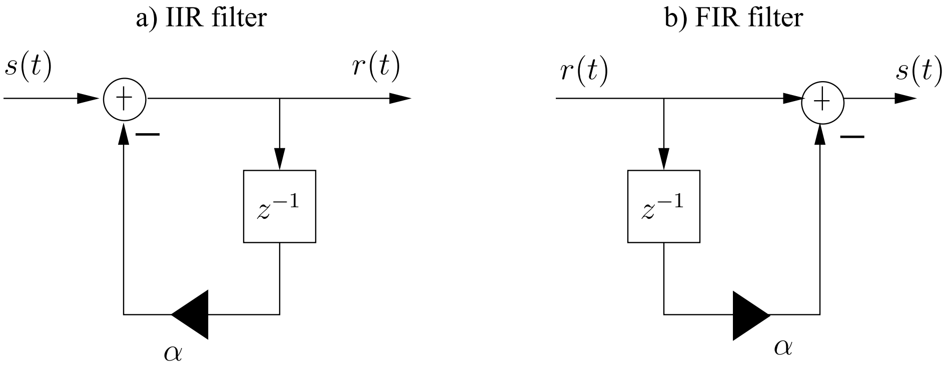


Figure 3.2: Simplified IIR and FIR filter structures to model scalar IRI and the cancellation process.

In automatic control [51], there are z-transform transfer function interpretations for matrix type of signal representations as in (3.7). However, we felt that explaining these relations using state space representations as what follows in Section 3.1.2 is more amenable for further analysis.

3.1.2 Full IRI Cancellation in a $1 \times 2 \times 1$ Network

If a set of signals forwarded by R_i during $TS = t$ is traced, it can be seen that two versions of these signals are delivered to the destination during two consecutive time slots, $TS = t$ and $TS = t + 1$. The first version follows the black path shown in Fig. 3.3. The signals in this version are attenuated by the channel between the forwarding relay and the destination, and arrive at the destination at $TS = t$. The other version arrives at the destination at $TS = t + 1$, following the blue path shown in Fig. 3.3. On the blue path, the signals are attenuated by the channel gain between the two relays, and are then amplified by the other relay and attenuated again by the channel gain between the destination and the other relay.

In mathematical terms, the two consecutive signals that arrive at the destination

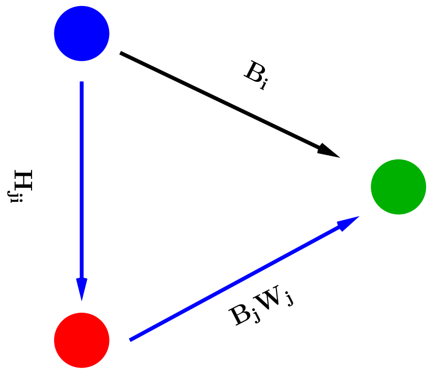


Figure 3.3: The IRI effects at the destination as a result of two path reception along:(i) the direct path and (ii) through the other relay path.

can be expressed as:

$$\mathbf{y}(t+1) = \mathbf{B_i W_i r_i}(t) + \mathbf{n_D}(t+1) \tag{3.13}$$

$$\mathbf{y}(t+2) = \sqrt{P}\mathbf{B_j W_j A_j S}(t+1) + \mathbf{B_j W_j H_{ji} W_i r_i}(t) + \mathbf{B_j W_j n_{R_j}}(t) + \mathbf{n_D}(t+2) \tag{3.14}$$

From the perspective of the destination, the set of signals in (3.13) contains no desired signals during $TS = t+2$, however it can be used to filter out the IRI signals from the desired signals in $\mathbf{y}(t+2)$.

Thus, the second term in (3.14) can be matched through algebraic manipulation to (3.13):

$$\mathbf{B_j W_j H_{ji} B_i}^{-1}\mathbf{y}(t+1) = \mathbf{B_j W_j H_{ji} W_i r_i}(t) + \mathbf{B_j W_j H_{ji} B_i}^{-1}\mathbf{n_D}(t+1) \tag{3.15}$$

If (3.15) is subtracted from (3.14), the received signal at D during $TS = t+2$ becomes:

$$\hat{\mathbf{y}}(t+2) = \sqrt{P}\mathbf{B_j W_j A_j S}(t+1) + \mathbf{B_j W_j n_{R_j}} - \mathbf{B_j W_j H_{ji} B_i}^{-1}\mathbf{n_D}(t+1) + \mathbf{n_D}(t+2) \tag{3.16}$$

Thus, the receiver needs to use signals from two consecutive time slots in order to filter out the IRI which accompanies the desired signals. In other words, the receiver makes two copies of the signals received from R_i at $TS = t$, as illustrated in Fig. 3.4.

The first copy is multiplied by $\mathbf{B_j W_j H_{ji} B_i}^{-1}$ and delayed in order to use it as an IRI filter for the next signals (see the lower line of Fig. 3.4), while the other copy (indicated by the upper line in Fig. 3.4) is passed into the filter to remove the IRI so that the signals will be ready for further manipulations. The filter structure in Fig. 3.4) is the matrix version of the FIR filter for the scalar signals in Fig. 3.2 where we presented simplified interpretation for the cancellation of the IRI effects.

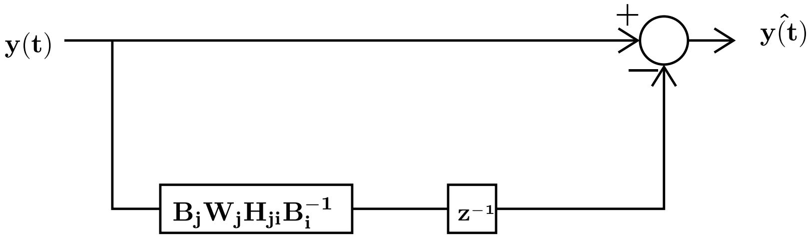


Figure 3.4: Filter to remove IRI.

An important aspect of the IRI filter developed in this section is that it depends only on two consecutive time slots, which implies that the receiver does not have to store any signal for more than one time slot. Another implication, together with the concept of opportunistic listening which will be introduced in the next section, is that more than one receiver can be served without wasting time slots for flushing purposes. Therefore, the scheme developed in this section can be generalized to more flexible systems, as discussed in the next section.

3.2 Opportunistic Listening and a Generalized IRI Filter

The IRI cancellation scheme developed in Section 3.1 provides a level of flexibility that makes it suitable not only for 1 × 2 × 1 networks, but also for more generalized networks. One reason for this is that the scheme depends only on the current time slot (which is the desired time slot) and the previous time slot, which is used to filter out the IRI. Thus, any receiver that can receive signals over two consecutive time slots can queue to be served by the relays. The features of wireless communications that cause IRI are also those that make it possible for receivers to hand over the service, allowing a new receiver to benefit from the scheme without wasting time initiating a new connection. When relay R_i in Fig. 3.5 retransmits the signals it received in the previous time slot, it broadcasts them in an omnidirectional fashion. This causes IRI at R_j, and also allows any receiver in the range of the relay to listen to the signals, e.g., D_j in Fig. 3.5. These signals are intended for another receiver, D_i, and are not of particular interest to D_j. However, D_j listens to these signals not for its own data decoding purposes, but rather in order to use them to remove the interference from the next signals that will be sent from R_j to D_j. Figure 3.5 visualizes only the two relay transmission stage and does not account where the source data are coming from.

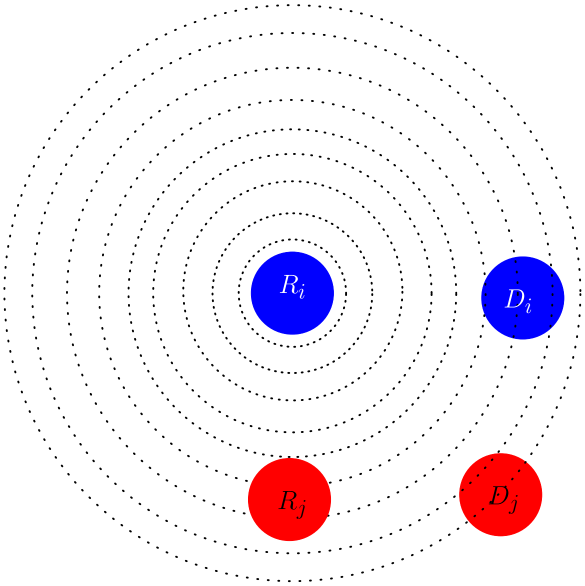


Figure 3.5: Signals causing interference at R_j used by D_j to remove the interference.

In addition, the IRI cancellation scheme from Section 3.1.2 is not coupled to particular relays. Even though the number of relays must be at least two in order to maintain successive relaying, the utilization of two particular relays and their continued use is not essential to the scheme. It should be noted that the relays are not directly involved in processing the signals to cancel the IRI. Therefore, a new relay could be utilized in the successive relaying developed in the previous section.

Finally, there is no restriction on the number of sources that can take turns benefiting from this scheme. In fact, with this scenario, the source is the communication node which is not involved in any processing to facilitate the interference-free communication process, in the sense of the required post-processing at the destination, or even the sharing of any information. All the source does is to send its signals to the desired relay. In this model, the relays have to know the receiving CSI channel gains matrices $\mathbf{A}$s in order to convey them to the destination. However, there are no restrictions on the sources except for transmitting their data signals to the destination (and may be pilot tones for channel estimation at the relays).

In the following subsections, examples of networks that can benefit from the IRI cancellation scheme developed in 3.1.2 are presented. These examples are of interest not only for demonstrating the flexibility of the proposed scheme, but also for their practical applications.

3.2.1 Proposed IRI Cancellation in a $2 \times 2 \times 2$ Network

Networks with two sources and two destinations are extensively studied by researchers. Whether each source sends messages to both destinations, or each source is associated with only a single destination, at least two relays are required to maintain successive relaying. Thus, a two-source-two-destination network with two relays forms a $2\times2\times2$ network.

Figure 3.6 illustrates a $2 \times 2 \times 2$ network which consists of two one-to-one relay networks. In the figure, the first one-to-one relay network is depicted in black, and the second one-to-one relay network with the blue nodes represent a similar scenario. However, the network shown in black and that shown in blue must share channel resources and use the same spectrum. They could act conservatively and not allow two nodes (sources and relays) to transmit at the same time, and hence accept a prelog reduction factor in the data rate due to relaying extra time. On the other hand, they could cooperate and transmit back-to-back (one source at the time and one relay at the time), utilizing an IRI cancellation scheme.

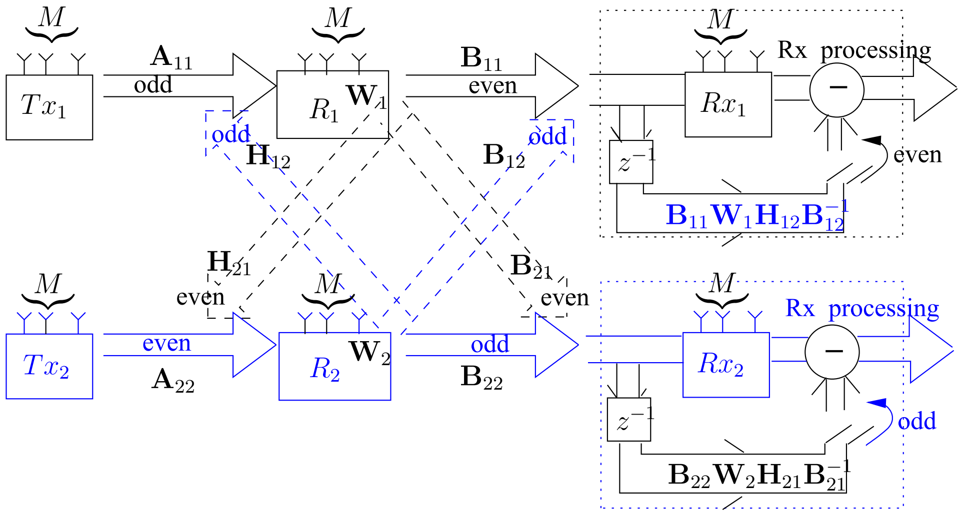


Figure 3.6: IRI cancellation in an AF MIMO 2 × 2 × 2 network.

Building on the 1 × 2 × 1 configuration, the 2 × 2 × 2 network splits the single source into two independent sources, where both sources have the same capability as the single source in the 1×2×1 network. The restrictions on the sources are loose, and as long as they transmit back-to-back or alternately, the two sources will not affect the IRI scheme. In this scenario, the difference is in the assignment of channels between the sources and the relays. As illustrated in Fig. 3.6, the channel between the black source and the black relay is labelled $\mathbf{A_{11}}$ and the channel between the blue source and the blue relay is labelled $\mathbf{A_{22}}$. The configuration shown in Fig. 3.6 represents a simple network which assumes that each source is associated with one destination and that no cross communication is performed. This simple representation helps to introduce the implementation of the proposed scheme in a 2×2×2 network in a clear

Table 3.2: $2 \times 2 \times 2$ network parameter definitions

Symbol	Definition
$\mathbf{A_{11}}$	The channel from S_1 to R_1
$\mathbf{A_{21}}$	The channel from S_1 to R_2
$\mathbf{A_{12}}$	The channel from S_1 to R_2
$\mathbf{A_{22}}$	The channel from S_2 to R_2
$\mathbf{B_{11}}$	The channel from R_1 to D_1
$\mathbf{B_{21}}$	The channel from R_1 to D_2
$\mathbf{B_{12}}$	The channel from R_2 to D_1
$\mathbf{B_{22}}$	The channel from R_2 to D_2
P_1	S_1 signal power
P_2	S_2 signal power

way. After the application of the proposed scheme to a simple $2 \times 2 \times 2$ network is demonstrated, it will be applied to more complex scenarios. Thus, although the cross channels between the black source and the blue relay and vice versa do not participate in the network configuration shown in Fig. 3.6, they are involved in the more complex scenarios. The cross channel between the black source and the blue relay is therefore labelled $\mathbf{A_{21}}$, and the channel between the blue source and the black relay is labelled $\mathbf{A_{12}}$.

In the second hop of the $2 \times 2 \times 2$ network, the single destination is likewise replaced by two destinations with capabilities identical to those of the destination in the $1 \times 2 \times 1$ network. The primary modification that affects the IRI cancellation scheme occurs in this part of the network. In the $1 \times 2 \times 1$ network, the single destination listens continuously to the signals retransmitted by both relays, since both relays are delivering desired signals to this destination. The two destinations in the $2 \times 2 \times 2$ network shown in Fig. 3.6 also have to listen to both relays continuously.

However, in this scenario the purpose of listening differs from the situation with the single destination in the 1 × 2 × 1 network. The signals transmitted by the blue relay shown in Fig. 3.6 are not intended for the black destination. However, the signals arrive at the black destination, which in fact requires them in order to eliminate the IRI effects in the next time slot. Thus, the black destination makes use of opportunistic listening to store non-intended signals, so as to utilize them for IRI cancellation in the next time slot.

Splitting the destination into two destinations creates four new channels between the relaying nodes and the destination nodes. As shown in Fig. 3.6, the channel between the black relay and the black destination is labelled $\mathbf{B_{11}}$ and the channel between the blue relay and the blue destination is labelled $\mathbf{B_{22}}$. Similarly, the cross channel between the black relay and the blue destination is labelled $\mathbf{B_{21}}$ and the cross channel between the blue relay and the black destination is labelled $\mathbf{B_{12}}$. A summary of the new channels in the 2 × 2 × 2 network is presented in Table 3.2.

The modifications to the network introduce minor changes to the received signals in (3.16). However, these changes are in the channel assignments and possibly in the quality of the cross channels, and are not in the core components of the IRI cancellation scheme. In other words, the receiver uses signals that arrive at $TS = t$ through cross channels, utilizing the concept of opportunistic listening in order to remove the IRI from the desired signals that are received along the desired path at $TS = t + 1$. In accordance with the model shown in Fig. 3.6, the signals received by the black destination during an odd time slot $TS = t_o$ are transmitted by the blue relay. These signals are intended for the blue destination, but through opportunistic listening the black destination receives:

$$\mathbf{y_1}(t_o) = \mathbf{B_{12}}\mathbf{W_2}\mathbf{r_2}(t_o - 1) + \mathbf{n_D}(t_o) \tag{3.17}$$

The black destination utilizes the signals in (3.17) to deal with the interference accompanying the desired signals along the black path at $TS = t_o + 1$:

$$\begin{aligned}\mathbf{y_1}(t_o{+}1) = {} & \sqrt{P_1}\mathbf{B_{11}}\mathbf{W_1}\mathbf{A_{11}}\mathbf{S}(t_o) \\ & +\mathbf{B_{11}}\mathbf{W_1}\mathbf{H_{12}}\mathbf{W_2}\mathbf{r_2}(t_o{-}1)+\mathbf{B_{11}}\mathbf{W_1}\mathbf{n_{R_1}}(t_o)+\mathbf{n_D}(t_o{+}1)\end{aligned} \tag{3.18}$$

Therefore, destination D_1 multiplies (3.17) by $\mathbf{B_{11}}\mathbf{W_1}\mathbf{H_{12}}\mathbf{B_{12}^{-1}}$ and subtracts it from (3.18), yielding:

$$\begin{aligned}\hat{\mathbf{y_1}}(t_e) = {} & \sqrt{P_1}\mathbf{B_{11}}\mathbf{W_1}\mathbf{A_{11}}\mathbf{S}(t_o) \\ & + \mathbf{B_{11}}\mathbf{W_1}\mathbf{n_{R_1}}(t_o)-\mathbf{B_{11}}\mathbf{W_1}\mathbf{H_{21}}\mathbf{B_{12}^{-1}}\mathbf{n_D}(t_o)+\mathbf{n_D}(t_e)\end{aligned} \tag{3.19}$$

in which $t_e = t_o + 1$. Similarly, by following the same procedure, destination D_2 can manipulate two consecutive time slots to obtain:

$$\begin{aligned}\hat{\mathbf{y_2}}(t_o) = {} & \sqrt{P_2}\mathbf{B_{22}}\mathbf{W_2}\mathbf{A_{22}}\mathbf{S}(t_e) \\ & +\mathbf{B_{22}}\mathbf{W_2}\mathbf{n_{R_2}}(t_e)-\mathbf{B_{22}}\mathbf{W_2}\mathbf{H_{12}}\mathbf{B_{21}^{-1}}\mathbf{n_D}(t_e)+\mathbf{n_D}(t_o)\end{aligned} \tag{3.20}$$

in which $t_o = t_e + 1$.

In the configuration described above, it is assumed that each source is associated with a particular receiver, and hence that interaction between the blue path and the black path shown in Fig. 3.6 occurs only in the second hop. However, the system is sufficiently flexible to deliver messages from any source to any destination, forming a model that is a hybrid between the 1 × 2 × 1 and 2 × 2 × 2 models. Depending upon the flow of signals, the network takes the form of one of the four models summarized in Table 3.3.

The first two models in Table 3.3 have a single destination and follow the scheme in the form of equation (3.16), while the last two models have two destinations and follow the scheme in the form of equation (3.20). Furthermore, when each source has

Table 3.3: Signal flow during two consecutive transmissions in various networks

Signal flow at t	Signal flow at t+1	Network model
S_i to D_i	S_i to D_i	$1 \times 2 \times 1$
S_i to D_i	S_j to D_i	$2 \times 2 \times 1$
S_i to D_i	S_i to D_j	$1 \times 2 \times 2$
S_i to D_i	S_j to D_j	$2 \times 2 \times 2$

independent messages for each destination, this is referred to as an x-channel model. A network can also be derived by concatenating multiple models from Table 3.3. For example, a source can send to two destinations, at $TS = t$ and $TS = t + 1$, forming a $1 \times 2 \times 2$ network. Then a second source can pick the transmission line at $TS = t + 2$ and $TS = t + 3$, with the network becoming $2 \times 2 \times 2$ at $TS = t + 1$ and $TS = t + 2$ and changing to $1 \times 2 \times 2$ at $TS = t + 2$ and $TS = t + 3$.

3.2.2 Proposed IRI Cancellation in a Downlink Cellular Network

The main idea of the proposed IRI cancellation method in the $2 \times 2 \times 2$ network is that the receivers actively overhear the signals at two TSs: (i) one of them representing data of interest along the data path affected by the crosstalk and (ii) the other being used to obtain the reference signal deployed in the cancellation process. With this as the driving principle for new applications, it is of particular interest to apply the developed scheme in a cellular-type network. This interest meets the objectives of the work in this thesis. While the general goal of the thesis is to increase the data rate in various networks, users with limited resources in cellular networks were the main focus of attention in the present research. For example, the problem of users with a limited number of antennas is addressed in Chapter 2, and the problem of users located far from the base station (BS) where the signal quality of the direct link becomes poor and unreliable is discussed in this subsection. These two problems combined will also be considered in the following chapter.

This subsection considers a $1 \times 2 \times K$ network representing a single BS serving K users via two relays, as shown in Fig. 3.7. Such a network can be obtained from the $2 \times 2 \times 2$ system considered above by consolidating the two transmitters shown in Fig. 3.6 into one BS, and splitting the two receivers into K mobile stations (MSs). Note that these users suffer from a very poor direct link, and that the only path which can provide them with better quality service is through the relay nodes. In the case of two dedicated relays within one cell sector, these two relays are positioned in a fixed location between the users and the base station. The optimal location of these relays is beyond the scope of this thesis. However, it is assumed that either relay can serve any of the K users, and that therefore all of the users can listen to either of the two relays.

In order to incorporate the proposed IRI cancellation scheme into a downlink cellular system, the $1 \times 2 \times K$ network which represents the signal flow from the BS to the MSs through relays can be interpreted as multiple cascaded networks of the type $1 \times 2 \times 2$ shown in Table 3.3. The network has a single source, S_i, and K destinations. If it is assumed that the destination served during $TS = t$ for any given t is D_i, and that the destination to be served in $TS = t + 1$ is D_j for $j \neq i$, then the $1 \times 2 \times K$ network becomes a cascade of networks of the type $1 \times 2 \times 2$ indicated in Table 3.3. Thus, the cellular network can be considered in the context of the discussion presented in Subsection 3.2.1. Any user that can listen to the two relay nodes can therefore queue for service by using the IRI cancellation scheme developed in equation (3.19).

While the above discussion focuses on the downlink path of the system, the uplink

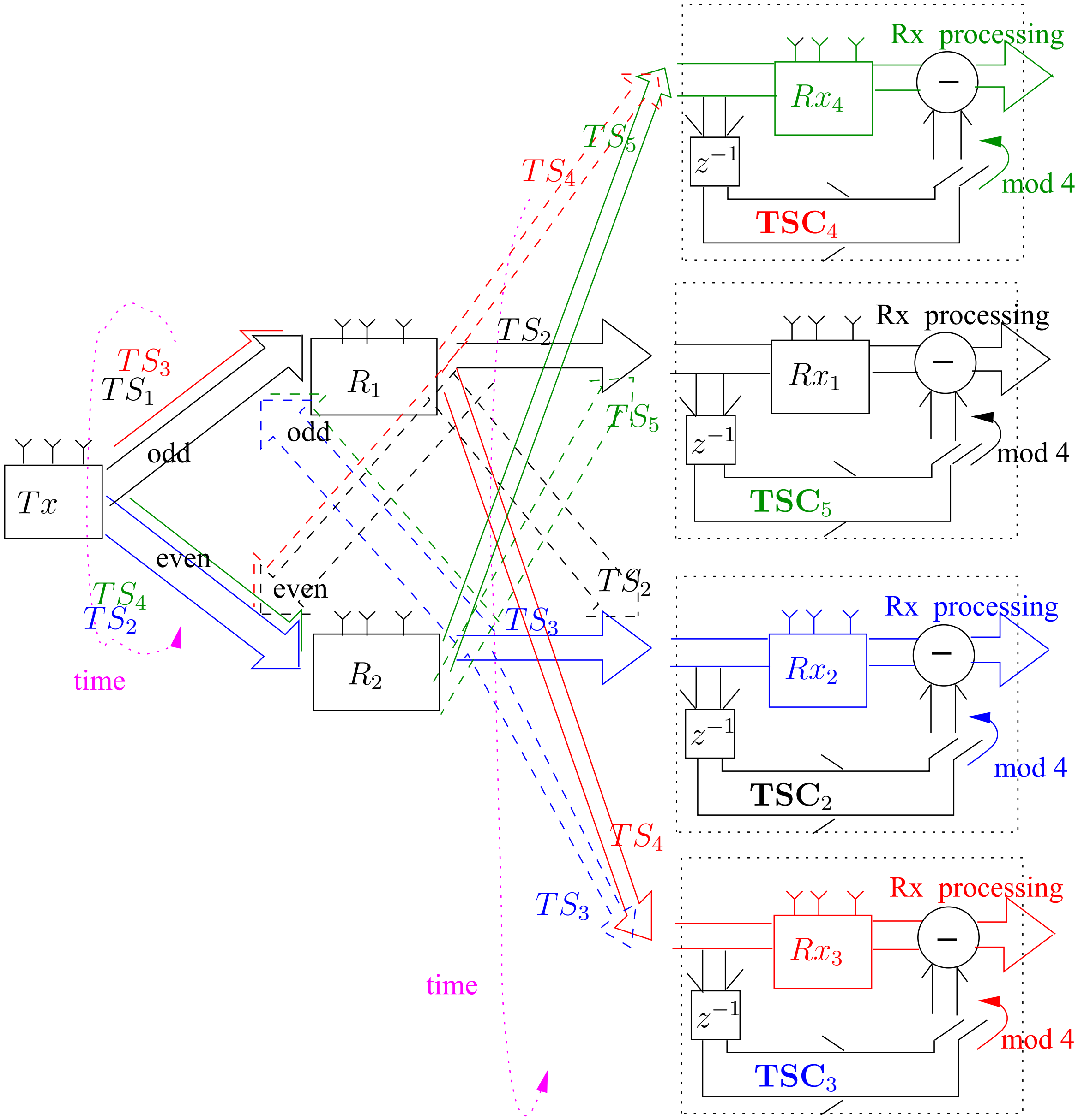


Figure 3.7: TS scheduling and IRI cancellation in an AF MIMO $1 \times 2 \times K$ network.

path can likewise be interpreted as a cascade of networks of the type $2 \times 2 \times 1$ shown in Table 3.3. In fact, the uplink path, or $k \times 2 \times 1$, does not require the MSs, which are the sources in this case, to participate in two consecutive time slots. When an MS acts as the source, it queues for a time slot and sends during the allocated time slot only. This is the opposite of the case where the MS acts as a destination; in order for the MS to extract useful information, it has to listen during the time slot preceding its allocated time slot, as well as during the allocated time slot.

In this network it is assumed that all nodes are equipped with the same number of antennas. Cellular systems in which MSs are equipped with only a single antenna are addressed in the next chapter. Furthermore, the discussion so far has focused on one path of data flow, from the source to the destination. The next subsection considers a network of three transceivers that share information through two intermediate nodes.

3.2.3 Three Transceivers and Two Relays

A particularly interesting aspect of the proposed IRI cancellation scheme is its applicability to a three-transceiver network which communicates via intermediate communication nodes, as illustrated in Fig. 3.8. Unlike the networks discussed previously, where the flow of data takes place in one direction, from source to destination, the network represented in Fig. 3.8 has traffic flow in both directions, in and out, and the sources also serve as the destinations. Furthermore, each source communicates with both of the other two sources, and over one cycle each transceiver sends twice and receives twice.

Extensive research has been done on similar Y-channel networks in the field of multi-way communications. However, most studies focus on signal alignment using a single relay node. Although the y-channel networks allow all transceivers to send simultaneously and receive simultaneously via cooperative precoding at the transceivers and beamforming at the relay, it is necessary for all instantaneous CSIs to be available

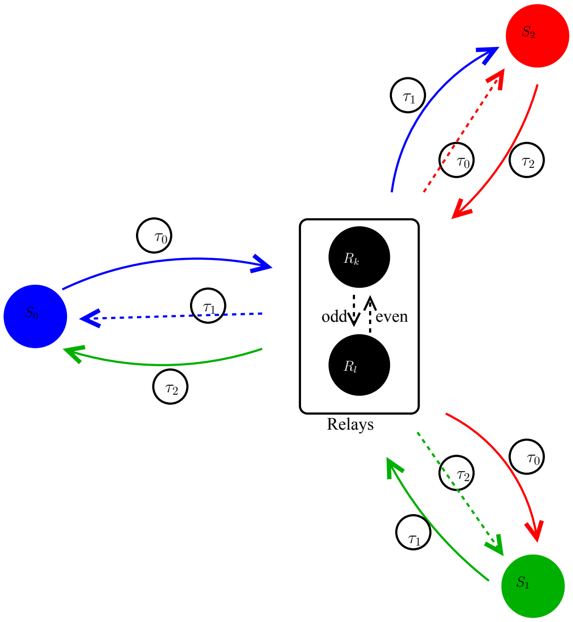


Figure 3.8: Three Transceiver and two relay Network.

globally over the network. This prerequisite may not be practical, since the signal alignment process is based on manipulating the CSI over the network, and hence is highly sensitive to any change in network channel conditions.

Following proper scheduling and making use of opportunistic listening, successive relaying may provide a more practical method to help three transceivers to communicate efficiently. The entire IRI cancellation process is performed at the receiving end of the transceiver, and any CSI requirements can be fed to the receiver. Moreover, it is assumed that all nodes are equipped with the same number of antennas and that the number of data streams sent or received in any particular time slot is equal to the number of antennas at a single node. This includes the SISO case that is not applicable in a y-channel network. Furthermore, any signals sent can be made available at all nodes through further manipulation of the signals received and heard opportunistically.

If it is assumed that each transceiver always has signals to be shared with the other two transceivers, then successive relaying according to the proposed IRI cancellation scheme continues without interruption. Figure 3.8 presents the signal sharing criteria. At $TS = \tau_i = T + i$, where T denotes the cycle number and $i \in \{0, 1, 2\}$, S_i sends signals to the relay nodes. Then at $TS = \tau_{i+1}$, if $T + i$ is odd, R_k relays the signals, causing interference to R_l. Otherwise, R_l forwards the signals, causing IRI at R_k. These signals, as well as their IRI, arrive at S_{i+2} along the solid line shown in Fig. 3.8 as received signals, and arrive at S_i along the dashed line shown in Fig. 3.8, which is the opportunistic listening path. S_{i+1} operates in half-duplex mode and is busy sending signals, and therefore does not receive during $TS = \tau_{i+1}$.

In accordance with the configuration shown in Fig. 3.8 and the application of the developed scheme, it appears as if each transceiver sends only to one of the other transceivers and receives only from one. In other words, it seems as if S_i sends desired signals only to S_{i+2} and receives desired signals only from S_{i+1}. However, the signals opportunistically heard by S_i at $TS = \tau_{i+1}$ can not only be used to remove the IRI from the received signals at $TS = \tau_{i+2}$ but can also be manipulated to extract signals originally sent by S_{i+2} and supposed to be received at $TS = \tau_{i+3} = \tau_i$. The desired signals transmitted by S_{i+1} can be extracted by directly applying the developed scheme in (3.16). If (3.13) and (3.14) are rewritten to accommodate the adjustment in the network, the received signals at S_i during $TS = \tau_{i+1}$ and S_{i+2} can be expressed as:

$$\mathbf{y}(\tau_{i+1}) = \mathbf{B_{ik}W_k r_k}(\tau_i) + \mathbf{n_i}(\tau_{i+1}) \tag{3.21}$$

$$\begin{aligned}\mathbf{y}(\tau_{i+2}) = {} & \sqrt{P}\mathbf{B_{il}W_l A_{li+1}S_{i+1}}(\tau_{i+1}) \\ & + \mathbf{B_{il}W_l H_{kl}W_k r_k}(\tau_i) + \mathbf{B_{il}W_l n_{R_l}}(\tau_{i+1}) + \mathbf{n_i}(\tau_{i+2})\end{aligned} \tag{3.22}$$

It is assumed here, without loss of generality, that $TS = \tau_i$ is even. Thus, the signals received at S_i during $TS = \tau_{i+2}$ after applying the IRI cancellation scheme are:

$$\begin{aligned}\hat{\mathbf{y}}(\tau_{i+2}) = &\sqrt{P}\mathbf{B_{il}W_{l}A_{li+1}S_{i+1}}(\tau_{i+1})\\ &+\mathbf{B_{il}W_{l}n_{R_l}}(\tau_{i+1}) - \mathbf{B_{il}W_{l}H_{lk}B_{ik}^{-1}n_{i}}(\tau_{i+1}) + \mathbf{n_{i}}(\tau_{i+2})\end{aligned} \tag{3.23}$$

After processing the received signals in (3.23), transceiver S_i knows its signals and the signals transmitted by S_{i+2}, as well as the signals originally sent by S_i itself. The signals that are missing are those sent by S_{i+1}, which are the signals broadcast by the relay nodes while S_i is busy transmitting to the relay nodes. Since this operation is in half-duplex mode, the signals cannot be heard by S_i. However, S_i has sufficient information to enable it to extract these missing signals, because these signals are included in the signals received during the opportunistic listening period. These signals travel along the dashed path shown in Fig. 3.8, and are expressed in compact form in equation (3.21). When this equation is expanded to focus on the desired missing signals, it becomes:

$$\begin{aligned}\mathbf{y}(\tau_{i+1}) = &\mathbf{B_{ik}W_{k}A_{ki}x}(\tau_i) + \sqrt{P}\mathbf{B_{ik}W_{k}H_{kl}W_{l}S_{i-1}}(\tau_{i-1})\\ &+ \mathbf{B_{ik}W_{k}H_{kl}W_{l}H_{lk}W_{k}r_{k}}(\tau_{i-2}) + \mathbf{B_{ik}W_{k}n_{R_k}}(\tau_i) + \mathbf{n_{i}}(\tau_{i+1})\end{aligned} \tag{3.24}$$

The expanded form of (3.24) can be expressed in terms of received and transmitted signals at S_i:

$$\begin{aligned}\mathbf{y}(\tau_{i+1}) = &\mathbf{B_{ik}W_{k}A_{ki}x}(\tau_i) + \sqrt{P}\mathbf{B_{ik}W_{k}H_{kl}W_{l}S_{i-1}}(\tau_{i-1})\\ &+ \mathbf{B_{ik}W_{k}H_{kl}W_{l}H_{lk}B_{ik}^{-1}y}(\tau_{i-1}) + \mathbf{B_{ik}W_{k}n_{R_k}}(\tau_i) + \mathbf{n_{i}}(\tau_{i+1})\end{aligned} \tag{3.25}$$

Thus, the received signals $\mathbf{y}(\tau_{i+1})$ are comprised of an attenuated version of $\mathbf{x}(\tau_i)$, an attenuated version of $\mathbf{y}(\tau_{i-1})$, and the signals of interest to S_i. Moreover, the received signals $\mathbf{y}(\tau_{i-1})$ for the current cycle are originally the received signals $\mathbf{y}(\tau_{i+2})$ in the previous cycle. These are the signals that arrive at S_i along the solid line shown in Fig. 3.8.

The filtering process for the whole system in this network is presented as a schematic block diagram in Fig. (3.9), where $\alpha_1 = -\mathbf{B_{il}W_{l}H_{lk}B_{ik}^{-1}}$, $\alpha_2 = -\mathbf{B_{ik}W_{k}A_{ki}}$,

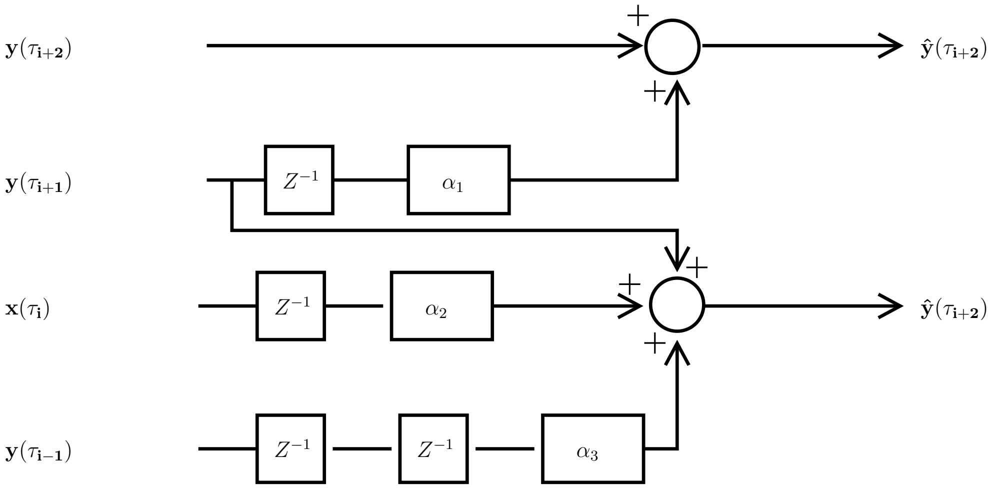


Figure 3.9: Schematic diagram illustrating how the desired signals can be recovered by filtering out the accompanying interference.

and $\alpha_3 = -\mathbf{B_{ik}W_kH_{kl}W_lH_{lk}B_{ik}^{-1}}$. It should be noted that the unprocessed copy of $\mathbf{y}(\tau_{i+1})$ is an essential ingredient for extracting the desired received signals from the other two transceivers. After adding $\alpha_2\mathbf{x}(\tau_i)$ and $\alpha_3\mathbf{y}(\tau_{i-1})$, the filtered copy of the received signals $\mathbf{y}(\tau_{i+1})$ becomes:

$$\begin{aligned}\hat{\mathbf{y}}(\tau_{i+1}) &= \sqrt{P}\mathbf{B_{ik}W_kH_{kl}W_lS_{i-1}}(\tau_{i-1}) \\ &\quad + \mathbf{B_{ik}W_kH_{kl}W_ln_l}(\tau_{i-1}) + \alpha_3\mathbf{n_i}(\tau_{i-1}) + \mathbf{B_{ik}W_kn_{R_k}}(\tau_i) + \mathbf{n_i}(\tau_{i+1})\end{aligned} \tag{3.26}$$

The result shown in (3.26) extends the flexibility of the IRI cancellation scheme, so that it can be applied to further networks.

Applications of the proposed IRI cancellation scheme are not limited to the networks discussed above. These networks have been selected to provide insight into the power and flexibility of the developed scheme. In addition, further capabilities and extensions have been developed during the discussion of the network in this section. The next section analyzes the performance of the proposed scheme in terms of capacity and BER, for various scenarios.

3.3 Performance Analysis

This section examines the performance of the proposed scheme. Even though the scheme is applicable to multiple networks, its performance can be analyzed from the receiver point of view, which is common to all the networks with two exceptions. The first exception is in the $1\times2\times1$ network, where the two relays serve the same receiver, since this does not provide the flexibility of assigning a particular relay to a certain receiver in order to improve the quality of the received signals. The second exception is in the three-transceiver network, where the signal processing differs slightly from that in the main scheme. In particular, in the processing of the second set of received signals, these signals are passed through two consecutive filtering processes, which thus affect the quality of these signals.

The performance of the scheme will be presented in the context of capacity versus SNR, and BER versus SNR, where all nodes are equipped with two antennas. It should be noted that although the performance of the scheme is presented here using nodes with two antennas, the scheme is more general and is applicable to any number of antennas. Moreover, it is assumed that the transmit power is distributed evenly among all transmitting antennas, and that the total transmit power is equally divided between the source and the relay. From a power allocation point of view, dividing the power equally between the transmitter and the relay is the optimal setting when the relay is positioned midway between the source and the destination. In this scenario, with a high SNR, the amplifying factor of the relay compensates for loss in signal power due to channel attenuation between the source and the relay.

The following subsection discusses the performance of the scheme in terms of capacity versus SNR.

3.3.1 Capacity of Successive Relaying with IRI Cancellation at the Receivers

In general, the focus of successive relaying is to increase the throughput of a system by reducing, if not eliminating, the effects of the prelog factor. Thus, the primary motivation for studying successive relaying schemes is to improve the capacity of the network by multiplexing more signals in the available resources, e.g., the time slots in the networks of interest. Researchers usually focus on improving the capacity of the system when the environment of the network maintains sufficient reliability or, in other words, when the system exhibits a high SNR. Thus, the first measurement discussed here in order to evaluate the scheme is the improvement in throughput.

Figure 3.10 presents the capacity performance of the proposed successive relaying scheme (plotted as a dashed blue line) as compared to a conventional scheme (plotted as a solid orange line with stars) in a Rayleigh fading environment. To avoid interference between the nodes, the conventional relaying approach does not allow two nodes to transmit at the same time. As the SNR increases, the capacity of the proposed scheme diverges sharply from that of the conventional approach, with the proposed scheme clearly outperforming the conventional approach. For example, when the SNR is in the range of 40 dB, the capacity of the proposed scheme is approximately double that of the conventional approach.

In fact, the capacity of the proposed scheme is close to that of a full-duplex relaying approach (plotted in Fig. 3.10 as a dashed yellow line with plus signs). The full-duplex system used for comparison here is assumed to be ideal in the sense that it continuously relays signals without causing self-interference. In other words, the comparison in Fig. 3.10 is not with practical full-duplex relaying that suffers from self-interference, but rather with an ideal full-duplex system, where the signals are accompanied only by AWGN. In Fig. 3.10 it can be seen that there is a constant gap

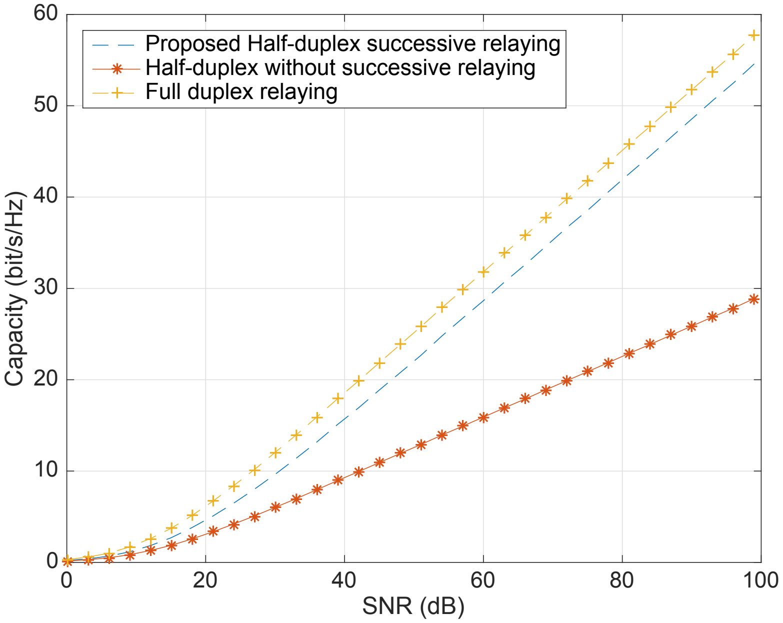


Figure 3.10: Capacity comparisons between the proposed successive relaying scheme, a conventional relaying scheme with no successive relaying, and full-duplex relaying with no self-interference effects.

between the performance of the proposed scheme and that of the ideal full-duplex approach. This gap reflects the cost of the introduced noise term which, since it is not only introduced but also amplified, is the key challenge of the proposed scheme. The impact of this term is discussed further in the next subsection, which analyzes the BER performance.

The capacity of the proposed scheme is slightly different when it is applied to the three-transceiver network discussed in Subsection 3.2.3. There are two sets of signals that must be recovered by transceiver S_i: One set of signals sent from S_{i+1}, and another set sent by S_{i+2}. The signals transmitted by S_{i+2} pass through a filtering process which is the same as that used for the signals in the one-way flow networks. Thus, the capacity achieved by these signals can be expected to equal that of the proposed scheme shown in Fig. 3.10. The capacity of these signals is also presented in Fig. 3.11 (plotted as a dashed blue line). However, the set of signals which arrives

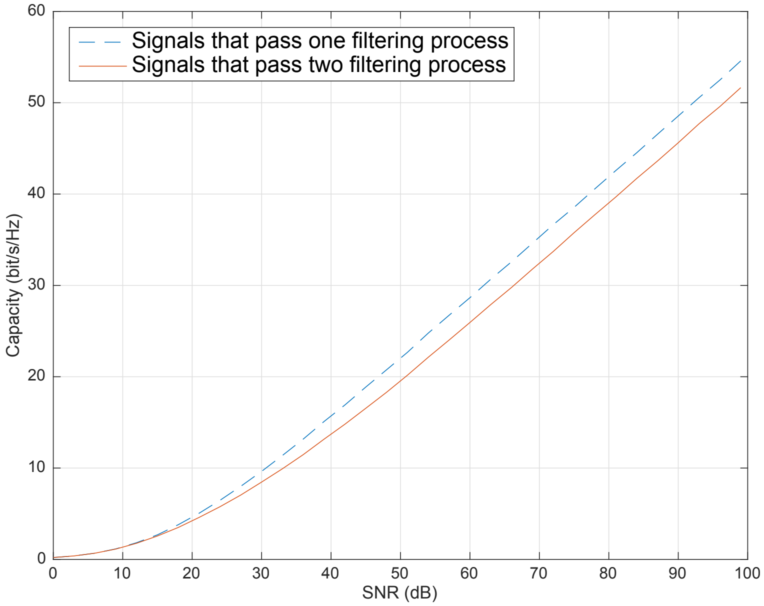


Figure 3.11: Capacity limits in the three-transceiver network.

at S_i from S_{i+1} passes through two stages of filtering, and has to travel further before arriving at S_i. Thus, the capacity achieved by the signals sent by S_{i+1} (plotted as a solid red line in Fig. 3.11) is slightly lower than that of the signals arriving from S_{i+2}. This affects the capacity of the network.

The next subsection discusses the performance of the proposed scheme from the point of view of BER.

3.3.2 BER Performance of Successive Relaying with IRI Cancellation at the Receivers

The improvement in throughput comes at a cost. In the proposed scheme, a new noise term is introduced and amplified, representing a key challenge in the performance of the scheme in terms of bit error rate (BER). The impact of the additional noise term on the performance of the scheme is analyzed by studying the relationship between the signal transmit power and the BER.

It is assumed that all channels in the network experience Rayleigh fading with unit variance. The impact of the additional noise during the IRI cancellation process was examined with the aid of MATLAB®. With these channel conditions and the adoption of maximum likelihood detection, as shown in Fig. 3.12 there is a noticeable degradation in the BER performance of the proposed scheme (plotted as a dashed blue line) as compared with a conservative relaying approach. This is due to the fact that the added noise is enhanced by a multiplicative factor equal to the inverse of the channels between the receiver and one of the relays.

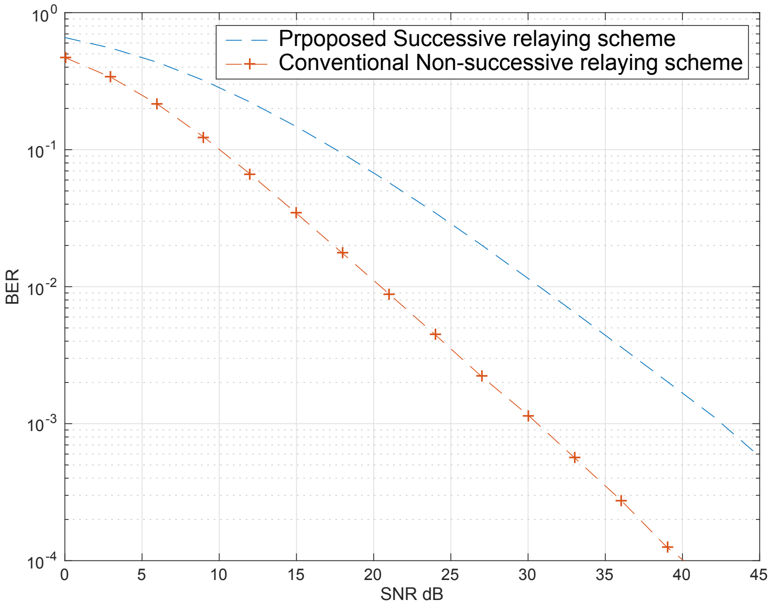


Figure 3.12: The effect on BER performance of the additional noise term introduced in the proposed scheme.

The impact of the channel inverse noise enhancement suggests the possibility of studying which relay should be chosen to serve which user. Choosing the assisting relay for each user may not be applicable for networks with a single destination, but could help to improve the performance of other networks. In order to investigate this aspect, the BER code was run while varying the quality of the channel between the desired destination and the relay which causes the interference: $\mathbf{B_{12}}$ in (3.14). The remaining channels are assumed to have Rayleigh fading with unit variance. The relative strength of $\mathbf{B_{12}}$ is set to be half, equal to, or double the strength of $\mathbf{B_{11}}$, which is the channel that connects the destination and the relay transmitting the desired signals.

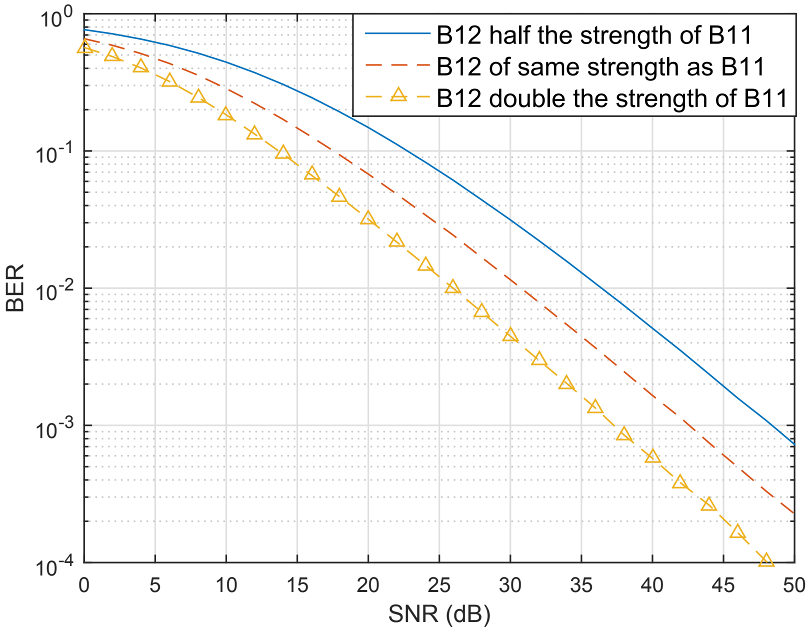


Figure 3.13: BERs of the proposed scheme, where $\mathbf{B_{12}}$ has a relative strength that is (i) half, (ii) equal to, or (iii) double the strength of $\mathbf{B_{11}}$.

As illustrated in Fig. 3.13, when the channel strength of $\mathbf{B_{12}}$ is about half that of $\mathbf{B_{11}}$, the BER performance (plotted in Fig. 3.13 as a solid blue line) exhibits degradation as compared to the scenario where both channels have approximately equal strength (plotted as a dashed red line). Moreover, the BER performance improves in the case where $\mathbf{B_{12}}$ is approximately double the strength of $\mathbf{B_{11}}$ (plotted as a dashed yellow line with triangles). This means that choosing the relay with better channel quality to transmit these signals improves the BER performance.

Even though the proposed scheme has broad applications and can be used in various networks, it faces a practical challenge when one of the relays is in the region of the other relay but out of range of the destination. In this situation, the destination would receive the desired signals accompanied by interference that cannot be filtered out. A problem also arises when there are fewer antennas at the destination than at the relays, since inverting the channel matrix requires the relay and destination to have an equal number of antennas. The following chapter addresses this challenge and presents a cellular network which benefits from the technique with a heterogeneous number of antennas in the network.

This next section presents the chapter summary.

3.4 Summary

This chapter has presented a new scheme to enhance the throughput of wireless relay networks through co-utilization of the available resources. In the scheme, spatial channel reuse by the source and one of the relays causes inter-relay interference, which is dealt with at the destination. The proposed scheme is generalized from single-source single- destination networks to multiple networks that can be adopted in practical applications.

Chapter 4

Source Pre-coding to Cancel IRI in Networks with Successive Relaying

Chapter 3 presented two-hop MIMO networks with AF successive relaying where the inter-relay interference (IRI) is fully-canceled at the destinations. In these systems with half-duplex relays, through TDMA scheduling, the wireless medium in both hops is always utilized through the simultaneous transmission of the source(s) and one of the relays which in turn leads to IRI. The developed in Chapter 3 IRI cancellation scheme through the post-processing at the destination is only applicable if (i) the destinations can opportunistically overhear signals transmitted by the relays contributing to IRI; (ii) the number of receive antennas at the destinations and relays are the same and (iii) destinations are aware of CSI between the relays and the local CSI between the destination and the assisting relays. However, in some network topologies with successive relaying these assumptions may not hold and one may need to resort to the alternative IRI cancellation schemes.

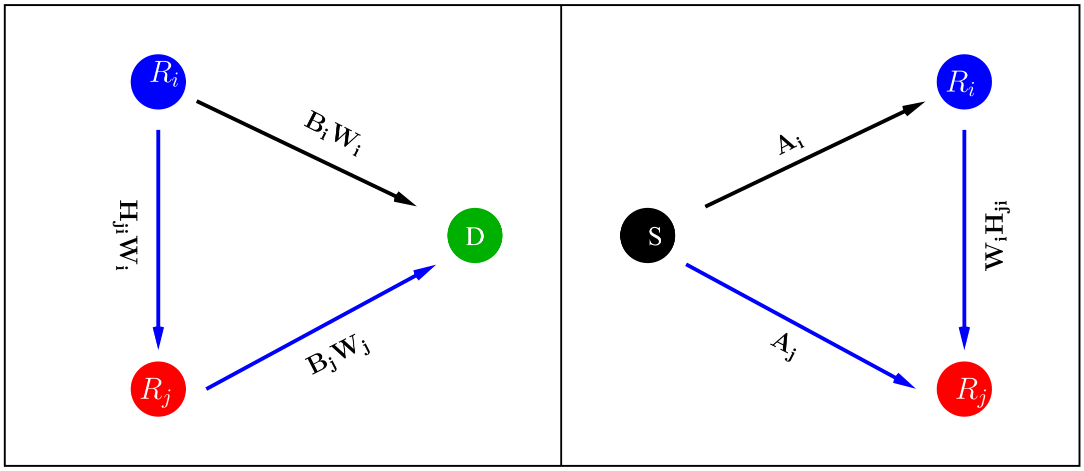
In this chapter, the processing to mitigate IRI is performed exclusively at the source using CSI between the source and the multiple relays supporting corresponding receivers. The proposed linear precoding exploits the principles of signal alignment at the BS and considers recursive characteristics of the IRI analyzed in Chapter 3. The scheme performance is affected by the noise accumulation which is addressed in this chapter by (i) time scheduling of relays' participation (depending on their positions in the network topology) so as to benefit from spatial attenuation of signals and (ii) periodically re-initializing source transmissions. Initial development of the IRI cancellation, described in Section 4.1, is carried out in the context of downlink transmissions in the single-cell cellular network, where the source and the destinations are the BS and MSs, respectively, and where the reception at each MSs is assisted by one dedicated AF relay. Subsequently, it will be demonstrated that the proposed scheme could also be deployed on the uplink. In the system model adopted in Section 4.1, the number of antennas at the BS, relays and MSs is set to M. The developed system model is then extended in Section 4.2 to support MSs each equipped with a single antenna. In this system, the relays and the BS are still equipped with M antennas, however, to deliver M independent messages to the MSs in one transmission cycle supporting all the users, the individual relays will transmit in M time slots as opposed to one time slot. In this system, the source precoding IRI cancellation scheme is identical to that developed in Section 4.1 except that the BS is incorporates appropriately chosen input signals as part of the pre-processing. In this modified system, the developed solution increases the number of independent streams multiplexed in the network from one stream per MS in TDMA (limited by the single antenna at the MS) to M streams in one transmission cycle.

Section 4.3 shows the ability of the new scheme to full IRI cancellation, and also demonstrates the performance trade-offs between the bit error rate (BER) improvements and the loss in throughput efficiency due to flushing. Finally, the chapter is summarized in Section 4.4

4.1 IRI Cancellation at the Source

Relays in our MIMO networks with successive relaying are limited to retransmit their received signals and since and we considered IRI cancellation at the destination in Chapter 3, in this chapter the focus is on the IRI cancellation at the source. This is because in some network configurations, we may not have sufficient conditions to perform IRI cancellation at the destination as discussed in Section 4.1.1. The underlying principle for the IRI cancellation at the source alone is the observation that relays re-transmit interference-free signals. This is because the destinations are not performing any post-processing, and the primary relay assisting the reception of its corresponding destination must receive interference free signal that is later retransmitted to the destination. It is our critical assumption in this chapter that the destinations are not overhearing signals from other relay transmissions except for the one that they are associated with and as such we cannot perform the IRI cancellation at the destination. In the network scenario considered in the present chapter, destinations cannot (or just do not) benefit from the opportunistic listening as presented in Chapter 3.

For the IRI cancellation at the source, the source has to pre-code its signals in order for the relays to transmit interference free signals. Before we propose the IRI cancellation scheme at the source, we contrast the IRI effects and possible means of IRI cancellation at the source and the destination with the help of Figs. 4.1(a) and 4.1(b). In these figures, we follow the convention for channel gain matrices $(\mathbf{A_i}, \mathbf{B_i}, \mathbf{H_{ij}})$ between different nodes (S for source, D for the destination and R_i for the relay) in the network as presented in Chapter 3. For the IRI cancellation at the destination in Fig. 4.1(a), the destination makes use of the fact that the IRI arrives at the destination in two consecutive time slots from two different paths as discussed in the previous chapter. For the IRI cancellation at the source in Fig. 4.1(b), to obtain the interference-free signals at the relays, the source can cancel the IRI effects by sending the signals “twice in two time slots”: one to be transmitted to the destination (primary relay) and the other to zero-force IRI at the other relay. In Fig. 4.1(b) relevant to this chapter, the source sends signal along the black arrow to R_i (the primary relay for the destination) in time slot τ. This signal leaks in time slot $\tau + 1$ to R_j reception from the source when R_i re-transmits its signal to the associated destination. If nothing was done, the R_j signal to be re-transmitted at $\tau + 1$ would not be interference free. However, the interfering signal from R_i to R_j at $\tau + 1$ is known to the source that can send at $\tau + 1$ a modified (based on the CSI, i.e. $\mathbf{A_i}, \mathbf{A_j}, \mathbf{H_{ij}}$) copy of the signal that was sent at τ to R_i along with the new signal to R_j — this is in order to cancel the IRI effect at R_j.



(a) (b)

Figure 4.1: The IRI effects as a results of two path reception along:(i) the direct path and (ii) through the other relay path.

In our system model in this section we consider the following topology for the network as visualized in Fig. 4.2 where relays are spaced to cover a sector (a pizza slice) of a circle (a cell) while the source is placed in the middle of the cell. The source on a first reading can be interpreted as the BS so in this case we talk about the downlink transmissions. Our tacit assumption here is that the BS and relays

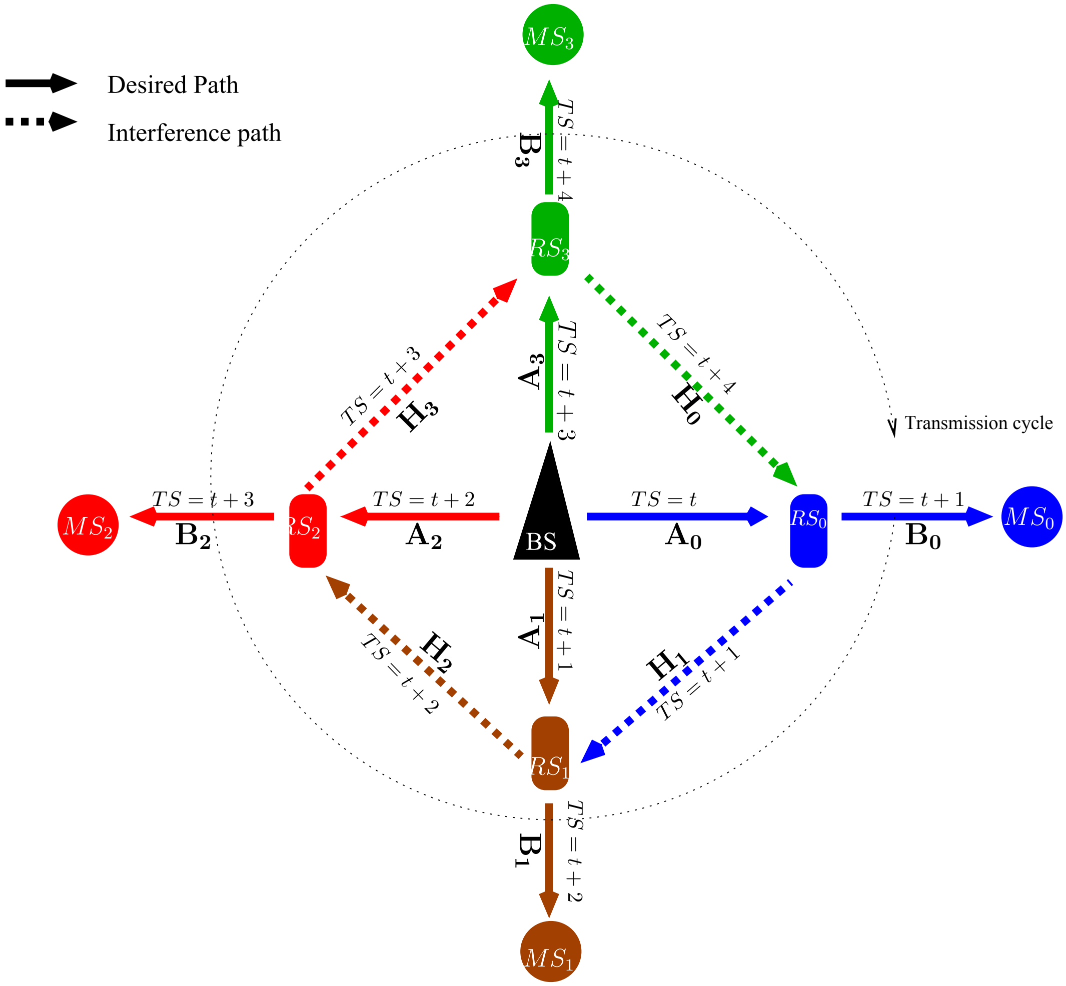


Figure 4.2: TS scheduling with desired and IRI paths in AF MIMO $1 \times K \times K$ network.

form a fixed infrastructure with controlled positions for these devices. Figure 4.2 shows only four relays evenly spaced along the cell boundary. While the BS and RSs are in the fixed positions, the MSs are mobile devices and can move within their associated sectors. However, in our representation Fig. 4.2 to easier explain transmission ranges/gains the MSs and their associated relays are along the same radial direction.

We assume that the location of the RSs are properly selected to ensure (i) the source-relay channel quality to be good, (ii) one relay covers a unique sector (iii) and the relay receiving channel gain matrices experience independent slow fading. While we would like to have spatial separation among relays as to avoid IRI, the relay-to-relay channel gains causing IRI cannot be disregarded. This is because to ensure acceptable signal strength between the BS and the relays, the relays cannot be spaced too far from the BS which imposes limits on the spatial separation of relays. In other words, to obtain acceptable source-to-relay channel quality (reduced distance between BS and the RSs), we are increasing relay-to-relay channel gains (reduced distance between RSs) and cause the IRI which calls for IRI cancellation.

The source-based IRI cancellation scheme developed in this section is applicable to multiple (K) relays distributed evenly along a circle circumference and assisting K corresponding MSs. In Fig. 4.2 we work with four relays and time indexing is modulo 4 when the BS sends in clock-wise, round robin cycles following TDMA principles continuously to the $K = 4$ relays. At the same time, relays (actually one at a time with one time slot delay) retransmit clockwise to their corresponding destinations. In Fig. 4.2 we show with the continues, color-coded, lines the desired signal paths to the destinations and the dashed lines are used to show the potential interference paths. The interference paths are also following the clock-wise patterns because of time scheduling. This is because even though the counter clock-wise interference paths exist physically, their are not affecting the relays as one relay receives data only over one time slot out of $K = 4$ time slots.

Since the source-based IRI cancellation for one relay is repeated K times for every relay, working with $K = 4$ in Fig. 4.2 does not impose any limitations on our scheme. Though it should be observed that the more relays along the ring, the smaller the distance between the relays. Thus, the relay covers smaller sector size which on the positive side reduces its transmit power while on the negative side creates stronger IRI and more noise accumulation as elaborated later on. Furthermore, adding more relays increases the cost invested in the infrastructure which is not discussed in this work.

In the next subsection, we elaborate further on the system level aspects and the operation of the proposed transmission scheme. The derivations for the proposed source-based IRI cancellation are presented in subsection 4.1.2, while the noise accumulation phenomenon is addressed in subsection 4.1.3. The performance of the proposed scheme is discussed in subsection 4.3.

4.1.1 Successive Transmissions

In our system model, we consider a source or BS transmitting to K users (MSs) through K relay stations (RSs). We refer to this configuration as $1 \times K \times K$ network. It should be noted that this configuration is used to present the scheme in a clear and systematic way. However, in reality each relay serves a sector of the cell, in which a group of users are associated with a particular relay. Dedicating a relay to every user would create a mesh of relays in the cell that are inefficiently deployed in the system.

In Section 4.1, we initially discuss the situation where the BS is sending independent data to K active MSs in separate sectors through K RSs where the dedicated AF relay is forwarding interference free data to the assisted MS. Throughout the whole Section 4.1, we assume that the all nodes (the BS, RSs and MSs) in the network visualized in Fig. 4.2 are equipped in M antennas. We relax the latter assumption in Section 4.2 where we consider the MSs equipped with single antennas and we have there M active MSs per sector each assisted by the RSs serving the designated sector.

A simplified layout with $K = 4$ is illustrated in Fig. 4.2. Every MS in Fig. 4.2 receives signals from the BS via the corresponding RS along the "desired" data path, indicated by color-coded solid arrows. Following a clock-wise pattern, the BS transmits continuously to the RSs in round-robin cycles, where one cycle is comprised of K time slots (TSs). It should be noted that transmission continues, and that the last relay transmits its signals during the first time slot of the next cycle. Otherwise, each cycle would consist of $K + 1$ TSs.

The notation t represents how many times the first relay has been included in a full round-robin cycle before a flushing TS occurs. Moreover, in mod K arithmetic, $k \in \{0, 1, \cdots, K-1\}$ indicates the offset from the beginning of the transmission cycle. For the brevity of notation, we denote the k-th RS as R_k and the k-th MS as D_k (D for destination).

During time slot $t+k$, the BS transmits signals to RS_k which in turn amplifies and forwards the signals to D_k at the $t + k + 1^{th}$ time slot. We assume that all terminals are using omnidirectional antennas and in Fig. 4.2 arrows capture only the logical and time flow of signals. Hence, when R_k is transmitting to D_k in TS $t + k + 1$, this transmission is causing interference with the reception of the desired signal being transmitted from the BS in TS $t + k + 1$ to R_{k+1}. This IRI along the interference path is visualized in Fig. 4.2 using the color-coded dashed arrows.

Even though, in Fig. 4.2, the RSs and MSs are visualized as equi-spaced and equidistant from the BS, this topology is only chosen to clarify the initial concepts and the developed algorithm accounts for much more complicated topological distributions of RSs and MSs.

The channel gain matrix between the BS and the relay R_k is denoted as $\mathbf{A}_k$ and the channel gain matrix between R_k and R_{k+1} is given by $\mathbf{H}_{k+1}$. $\mathbf{B}_k$ is the channel gain matrix between R_k and the corresponding destination D_k. Note that our scheme has one directional flow of data and we are indexing the channel access TSs from the BS to the destinations using the same index as that identifying the receiving relay. When the relay retransmits to the associated destination (only once in the transmission cycle), the index increases (using mod K arithmetics). We assume that BS-RS and RS-RS channel state information (CSI) are made available at the BS. The channels between the RSs and their corresponding MSs are not affecting the proposed

IRI cancellation scheme at the BS.

With the above configuration, the received signal at relay R_{k+1} during TS $t+k+1$ is:

$$Y_{R_{k+1}}=\sqrt{P_{k+1}}\mathbf{A}_{k+1}\mathbf{x}_{k+1}+\mathbf{H}_{k+1}\hat{\mathbf{x}}_{R_k}+\mathbf{n}_{R_{k+1}} \tag{4.1}$$

In order, the terms in (4.1) are: (i) signal received along the desired path; (ii) IRI received along the interference path, and (iii) additive white gaussian noise (AWGN) at the R_{k+1} and denoted as $\mathbf{n}_{R_{k+1}}$. Note that for clear presentation we omitted the time index in (4.1). Thus, all of the signals introduced in (4.1) arrives at R_{k+1} during $t+k+1$.

The desired path signal $\sqrt{P_{k+1}}\mathbf{A}_{k+1}\mathbf{x}_{k+1}$ consists of an encoded signals vector $\mathbf{x}_{k+1}$ scaled by the BS transmit power $\sqrt{P_{k+1}}$ and the gain matrix $\mathbf{A}_{k+1}$. Moreover, the encoded signals vector $\mathbf{x}_{k+1}$ is designed to carry the desired information to the $k+1$-th MS ($\mathbf{s}_{k+1}$) and to remove– at the $k+1$-th relay– the interference arriving along the interference path.

Along the interference path, the arriving signal vector $\mathbf{H}_{k+1}\hat{\mathbf{x}}_{R_k}$ is composed of the signal $\hat{\mathbf{x}}_{R_k}$ transmitted by R_k, which is attenuated by the channel gain $\mathbf{H}_{k+1}$. Note that $\hat{\mathbf{x}}_{R_k}$ contains not only the desired signal intended for D_k but also an accumulation of noise terms from the time slots preceding $t+k+1$. The accumulated noise is limiting factor in this scheme and we will discuss its effects on the system performance in Section 4.1.3. The notation used when presenting the received signals at R_{k+1} is summarized in Table 4.1.

In an interference free environment, the received signal by R_{k+1} consists of the desired signal- attenuated by A_{k+1}- transmitted by the BS and this is the signal to be relayed to D_{k+1}. This signal when retransmitted is only accompanied by the AWGN noise at the relay. The task of the pre-coding signals by the BS is to deal with the interfering signals so that the received signal by D_{k+1} is only accompanied with terms of AWGN which cannot be controled at the BS. In other words, the relays are

Table 4.1: A summary of notation used to represent signals at R_{k+1}.

Symbol	Definition
BS	Source (BS)
R_k	Relay station k
D_k	Destination (MS) k assisted by R_k
$\mathbf{A_{k+1}}$	Channel gain matrix from the BS to R_{k+1}
$\mathbf{H_{k+1}}$	Channel gain matrix from R_k to R_{k+1}
$\mathbf{B_{k+1}}$	Channel gain matrix from R_{k+1} to D_{k+1}
P_{k+1}	Power of the signals intended for D_{k+1}
$\mathbf{x_{k+1}}$	Pre-coded signal transmitted by the BS
$\hat{\mathbf{x}}_{R_k}$	Transmitted signal by R_k
$\mathbf{n}_{R_{k+1}}$	AWGN at R_{k+1}

expecting interference-free signals to arrive at their ends. Hence, our objective is to design precoded symbols $\mathbf{x}_{k+1}$ at the BS to achieve this goal.

In the next subsections, we introduce first the criteria for removing the interference through pre-coding at the transmitter. Then we show how to limit the variance of the accumulative noise in the relays and discuss its impact on the system performance.

4.1.2 Interference Cancellation at the Transmitter

The objective of the proposed interference cancellation scheme is to deliver interference-free messages $\mathbf{s}_k(t+k)$ from the BS destined to D_k. That is, the desired signals should arrive at their destinations accompanied only with terms of attenuated additive white Gaussian noise. Therefore, despite the combination of signals that arrives at RS_k during $t+k$, the received signal at D_k in the TS $t+k+1$ should be:

$$Y_{D_k}(t+k+1) = \sqrt{P_k}\mathbf{B}_k\mathbf{F}_k\mathbf{A}_k\mathbf{s}_k(t+k) + \mathbf{B}_k\mathbf{F}_k\mathbf{w}_{R_k}(t+k) + \mathbf{n}_{D_k}(t+k+1) \qquad (4.2)$$

where the desired message vector $\mathbf{s}_k(t+k)$ is scaled by the channel gain matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ which are the attenuation on two hops along the desired path, P_k is the power at the BS used to transmit to R_k and $\mathbf{F}_k$ is an amplifying matrix at R_k (see Fig. 4.2). The $\mathbf{n}_{D_k}(t+1)$ is the omnipresent AWGN at D_k, and the term $\mathbf{w}_{R_k}(t+k)$ is the noise accumulated at the k-th RS which is also amplified by $\mathbf{F}_k$ and attenuated by $\mathbf{B}_k$ along the second hop.

The accumulation of noise $\mathbf{w}_{R_k}(t+k)$ at the relay develops due to the fact that the AF relays do not filter out their noise (from their receiving front ends) when they relay their signals. That is, the instantaneous noise at the relay R_i is amplified by the amplifying gain of the relay and transmitted along with the desired signals to the destinations. However, this noise term also accompanies the interference signals at the next relay R_j. While the interference is known to the transmitter and a pre-coding method could remove its effect, the noise term is random and no processing at the BS can cancel this term. Thus, a new noise term is introduced in the system every relaying period and there is no method to remove it at the AF relays. Except for breaking of the accumulation of noise through restarting the cycles (or "flushing off" the accumulated noise), the accumulated noise level will increase in the system as the number of relaying cycles increases. Note however, that the channel gains between the relays are not strong and the effect of individual noise terms starts relatively strong and then it diminishes as time passes.

Examining the signal received at D_k in (4.2), the forwarded signal by R_k in TS $t+k+1$ is:

$$\hat{\mathbf{x}}_{R_k}(t+k+1) = \sqrt{P_k}\mathbf{F}_k\mathbf{A}_k\mathbf{s}_k(t+k) + \mathbf{F}_k\mathbf{w}_{R_k}(t+k) \tag{4.3}$$

The signals in (4.3) arrives at R_{k+1} as $\hat{\mathbf{x}}_{\mathbf{R_k}}$ at TS $t+k+1$. Substituting (4.3) into

(4.1), the received signal at the relay R_{k+1} becomes:

$$\begin{aligned} Y_{R_{k+1}}(t+k+1) &= \sqrt{P_{k+1}}\mathbf{A}_{k+1}\mathbf{x}_{k+1}(t+k+1) \\ &+\sqrt{P_k}\mathbf{H}_{k+1}\mathbf{F}_k\mathbf{A}_k\mathbf{s}_k(t+k) + \mathbf{H}_{k+1}\mathbf{F}_k\mathbf{w}_{R_k}(t+k) + \mathbf{n}_{R_{k+1}}(t+k+1). \end{aligned} \tag{4.4}$$

Therefore, examining (4.4) and (4.2) with D_{k+1} rather than D_k, the transmitter needs to design $\mathbf{x}_{k+1}(t+k+1)$ so that

$$\sqrt{P_{k+1}}\mathbf{A}_{k+1}\mathbf{s}_{k+1}(t+k+1) = \sqrt{P_{k+1}}\mathbf{A}_{k+1}\mathbf{x}_{k+1}(t+k+1)+\sqrt{P_k}\mathbf{H}_{k+1}\mathbf{F}_k\mathbf{A}_k\mathbf{s}_k(t+k) \tag{4.5}$$

or

$$\mathbf{x}_{k+1}(t+k+1) = \mathbf{s}_{k+1}(t+k+1) - \frac{\sqrt{P_k}}{\sqrt{P_{k+1}}}\mathbf{A}_{k+1}^{-1}\mathbf{H}_{k+1}\mathbf{F}_k\mathbf{A}_k\mathbf{s}_k(t+k). \tag{4.6}$$

The developed precoding in (4.6) is designed to fully cancel the interference from one RS into another due to forwarding of data symbols. This can be seen by inductive reasoning when examining (4.4) and following in time $\mathbf{w}_{R_k}(t+k)$ which contains only the terms due to AWGN at the previously transmitting relays and represents the noise accumulation of the scheme. Note that (4.6) does not deal with the noise which follows random characteristic. The noise term is discussed in more details in the next section.

In addition, the proposed scheme provides some freedom to further manipulate the desired transmitted signals to improve system capacity. The BS can make use of the BS-RS channel knowledge to break the MIMO channel into number of parallel data streams. Then the RS assisted with the knowledge of RS-MSs CSI can serve multiple single antenna MSs equal to the number of parallel data streams available at the RS. As a demonstration, let $\hat{\mathbf{A}}_k = \sqrt{P_k}\mathbf{A}_k$ and the singular value decomposition of $\hat{\mathbf{A}}_k$ be $\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^H$. Then $\hat{\mathbf{s}}_{\mathbf{k}} = \mathbf{U}_k^H\mathbf{s}_k$. Moreover, the amplifier at the relay $\mathbf{F}_k = \mathbf{V}_k\alpha$ which makes the received signals at R_{k+1} when both the BS and R_k are transmitting to be equal to:

$$\begin{aligned}Y_{r_{k+1}}(t+k+1) = \sqrt{P_{k+1}}\mathbf{A}_{k+1}\mathbf{x}_k(t+k+1)+ \\ \mathbf{H}_{k+1}\alpha_k\mathbf{\Sigma}_k\mathbf{s}_k(t+k)+\alpha_k\mathbf{V}_k\mathbf{w}_k + \mathbf{n}_k(t+k+1).\end{aligned} \tag{4.7}$$

and the encoded signals in (4.6)

$$\mathbf{x}_{k+1}(t+k+1)=\hat{\mathbf{s}}_{k+1}(t+k+1) - \frac{1}{\sqrt{P_{k+1}}}\mathbf{A}_{k+1}^{-1}\mathbf{H}_{k+1}\alpha_k\mathbf{\Sigma}_k\mathbf{s}_k(t+k). \tag{4.8}$$

Note that we assumed here that P_{k+1} is a scaler and P_k is already incorporated in Σ. Note also that for modulations with multiple power levels, the variation can be accounted for in the signal vector **s** directly.

For completeness, we mention that in some scenarios the RSs receive their signals using M antenna while transmit their signals using $\hat{M} < M$. In this case, the BS assumes that the unused antenna by RS to be transmitting zeros when it calculates the pre-coded signals.

In the next section, we discuss the accumulation of noise in the developed cancellation scheme resulting from AF nature of our relays.

4.1.3 Mitigating the Effects of Noise Accumulation

Noise accumulation has been identified already as a major drawback in AF alternative relaying schemes. AF relays do not filter their AWGN from the RF front-end. In fact, they enhance it as they amplify both the information bearing portion of the signal and noise the same way. Moreover, the AF alternative relaying keeps the noise term cycling form one relay node to another and in every TS an additional noise term is introduced.

When deploying relays in cellular system, the number of relays and their spatial location, i.e., the state of channels, play a key role in determining the system performance. We examine here the impact of relays on noise accumulation represented by $\mathbf{w}_{R_k}(t+k)$ only in some simplified configurations. Specifically, when the relays are distributed evenly along a circumference of a circle centered at the BS, at least three RS are required to cover the service area.

Also, the more RS along the RS circle, the smaller designated area per RS. This in turn reduces the required relaying power and minimizes the enhancement of the accumulative noise. Power control and scheduling relay transmissions with more favorable channel conditions that minimize noise accumulation is one area in our scheme that could be pursued to mitigate the accumulated noise and is left for further investigations. In what follows and in our result analysis, we assume comparable channel gains in the data and interference paths.

While the AWGN noise accumulation affects the overall system performance, the problem may be mitigated by limiting the accumulated noise variance to the acceptable level or periodically re-initializing the transmission scheme as it is done when dealing with error propagation in some systems. Paying a close attention to the noise accumulation by expanding the accumulated noise at R_k and time slot $t+k$ when $K = 4$ relays are sequentially taking turns, i.e., TSs, in relaying the signals:

$$\begin{aligned}\mathbf{w}_{R_k}(t+k) &= \mathbf{n}_{R_k}(t+k) + \mathbf{H}_k\mathbf{F}_{k-1}\mathbf{n}_{R_{k-1}}(t+k-1) \qquad (4.9)\\ &+\mathbf{H}_k\mathbf{F}_{k-1}\mathbf{H}_{k-1}\mathbf{F}_{k-2}\mathbf{n}_{R_{k-2}}(t+k-2)\\ &+\ldots+\mathbf{H}_k\mathbf{F}_{k-1}\mathbf{H}_{k-1}\mathbf{F}_{k-2}\ldots\mathbf{H}_2\mathbf{F}_1\mathbf{n}_{R_0}(0) \qquad (4.10)\end{aligned}$$

With AWGN at the relays, $\mathbf{n}_{R_i}(t+i)$ being i.i.d., the accumulated $\mathbf{w}_{R_k}(t+k)$ Gaussian noise depends mainly on the relationship between the interference channel conditions, i.e., $\mathbf{H}$ between the transmitting relay and the receiving relay and the amplifier $\mathbf{F}$ of the transmitting relay in a geometric progression fashion. Thus, it is the first two terms in (4.9) that determine how the performance of the scheme degrades as the relaying loop progresses.

In the most comon topologies, the relays are placed at least halfway between the

BS and the boundary of the cell, so the minimum distance between the transmitting RS and the receiving RS can be kept below $\sqrt{2}d$ where d is the distance between the BS and the RS ring. When the number of relays is greater than four, then the scheduling of BS/RS transmissions to minimize the variance of the accumulated noise in (4.9) follows the procedure below:

1. Pick the first RS.
2. The next RS is the RS that has the minimum distance with the current RS with the condition that the distance between them is greater than or equal to $\sqrt{2}d$
3. Continue selecting RSs following step (2) and moving in circle along the ring of the RSs.
4. When the circle turns 360°, the next RS is the next unused RS the inter-distance nol smaller than $\sqrt{2}d$.
5. Repeat step number (3) and (4) until all RS are selected.

In the case of four and three RSs, the distances between the adjacent RS are already greater than or equal to $\sqrt{2}d$ and the scheduling mechanism described does not violate the minimum distance of $\sqrt{2}d$ condition. It is worth observing, that the scheduling of transmission mechanism as well as our IRI cancellation assumes perfect knowledge of CSI — in real-world applications, CSI estimation would have to be addressed and is beyond the scope of this paper.

In the worst case scenario, the channel between the transmitting and receiving relays is only affected by the deterministic signal path loss and the noise power travelling from one relay to another is degraded with factor of $\frac{1}{2d^2}$. We consider this conservative assumption in our evaluation to the accumulating noise.

Finally, to deal with noise accumulation, we could periodically (after a certain number of transmission cycles or TSs) re-initialize the transmissions by pausing the transmission of the BS for one TS. This would “flush out” the accumulated noise at RSs. This solution trades off the the TS utilization with limiting the variance of the accumulated noise or equivalently the worst BER in the last cycle before accumulated noise flushing. Here if the flushing occurs every L TSs (turns), the time utilization by BS (or equivalent spectral efficiency) drops to $\frac{L}{L+1}$, where L is the number of successive TSs (turns) before flushing.

4.2 Increasing the Throughput of Single Antenna Users

The work in Chapter 3 and Section 4.1 focuses on networks deploying nodes with equal number of antennas. Interestingly, the IRI cancellation techniques developed in these parts of the thesis serve as building stones in this section to increase the data rate in networks with single antenna users. Specifically, in the present section, we deal with the case when the BS and relays are equipped with M antennas and the single antenna MSs are located close to the edge of the cell. Moreover, one relay, in its associated sector, aids the reception of M MSs. While in Chapter 2 the downlink path of a cellular type network achieves $\frac{MK}{M+K-1}$ pre-log factor in capacity calculations as in 1.4 through time scheduling and distance separation, the pre-log factor in this section is raised to M following a novel approach through careful scheduling and incorporating new IRI cancellation techniques.

In Section 4.2.1, we present our system model. Then, in Section 4.2.2, we develop IRI cancellation at the BS for downlink transmissions in the underlying network considered in Section 4.2. In Section 4.2.3, we incorporate IRI cancellation at the receiver to assist the uplink process.

4.2.1 Single Antenna Users at the Cell Boundary

In this section, we consider K single antenna users (MSs) that are located close to the boundary of a cell and they are served by a BS equipped with M antennas as shown in Fig. 4.3. Because of distance separation, we assume there is no connectivity between the BS and MSs and the received signal power from the BS at the MS has a negligible effects on the signal processing at the MSs. Therefore, the BS transmissions are assisted by L dedicated RSs that are distributed evenly along a ring between the center of the cell and the boundary.

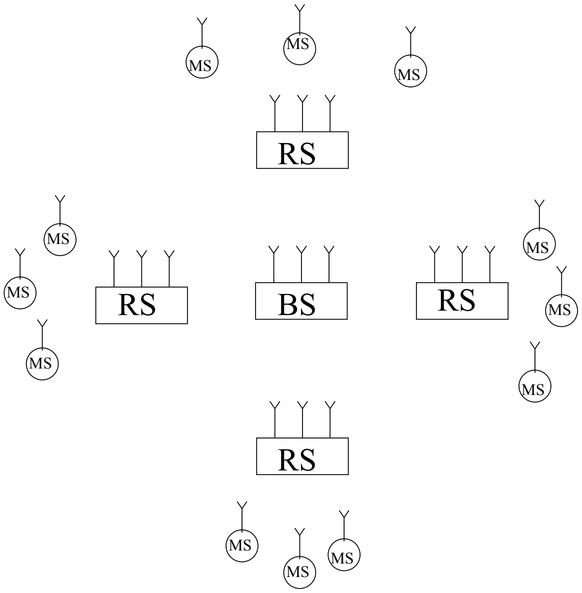


Figure 4.3: $1 \times L \times K$ network layout.

In this $1 \times L \times K$ network, the RSs are also equipped with M antennas and operate in half-duplex mode. We continue to assume that, in the downlink transmissions, the RSs are using the same frequency band as the BS in order to maintain high bandwidth efficiency. In one sector ("the pizza slice"), there are M active single antenna MSs which reception is assisted by the dedicated single relay serving the particular sector. With this $K = L \cdot M$.

Fig. 4.3 shows the layout of the network in the case when $M = 3, L = 4$ and $K = 12$ and it is just a basis to demonstrate the operational principles of our algorithms without imposing any limitations on the actual position of RSs and MSs. What is important, in the general network configurations, is that the relay signals are only overheard by the MSs associated with the dedicated RSs and that the MS's reception from relays is not affected by the transmissions from the BS which could be accomplished by power control in the system. Because in practical systems, the number of relays may be higher than $L = 4$ to support more users, we cannot ignore the possibility of signals transmitted from one relay affecting the reception of MSs that are not associated with that relay and are in the adjacent sectors.

Moreover, we assume that the BS-RS and RS-RS channel gain matrices are not changing when they corresponding nodes are involved in the communication process. All receiving channels are known to their associated relays and this CSI is shared with the BS. We do assume the reciprocity of channels, i.e., $H_{ij} \neq H_{ji}$ where H_{ij} is the channel between the receiving node i and transmitting node j. However, if the channels are reciprocal, the required amount of CSI data to be shared with the BS is reduced to the RS-RS channels which parameters are required in performing the source (BS) based IRI cancellation.

In the IRI cancellation schemes developed in the next two subsections, the MSs are not involved directly in the IRI cancellation process. However, they will follow two different scheduling patterns in the uplink and downlink modes. The RS-MS CSI is important to the RSs in order to set the amplification gain on the uplink. The knowledge of the RS-MS CSI is also necessary for the MSs to decode their signals in the downlink phase. Note that assume Rayleigh fading channel in our result analysis; however, the proposed scheme is flexible to fit other type of fading.

We assume in this section that the relays operate using decode-and-forward (DF) in the downlink and amplify-and-forward (AF) in the uplink. The reasons for adopting

DF for the downlink while AF for the uplink are on one hand that we are looking to provide all users with approximately equal level of service and the signals in the downlink arrive to the RSs interference-free since the cancellation is performed always by the BS. If AF is used in this phase, a MS would have an average service worse than its preceding MS and better than the following MS since the noise is accumulated as the successive relaying continues. DF on the other hand, eliminates the noise accumulation problem. The uplink phase; however, is using AF technique since the cancellation of the interference is left to the BS which is the only node in the network that can listen to both required signals in order to cancel the interference.

In the next subsection, we present the downlink IRI cancellation methodology of our successive relaying scheme.

4.2.2 IRI Cancellation on a Downlink

The signals from the BS to the MSs go through two phases in this subsection. These phases follow the path of a signal set from its initiation at the BS to the destination of these signals at their desired MSs. During the first phase, the BS sends M signals (the signal set or signal vector) to the listening RS within a single time slot. Figure 4.4 represents this phase as the BS sends $M = 3$ data streams (black arrows) to RS_3 during $TS = 3$. In this figure we index the TS and RS similarly as in Fig. 4.2 where we were discussing IRI cancellation at the source but the RS as forwarding the interference-free signal set to a single MS equipped with M antennas. Actually, Figs. 4.4 and 4.2, if we consider group of M single antenna MSs in a single sector as "a single destination" with the virtual antenna array. Note that the color-coded signal representations in Fig. 4.4 capture only the number of streams and their desired flow directions. In practice however signals are transmitted using omni-directional antennas and every receiving antenna gets a combination of the three signals and further manipulations has to be performed in order to separate these signals.

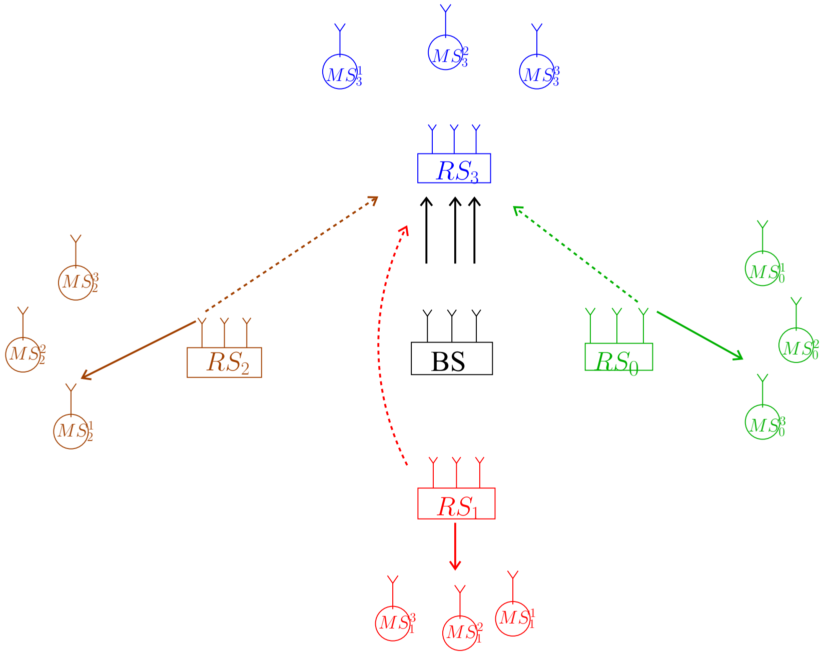


Figure 4.4: BS sending M data streams to a listening RS_3 while the other $M = 3$ RSs are relaying signals to their corresponding destinations and contributing to the IRI through a virtual antenna array.

Following TDMA, the BS sends a set of M signals to RS_k at $TS = k$ where indexing of TSs is modulo L though for convenience we still index the RSs with k. These independent signals at a given TS are intended for M different MSs in a given sectors. If no interference occurs, the signal (in a vector form) received at RS_k during $TS = k$ is:

$$Y_{R_k} = \sqrt{P_{BS}}\mathbf{A}_k\mathbf{s}_k + \mathbf{n}_k \tag{4.11}$$

.

In the second phase, the RSs assume that their received signals are interference-free since the BS performs the IRI cancellation. Thus, RS_k can decode the vector $\mathbf{s}_k$ which contains $M = 3$ independent signals. RS_k splits its received signals into three individual streams and uses one antenna to send a signal to the intended MS per time slot. As Fig. 4.4 captures the flow during only a single time slot, when one relay is receiving data from the BS, every other relay of the three $(L-1)$ transmitting RSs is serving a single MS with a single data stream during $TS = 3$. That is, RS_2 sends its first stream out of three to the first associated MS, RS_1 sends its second data stream to its second associated MS, and RS_0 sends its last stream to its third MS. The number of MS per RS does not have to be necessarily M and the number is used for illustration purposes only. With this three receiving MSs set up, RS_k uses TS_{k+1} to send signals to MS_1^k , TS_{k+2} to send signals to MS_2^k, and TS_{k+3} to send signals to MS_3^k where the subscript k (at the bottom of MS_i^k with $i \in \{1, 2, \cdots, M\}$ indicates the RS associated with the MS and i indexes the MSs served by a given RS. The received signals at these MSs are;

$$Y_{MS_i^k}(k+i) = \sqrt{P_{R_k}} B_i s_i^k + n_i \tag{4.12}$$

The scalar signal s_i in (4.12) is attenuated by only the scalar channel gain between the transmitting antenna at RS_k and MS_i^k and is affected only by AWGN n_i in its RF front end since the RS applies DF in this data flow direction and there is no noise accumulation at the RSs.

The two phases when transmitting signals through one RS usually overlap with the phases of other RS when transmitting other signaling sets since the communication medium is continuously utilized by the BS to send a set of signals to a listening RS and also used by the relays- except for the listening RS- to re-transmit their signals. When the transmitting RSs uses the same medium as the BS, their transmitted signals causes interference on the listening RS. This is illustrated with the dashed arrows leaking from the transmitting RS to the receiving signals in Fig. 4.4. These $M = 3$ dashed arrows leak the IRI to RS_3 when it receives the data from the BS represented with black arrows. With this observation, we established a link between the present section and the source based IRI cancellation in Section 4.1. In essence, in the present system, the IRI in a given TS is caused by the virtual antenna array where the individual antennas are from M RSs receiving from the BS on the previous M TSs. More specifically, when in Fig. 4.4 we had one RS contributing to a vector IRI (of size M) in a given TS because RS was transmitting vector signal to one MS equipped with M antennas in a single TS, in Fig. 4.4 we have a vector IRI signal of size M created by M different RSs. With this connection between systems in the present section and in Section 4.1, we can directly build the IRI cancellation at the source by considering properly modified inter-relay channel gain matrices where in the present section we actually should be working with the virtual relay to the relay channel gain matrices. Specifically, in the present section, when the IRI cancellation is not implemented yet at the BS, the received signal at RS_k during $TS = k$ is:

$$Y_{R_k} = \sqrt{P_{BS}}\mathbf{A}_k\mathbf{s}_k + \sqrt{P_{R_{k-1}}}\mathbf{H}_k^{k-1}s_1^{k-1} + \sqrt{P_{R_{k-2}}}\mathbf{H}_k^{k-2}s_2^{k-2} + \sqrt{P_{R_{k-3}}}\mathbf{H}_k^{k-3}s_3^{k-3} + \mathbf{n}_{R_k} \quad (4.13)$$

or rather

$$Y_{R_k} = \sqrt{P_{BS}}\mathbf{A}_k\mathbf{s}_k + \Sigma_{i=1}^{M}\sqrt{P_{R_{k-i}}}\mathbf{H}_k^{k-i}s_i^{k-i} + \mathbf{n}_{R_k} \quad (4.14)$$

where $\mathbf{H}_k^{k-i}$ is the $M \times 1$ channel vector from RS_{k-i} (actually its ith antenna) to RS_{k+1}. We should point out here that indexing of relays is using mod L arithmetic. The interference vector $\Sigma_{i=1}^{M}\sqrt{P_{R_{k-i}}}\mathbf{H}_k^{k-i}s_i^{k-i}$ is known to the BS and the BS can cancel the effects of these signals at RS_k by sending an interference cancellation signals along with the desired signals to RS_k. The transmitted signals by the BS then becomes:

$$X_{BS} = \sqrt{P_{BS}}\mathbf{s}_k - \mathbf{A}_k^{-1}\Sigma_{i=1}^{M}\sqrt{P_{R_{k-i}}}\mathbf{H}_k^{k-i}s_i^{k-i} \quad (4.15)$$

. The multiplication by $\mathbf{A}_k^{-1}$ is to remove the effects of the channel attenuation $\mathbf{A}_k$ between the BS and RS_k.

Even though the distance between RS_1 and RS_3 in Fig. 4.4 appears to be equal to the distance between the BS and MS, we assumed that the positions of the RSs are designed to maintain good channel equality with the BS. Therefore, we account here for the possibility that the channel $\mathbf{H_{31}}$ between RS_3 and RS_1 could carry some sort of interference between the two. If this channel is weak enough, then the BS could neglect the IRI between these distantly spaced RSs. Similar argument is applied to the IRI between RS_0 and RS_2.

For completeness, the second phases divide the network into M one-to-one networks. These networks are separated by sufficient distance to disregard the interference of each individual network (between different RSs and their assisted MSs) on the other. This is argued as the distance to non-associated MSs from the RSs is greater than from the BS which we already assumed has negligible channel gains with the MSs. If interference exists, however, it will affect the BER performance as a flooring effect. That is, RSs have to maintain power control to reach the optimal BER performance.

Next, we present the uplink transmission process in the network considered in this subsection with single antenna MSs to offer a full solution (downlink and uplink) to IRI problems in the underlying cellular network.

4.2.3 IRI Cancellation on an Uplink

The scheduling of uplink transmissions from the MSs to the BS differs from the scheduling in the downlink discussed in Section 4.2.2. The setup of the network in the downlink falls under the umbrella of the one-to-many broadcast channel. If the CSI is not available at the transmitting node, the transmission has to to be performed blindly (no pre-coding) which collapses the network to one-to-one channels with single data stream per time slot. Thus, we have solved this problem through proper transmission scheduling and decoupling the network into multiple one-to-one networks with tolerable interference.

In the uplink transmission, the network falls under the category of many-to-one multiple access channels. Multiple number of single antenna MSs communicate simultaneously in a given TS with a single RS equipped with the antenna array where the number of antennas equals to the number of the MSs. These subnetworks of MSs with the RS acting as data collectors are transmitting in their dedicated TSs and all of these subnetworks are then communicating with a single BS. Therefore, the communications from MSs to the destination at the BS is divided into two networks. The first network consists of multiple and identical many-to-one networks between MSs and RSs. Even though these multiple networks seems decoupled and a reveres process of the downlink could be adopted in this part of the network, this approach is not the favorable since all of the data would have to be queued in the middle hubs (RSs) of the network. Adopting this approach would only add more MS-RS interference to the network. Alternatively, these networks are stacked in a queue and the M MSs that are associated with RS_k have to send their messages during time slot $TS = k$. This is visualized in Fig. 4.5 during $TS = 0$ when MS_i^0 $(i \in \{1, 2, \cdots, M)$ in green send signals to RS_0 in a designated $TS = 0$. Thus, RS_k has to listen to its MSs only during $TS = k$. That is, M MSs associated with RS_k are grouped in a "super MS with M antennas" and the received signal (in a vector form) at RS_k during $TS = k$ is:

$$Y_{R_k} = \Sigma_{m=1}^{M} \sqrt{P_{MS_m}} \mathbf{A}_{km} s_{km} + \mathbf{n}_k \tag{4.16}$$

where $\mathbf{A}_{km}$ is a $M \times 1$ channel gain vector between the single antenna MS_k^m and RS_k with M antennas while s_{km} is a scalar representing the signal from MS_m. Relation in (4.16) can be expressed in a matrix form assuming "the super MS with virtual array" as:

$$Y_{R_k} = \sqrt{P_{MS}} \mathbf{A}_k \mathbf{s}_k + \mathbf{n}_k \tag{4.17}$$

The received signal in (4.17) does not account for the interference that is introduced on the second hop when transmitting from RSs to the BS as discussed next.

The second part of the network consists of cascaded one-to-one networks between the RSs and BS. In this network the number of antennas at the transmitters -RS- equals to the number of antennas at the receiver or BS. With this setup, the channels

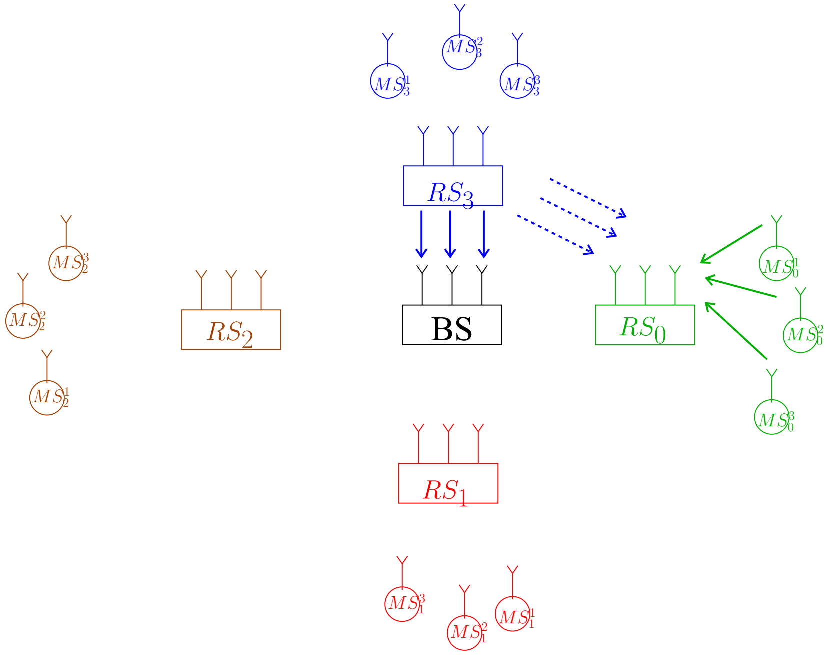


Figure 4.5: IRI effects in single antenna networks on the uplink where signals arrive to the RS through two paths:(i) direct path from assigned MSs and (ii) the other relay path.

at both sides of the RSs are utilized simultaneously. That is, while RS_k is listening to its MSs during $TS = k$, RS_{k-1} is forwarding its received signals using the same medium which causes the inter-relay interference we have been addressing in Chapter 4. This is represented during $TS = 0$ in Fig. 4.5 where while RS_0 is receiving its associated vector signal in green, RS_3 sends its vector signal to the BS causing IRI at RS_0 which represented as dashed blue arrows.

Note that the BS is receiving all the data from the RSs and therefore can utilize the scheme developed for IRI cancellation at the receiver in Chapter 3. That is, in order for the BS to extract the signals transmitted by the MSs associated with RS_{k+1}, it uses the signals received during $TS = k$ and $TS = k + 1$ and filters off the IRI before it passes the vector signal to the decoder.

4.3 Performance Analysis

In this section, we evaluate the performance of the proposed IRI cancellation at the transmitter. We first discuss the trade off between time saving and the power ratio invested to maintain certain level of BER. Then we discuss the impact of relay spacing on the BER performance of the scheme.

4.3.1 Performance of the IRI Cancellation at the Transmitter

In order to evaluate the performance of the proposed IRI cancellation scheme, we simulate a system with all nodes equipped with two antennas ($M = 2$) and $K = 4$ MSs unless stated otherwise. The transmitter sends pre-coded BPSK signals to the relays and we examine the BER versus total transmit power to noise ratio as the receivers perform maximum likelihood detection. In this scheme, the BS uses approximately two third of the total transmit power because it transmits two signal components to the RSs: one component is the information-bearing signal for the desired path towards the destination MS and the second component is to cancel the IRI at the affected RS.

The nature of the network separates the RSs at least by $\sqrt{2}d$ where d represents the distance between one RS and the BS. In order to account for the distance separation, we set the variance of the complex channel gain coefficient between two adjacent RSs to $\frac{1}{\sqrt{2}\sigma_A^2}$ which is $\sqrt{2}$ times smaller than the variance of the complex channel gain coefficient between the BS and any RS.

Figure 4.6 shows the impact of the noise accumulation on the average BER performance at MSs before we deploy the noise flushing procedure. The BER is plotted as a function of SNR after the completion of the first, second and third relaying cycle. A cycle is equal to four relaying turns or equivalently when all RSs completed their

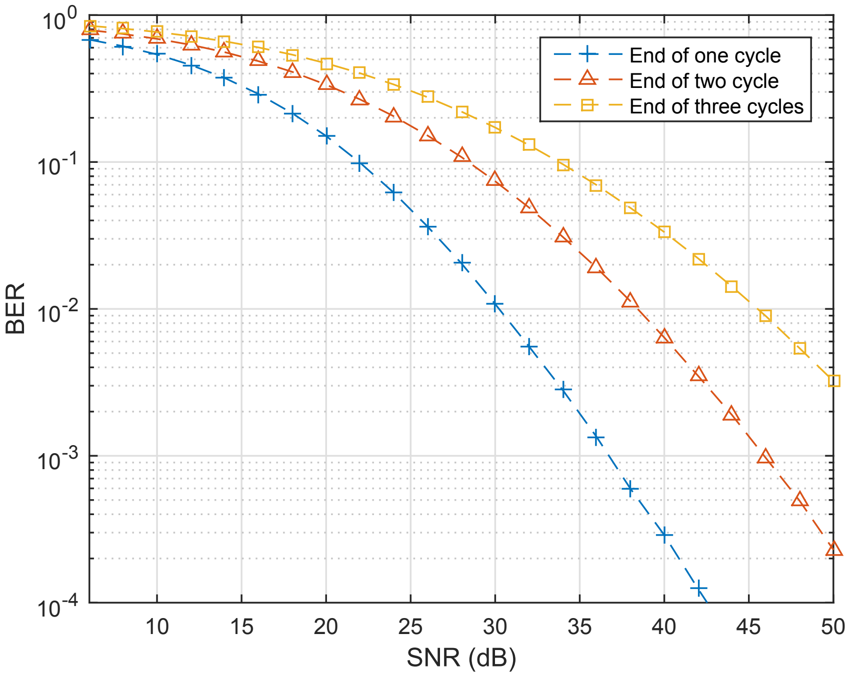


Figure 4.6: The BER vs. SNR performance of the proposed scheme using BPSK after a single cycle, two cycles, and three cycles

turns to relay signals to their destinations. The BER is plotted in different colors as an averaged BER (blue with plus sign for one cycle, red with triangles for the end of the second cycle, and yellow with squares for the end of the third cycles).

By the end of the first cycle, the BER approaches 10^{-3} around SNR of $37dB$, and the same BER performance is obtained by the end of the second cycle at around $46dB$. The performance by the end of the third cycle suffer from high degradation due to the noise accumulated at the RSs. We consider here the worst BER performance which occurs when serving the last user in the cycle. Actually, the average BER is dominated by the last few turns that have worse BER performance because our noise accumulation in consecutive terms is scaled by the contracting multiplicative terms. The average BER in this scenario does not provide an insight about the flushing point which is determined by the fall of the signals quality below a certain threshold.

The time saving by the end of the first cycle is equal to 30% to send the same messages as compared to the conventional or no alternative relaying scheme. This is equivalent to the prelog factor of $\frac{4}{5}$ versus $\frac{1}{2}$ in the conventional scheme. By the end of the second cycle, the time efficiency increases to $\frac{8}{9}$ which is an increment of only 8% on top of the saving after the end of the first cycle. Therefore, selecting the number of cycles when to perform flushing it is a trade off in the performance between the time saving (bandwidth efficiency) and the required SNR to achieve the targeted BER.

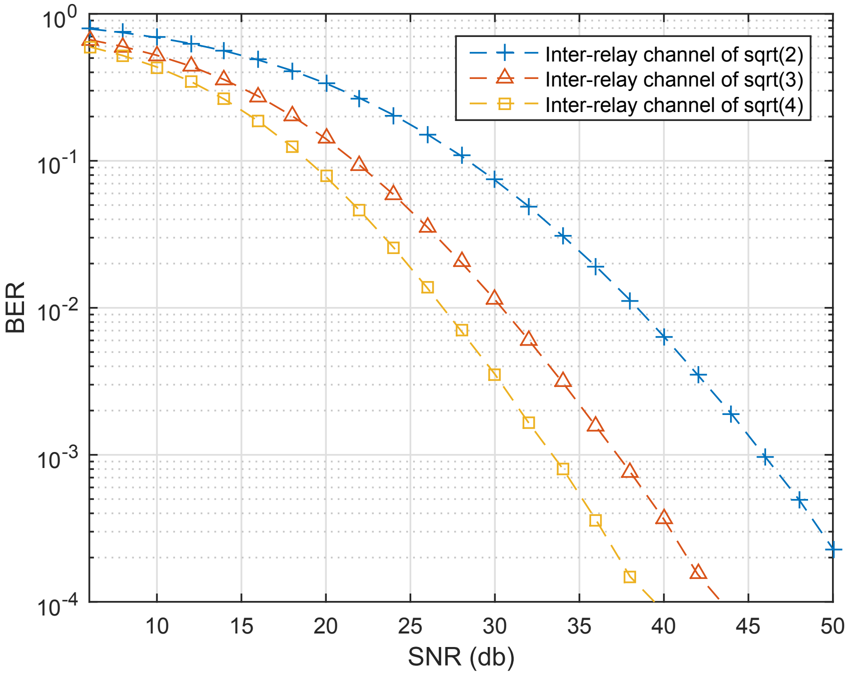


Figure 4.7: The BER performance vs. SNR of the proposed scheme for different spacing between RSs contributing to IRI by the end of the 8^{th} relaying turn.

Moreover, the number of RSs along the RS ring plays a key role in determining the performance of the system. Fig. 4.6 showed only the performance when the distance between the two successive relays is $\sqrt{2}d$ which corresponds to the four RSs spaced evenly. However, having more distributed relays along the RS ring helps serving more boundary users as well as provide the BS with some freedom to scheduling the BS/RS transmissions in order to improve the system performance. It should be understood here that the more RSs on the same ring, the smaller the spacing between the geographically adjacent relays. Saying that if we had large ("infinite") number of relays on a ring of radius d, when transmitting from the BS would not have two arrange the consecutive transmissions from the BS using geographically adjacent RSs as this would result in very strong inter-relay channels contributing to IRI. In this case, a better option would be to use to switch transmissions between RSs located along diagonals and every two TSs move on to the geographically adjacent RSs from the RS used two TSs before. This would result in a good distance separation between two RSs contributing to IRI of $2d$. Figure 4.7 shows how the BER performance improves by going from the spacing between RSs contributing to IRI of $\sqrt{2}d$ to a spacing of $\sqrt{3}d$ or even $\sqrt{4}d$. The results presented in Fig. 4.7 show the performance by the end of 8 relaying turns. The improvements in the BER performance follow the low of diminishing returns when increasing the distance between RSs contributing to IRI from $\sqrt{2}d \simeq 1.4d$ through $\sqrt{3}d \simeq 1.7d$ to $2d$.

4.4 Summary

In this chapter, we presented IRI cancellation schemes as implemented at the source, e.g., the BS in a single cell cellular networks. The need for for the IRI cancellation result from the reuse of TSs which one one hand leads to high bandwidth efficiency as measured through number of delivered messages per time slot and on the other hand to IRI which, if not mitigated, may may limit the BER performance.

Two network configurations were analyzed: the first one where MSs are equipped with the same number of antennas as the BS and the relays; and the second one for single antenna users. The BS has prior knowledge of the interference since all the transmitted signals originate at this source. As a result, the BS can send pre-coded signals to cancel the IRI along with the information bearing signal. By deploying precoding at the BS only, the proposed IRI cancellation for successive AF relaying in the first network configuration has the ability to fully cancel the inter-relay interference. The scheme sufferers from AWGN noise accumulation and solutions to deal with this problem have been proposed.

In the successive relaying networks with single antenna MSs on a downlink with the proposed time scheduling of transmissions, we made a critical observation that the IRI in a given TS is caused by the virtual antenna array where the individual antennas are from RSs receiving from the BS on the previous TSs. With this observation, we adopted the IRI cancellation scheme on a downlink developed in the first part of this chapter for the network with all nodes having the same number of antennas. To avoid noise accumulation we were using in the single antenna MSs network relays operating in the DF mode. To offer a complete solution (downlink and uplink) to IRI problems in networks with single antenna MSs, we also presented the uplink transmission scheme and the IRI cancellation by adopting here the scheme developed in Chapter 3 for the IRI cancellation at the destination (in this case the BSs).

The source based IRI cancellation schemes developed in the chapter relied on the knowledge of channels gain matrices related to the inter-relay channels and channels between the BS and RSs. learning about the relevant CSI could be accomplished by sending the training sequences and examining the effects at different points in a network along the data paths and the crosstalk paths. This may impose a limitation on the proposed schemes in fast changing channel environments. Since the BS and RSs have more processing power than the MSs, learning the relevant CSI at the BS and the RSs is a realistic task. The design of training vector sequences to learn the CSI at the receivers was not in the scope of this dissertation.

Chapter 5

Conclusions and Future Work

The main objective of the research in this dissertation was to increase the data rates in MIMO wireless networks with successive relaying. This objective was met by interference control through time scheduling of transmissions and relay interference cancellation. In this chapter, the contributions of this dissertation are summarized in Section 5.1, followed in Section 5.2 by suggestions for possible future work based on the results from this research.

5.1 Dissertation Summary

One of the fundamental drivers of communications and networking research is the goal of increasing communication rates over limited system bandwidth. The field of multihop transmissions considered in this dissertation is a special case of cooperative communications, which is emerging as a promising technology offering significantly higher spectral efficiency and reliability in wireless networks than is provided by conventional methods. Of particular interest are MIMO two-hop channels where amplify-and-forward (AF) relay terminals assist in the communication between the source and destination terminals. The research undertaken here focuses on signal processing aspects, where the distributed transmission and processing of messages combined with system-level control of time slotted transmissions contributes to optimized wireless system performance. Specifically, this dissertation provides potential solutions to enhance the efficiency of wireless networks with successive relaying by increasing the number of multiplexed data signals within a limited spectrum. Successive relaying overcomes losses in spectral efficiency by a factor of two, where the losses result from practical relays operating in half-duplex mode. Reusing the same spectrum with simultaneous multiple relay transmissions in successive relaying increases spectrum efficiency, however it also causes interference which, if not mitigated, may limit the overall wireless network performance. To deal with the relay interference, this dissertation first (i) proposes relaying strategies based on the time scheduling of transmissions and spatial separation between relays with MSs and BSs reusing the same spectrum, and then (ii) develops interference cancellation algorithms. Different network configurations are considered, with MSs equipped with multiple and single antennas. The existence (and lack) of a line of sight between the BS and MSs is incorporated into the proposed algorithms to increase the number of messages transmitted in the network in one time slot. Based on the CSI available at different nodes in the network and assuming the AF operation of relays, source- and destination-based inter-relay interference cancellation schemes have been developed. The common feature of these algorithms is that they exploit spatial spectrum reuse, opportunistic listening, and broadcast characteristics of wireless channels to cancel different types of interference. Uplink and downlink forwarding strategies are analyzed. The proposed solutions incorporate the time slotting of transmissions and the design of efficient algorithms to manage the IRI for universal frequency reuse. The results presented in this dissertation form a framework for the development of future networks, with successive relaying advancements integrating system-level and signal processing algorithmic designs, with a particular focus on time slot utilization. The author believes that exploration of the integration of signal processing algorithms with other system-level resources (such as power) within the scope of coordinated communications could lead to wider acceptance of cooperative networking in practical deployments of emerging classes of wireless networks, such as ad hoc and sensor networks and cellular networks with multiple hops.

The contents of the individual chapters of this dissertation and the research results can be summarized as follows:

Chapter 2 deals with single-antenna MSs in a cellular cooperative network, where the BS is equipped with M antennas. In this network, the MSs are assisted by a group of single-antenna relays (cooperative MSs that are not actively receiving their own data). The messages from the BS to the active MSs are delivered in two stages. In the first stage of a communication cycle, the proposed scheme benefits from the broadcast channel characteristic of the wireless medium, while in the second stage, the AF relays assist the receiver to construct a linear system of equations, so that each receiver can receive a total of M messages per communication cycle. In the second stage, MSs together with their associated cooperative MSs form a one-to-one network to receive multiple copies of the transmitted signals. Having multiple networks using the same spectrum at the same time causes interference at the receiving MSs. In Chapter 2, the problem of interference is solved through clustering and distance separation, to keep the interference below a certain threshold. The proposed system increases the number of multiplexed signals per time slot to $\frac{MN}{M+N-1}$ (where N is the number of active MSs in a communication cycle), as compared to one (1) in conventional TDMA systems.

Chapter 3 analyzes inter-relay interference in MIMO wireless networks with successive relaying. In these networks, the spectrum is utilized simultaneously by the source and one of the relays, causing what is known as inter-relay interference (IRI). This chapter develops a signal processing scheme that fully cancels the inter-relay interference at the receivers by exploiting opportunistic listening. The term opportunistic listening is used to refer to receiver operation where a destination listens to undesired signals in order to use them in the interference cancellation process. The proposed IRI cancellation scheme is not limited to a particular type of network, and is applied in various relay systems with two-hop forwarding, such as $1 \times 2 \times K$ and $K \times K \times K$.

The application of successive relaying and IRI cancellation at the destination is also generalized from downlink (unidirectional) data flow to multiway communications among three users equipped with multiple-input multiple-output (MIMO) transceivers that exchange messages with one another via two amplify-and-forward (AF) co-located relays.

Chapter 4 first proposes a successive AF relaying scheme that has the ability to cancel the inter-relay interference fully by deploying precoding at the BS only. The scheme benefits from the broadcast characteristic of the radio channel, and from knowledge at the BS of all the data contributing to inter-relay interference. The scheme suffers from AWGN noise accumulation, and solutions to deal with this problem are proposed.

While the first part of Chapter 4 deals with MSs equipped with M antennas, the second part extends the developed system model and IRI cancellation to support MSs equipped with a single antenna each. Here, work is done with downlink and uplink transmissions. Both cases benefit from the ability to collapse single-antenna MSs network configurations onto previously analyzed models by working with virtual antenna array concepts. Considerable improvements in data stream multiplexing are demonstrated, from 1 in conventional TDMA systems to M with the use of successive relaying, where M is the number of antennas at the BS and the relays.

5.2 Suggested Future Work

Many open problems have yet to be addressed in two-hop wireless relay networks. Some of these problems arise in connection with proposed new solutions using cooperative communications, while other problems are generic to all signal processing algorithms related to relaying systems, such as channel estimation and synchronization. In this dissertation, all algorithms and methods were analyzed first using mathematical tools and then verified through computer simulations. The main challenge facing new theoretical results in wireless network research is to find a path to apply these designs in practical systems.

The following topics related to the present study are suggested for future work.

1.Effects of Imperfect Channel Information

In the research presented in this dissertation, it has been assumed that perfect CSI is available at least at one node in the network, when applying relay interference cancellation. However, the availability of perfect CSI may be difficult to implement in practice, since knowledge about the channel is obtained through statistical estimation and feedback between the nodes. The problem of imperfect CSI in different network settings has been extensively studied and many algorithms to address this issue have been proposed in the literature [52]–[54]. Performance deterioration due to imperfect CSI in the developed algorithms could be investigated.

2.Synchronized Transmissions and Reception

An important challenge to IRI cancellation at the transmitter is the difficulty of synchronizing the arrival of the signals transmitted from the BS and the interference signals from the interfering relays. There are methods of achieving synchronous transmissions of the networking nodes, however in this thesis the synchronous reception of signals from terminals at different locations is required. Because of the variability in the propagation delays of the signals, this may pose a challenge, especially at high data rates. A potential solution here is to deploy orthogonal frequency-division multiplexing (OFDM), which extends the transmitted symbols duration and may help to alleviate the impact of signals arriving with different delays. This could be a subject of further study.

3.IRI Cancellation at Single-Antenna Relays

In the parts of this dissertation that study successive relaying, it is assumed that relays are always equipped with the same number of antennas as the node that performs the IRI cancellation. Analyzing networks where relays are equipped with a smaller number of antennas, and the impact of using multiple relays equipped with single antennas is an interesting problem, since it could contribute to making the IRI cancellation scheme more flexible in choosing from a bigger pool of relays so as to benefit from relay diversity.

4.Improving SNIR for Single-Antenna Users

In the investigation of relay-assisted downlink transmissions to support increased data rates for single-antenna users in Chapter 2, a factor limiting the scheme performance is interference from simultaneously transmitting “adjacent” relays. However, successive interference cancellation in a comparable system in Chapter 4 achieves further improvement in bandwidth efficiency. Merging both approaches to reduce the SNIR through spatial separation of interfering relays is a topic worthy of further investigation.

Bibliography

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* New York, NY, USA: Cambridge University Press, 2005.

[2] D. Bliss, K. Forsythe, and A. Chan, "MIMO wireless communication," *Linconlan Laboratory Journal*, vol. 15, no. 1, pp. 97–126, 2005.

[3] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-advanced," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 127–135, 2013.

[4] Z. Ding, S. Dasgupta, and R. Lopez, "Interference cancellation and blind equalization for linear multi-user systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP '00*, vol. 1, 2000, pp. 145–148.

[5] S. K. S. Yoshioka and F. Adachi, "Joint Tx/Rx signal processing for uplink distributed antenna network using single-carrier MU-MIMO," in *2015 10th Int. Conf. on Inform., Commun. and Signal Process. (ICICS)*, Dec 2015, pp. 1–5.

[6] S. Jafar, "Blind interference alignment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 3, pp. 216–227, June 2012.

[7] N. Lee and R. W. H. Jr., "Degrees of freedom for the two-cell two-hop MIMO interference channel: Interference-free relay transmission and spectrally efficient relaying protocol," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2882–2896, May 2013.

[8] A. Abrardo, G. Fodor, and B. Tola, "Network coding schemes for device-to-device communications based relaying for cellular coverage extension," in *2015 IEEE 16th Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, June 2015, pp. 670–674.

[9] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec 2004.

[10] V. Havary-Nassab, S. Shahbazpanahi, and A. Grami, “Optimal distributed beamforming for two-way relay networks,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1238–1250, March 2010.

[11] A. Paulraj, D. Gore, R. Nabar, and H. Bolcskei, “An overview of MIMO communications - a key to gigabit wireless,” *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[12] Y. Jing, “A relay selection scheme for two-way amplify-and-forward relay networks,” in *Int. Conf. on Wireless Commun. Signal Proc., WCSP*, Nov 2009, pp. 1–5.

[13] C.-X. Wang, X. Hong, X. Ge, X. Cheng, G. Zhang, and J. Thompson, “Cooperative MIMO channel models: A survey,” *IEEE Commun. Mag.*, vol. 48, no. 2, pp. 80–87, February 2010.

[14] A. Scaglione, D. Goeckel, and J. Laneman, “Cooperative communications in mobile ad hoc networks,” *IEEE Signal Proc. Mag.*, vol. 23, no. 5, pp. 18–29, Sept 2006.

[15] B. Rankov and A. Wittneben, “Spectral efficient protocols for half-duplex fading relay channels,” *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 379–389, February 2007.

[16] N. Zlatanov, V. Jamali, and R. Schober, “On the capacity of the two-hop half-duplex relay channel,” in *2015 IEEE Globe. Commun. Conf. (GLOBECOM)*, Dec 2015, pp. 1–7.

[17] F. Alhumaidi and J. Ilow, “Relay-assisted downlink transmissions to support increased data rates for single antenna users,” in *2014 IEEE Globe. Commun. Conf.(GLOBECOM)*, Dec 2014, pp. 4150–4155.

[18] ——, “Transmitter precoding to cancel inter-relay interference in AF systems with successive transmissions,” in *2016 IEEE 29th Canadian Conf. on Elect. and Comput. Eng. (CCECE)*, May 2016, pp. 60–64.

[19] ——, “Alternate AF MIMO relaying systems with full inter-relay interference cancellation,” in *Veh. Technol. Conf. (VTC Fall), 2015 IEEE 82nd*, Sept 2015, pp. 1–5.

[20] ——, “Increasing throughput in multi-way three-user MIMO networks using successive relaying and IRI cancellation,” in *2016 IEEE 12th Int. Conf. on Wireless and Mobile Comput., Networking. and Commun. (WiMob),*, 2016.

[21] F. Xiong, *Digital Modulation Techniques*, ser. Artech House telecommunications library. Artech House, 2000.

[22] J. Linnartz, *Narrowband land-mobile radio networks.* Artech House, 1993.

[23] H.-W. Chang and C.-W. Yu, "Optimizing one-to-one data communication on double-loop networks," in *The 2004 IEEE Asia-Pacific Conf. on Circuits and Syst. Proc. (APCCAS)*, vol. 1, Dec 2004, pp. 473–476 vol.1.

[24] C. Huang and S. Cui, "On the alternative relaying gaussian diamond channel with conf. links," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 758–768, February 2013.

[25] L. Lai and L. Huie, "Simultaneously generating multiple keys in many to one networks," in *2013 IEEE Int. Symp. on Inform. Theory Proc. (ISIT)*, July 2013, pp. 2394–2398.

[26] J. Pope and R. Simon, "Efficient one-to-many broadcasting for resource-constrained wireless networks," in *2015 IEEE 40th Conf. on Local Comput. Networks (LCN)*, Oct 2015, pp. 518–525.

[27] W. Han, A. Liu, and V. K. N. Lau, "Improving the degrees of freedom in MIMO interference network via PHY caching," in *2015 IEEE Globe. Commun. Conf. (GLOBECOM)*, Dec 2015, pp. 1–6.

[28] A. Goldsmith, *Wireless Communications.* Cambridge University Press, 2005.

[29] M. Krl, Y. Ji, S. Yamada, C. Borcea, L. Zhong, and K. Takano, "Extending network coverage by using static and mobile relays during natural disasters," in *2016 30th Int. Conf. on Advanced Inform. Networking and Applicat. Workshops (WAINA)*, March 2016, pp. 681–686.

[30] S. Zhang, X. G. Xia, and J. Wang, "Cooperative performance and diversity gain of wireless relay networks," *IEEE J. Select. Areas Commun.*, vol. 30, no. 9, pp. 1623–1632, October 2012.

[31] S. Zhao, M. Peng, and W. Wang, "An adaptive relay selection algorithm for optimizing diversity and multiplexing gains in wireless cooperative networks," in *2010 IEEE Wireless Commun. and Network. Conf. (WCNC)*, April 2010, pp. 1–6.

[32] D. Hwang, D. I. Kim, S. K. Choi, and T. J. Lee, "UE relaying cooperation over D2D uplink in heterogeneous cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4784–4796, Dec 2015.

[33] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. R. Liu, "Cooperative communications with relay-selection: when to cooperate and whom to cooperate with?" *IEEE Trans. Commun.*, vol. 7, no. 7, pp. 2814–2827, July 2008.

[34] H. J. Lee, S. H. Kang, and J. H. Kim, "A cooperative scanning mechanism for the mobile relay in the moving network environment," in *2010 7th IEEE Consumer Commun. and Networking Conf.(CCNC)*, Jan 2010, pp. 1–5.

[35] W. Sang, D. Shen, W. Ren, and X. Shuai, "A survey of capacity in cooperative relay networks," in *2011 Global Mobile Congress (GMC)*, Oct 2011, pp. 1–8.

[36] Q. T. Vien, H. X. Nguyen, and W. Tu, "Optimal relay positioning for green wireless network-coded butterfly networks," in *2013 IEEE 24th Annu. Int. Symp. on Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Sept 2013, pp. 286–290.

[37] E. Yilmaz, R. Knopp, and D. Gesbert, "On the gains of fixed relays in cellular networks with intercell interference," in *2009 IEEE 10th Workshop on Signal Process. Advances in Wireless Commun.(SPAWC)*, June 2009, pp. 603–607.

[38] B. Sunil and N. B. T, "Performance analysis and comparison of AF and DF relaying systems in rayleigh fading channel considering poisson interference field," in *2015 Int. Conf. on Innovations in Inform. Embedded and Commun. Syst. (ICIIECS)*, March 2015, pp. 1–5.

[39] M. Yu, J. Li, and H. Sadjadpour, "Amplify-forward and decode-forward: the impact of location and capacity contour," in *2005 IEEE Military Commun. Conf.(MILCOM)*, Oct 2005, pp. 1609–1615 Vol. 3.

[40] B. D. Moubagou and Y. Y. Chang, "Performance of AF and DF relays by employing half duplex and full duplex mode," in *2013 Int. Conf. on Measurement, Inform. and Control (ICMIC)*, vol. 02, Aug 2013, pp. 1379–1383.

[41] C. Geng and S. A. Jafar, "On the optimality of zero-forcing and treating interference as noise for k-user MIMO interference channels," in *2016 IEEE Int. Symp. on Inform. Theory (ISIT)*, July 2016, pp. 2629–2633.

[42] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k -user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp. 3425–3441, Aug 2008.

[43] T. Gou and S. A. Jafar, "Degrees of freedom of the k user $m \times n$ MIMO interference channel," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6040–6057, Dec 2010.

[44] Z. Sun, L. Yang, J. Yuan, and M. Chiani, “A novel detection algorithm for random multiple access based on physical-layer network coding,” in *2016 IEEE Int. Conf. on Commun. Workshops (ICC)*, May 2016, pp. 608–613.

[45] Q. Y. Liau and C. Y. Leow, “Study of relay position in two-path successive relaying with interference cancellation,” in *2014 IEEE Asia Pacific Conf. on Wireless and Mobile*, Aug 2014, pp. 254–259.

[46] O. Munoz, A. Agustin, and J. Vidal, “Cellular capacity gains of cooperative MIMO transmission in the downlink,” in *2004 Int. Zurich Seminar on Commun.*, 2004, pp. 22–26.

[47] C. Luo, Y. Gong, and F. Zheng, “Full interference cancellation for two-path cooperative communications,” in *Wireless Commun. and Networking Conf., WCNC*, April 2009, pp. 1–5.

[48] K.-H. Park and M.-S. Alouini, “Alternate MIMO relaying with three AF relays using interference alignment,” in *IEEE Int. Conf. on Commun. (ICC)*, June 2012, pp. 3526–3531.

[49] H. Park and J. Chun, “Inter-relay interference cancellation for AF MIMO two-path relay systems,” in *Veh. Technol. Conf. (VTC Fall)*, Sept 2012, pp. 1–5.

[50] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2009.

[51] T. Kaczorek, *Linear Control Systems: Analysis of Multivariable Systems.* New York, NY, USA: John Wiley & Sons, Inc., 1992.

[52] P. Aquilina and T. Ratnarajah, “Performance analysis of IA techniques in the MIMO IBC with imperfect CSI,” *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259–1270, April 2015.

[53] X. Chen and C. Yuen, “On interference alignment with imperfect CSI: Characterizations of outage probability, ergodic rate and SER,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 47–58, Jan 2016.

[54] X. Yin, X. Yu, Y. Liu, W. Tan, and X. Chen, “Performance analysis of multiuser MIMO system with adaptive modulation and imperfect CSI,” in *IET Int. Conf. on Inform. and Commun. Technol. (IETICT)*, April 2013, pp. 571–576.