

Capacity Enhancements to Wireless
MIMO Systems: Channel Feedback,
Keyhole Mitigation and Multi-User Support

August 2011

Oussama Souihli

A Thesis for the Degree of Ph.D. in Engineering

Capacity Enhancements to Wireless
MIMO Systems: Channel Feedback,
Keyhole Mitigation and Multi-User Support

August 2011

Graduate School of Science and Technology
Keio University

Oussama Souihli

Contents

List of Tables	v
List of Figures	vi
Abstract	viii
Acknowledgments	x
1 Introduction	1
1.1 Notations	3
1.2 The Communication Problem: A Capacity Perspective	4
1.2.1 The Channel Coding Theorem	6
1.2.2 Traditional Ways of Increasing Channel Capacity	9
1.2.3 Emergence of MIMO Technology: A Brief History	11
1.3 MIMO Technology: Key Concepts	12
1.3.1 System Model	12
1.3.2 Open-Loop MIMO Vs Closed-Loop MIMO	16
1.3.3 Full / Limited / Quantized CSI	19
1.3.4 Single-User MIMO Vs Multi-User MIMO	21
1.4 On The Channel Capacity of MIMO Systems	23
1.4.1 The Single-User Case	23
1.4.2 The Multi-User Case	25
1.5 Thesis Outline and Summary of Results	31
2 On Full CSI Feedback in Closed-Loop MIMO	37
2.1 Introduction	37
2.2 On The Coherence Time of A Wireless MIMO Channel	38

2.2.1	The Doppler Effect	39
2.2.2	Why Is It Necessary That The Fed-Back CSI Matches The True Channel In Closed-Loop MIMO ?	39
2.2.3	Then, How Often Should The CSI Be Fed Back To The Transmitter ?	40
2.2.4	Motivation of Our Research: The Need For A Fast CSI Feedback	40
2.3	System Model And Preliminaries	41
2.3.1	System Model	41
2.3.2	Echo-MIMO: An Overview	44
2.3.3	Comments on Echo-MIMO	45
2.4	Proposed Scheme	45
2.4.1	Transparent Inband Feedback (TIF)	46
2.4.2	TIF vs Echo-MIMO, A Comparative Study	47
2.5	Numerical Examples	50
2.5.1	Evaluation of the Estimation Accuracy	51
2.5.2	Evaluation of The Spectral Efficiency	52
2.6	Conclusion	54
3	On Relay-Assisted Keyhole Mitigation	55
3.1	Introduction	55
3.2	System Model and Preliminaries	58
3.2.1	System Model	58
3.2.2	Preliminaries	58
3.2.3	On Decode-and-Forward (DF) Relaying	60
3.3	On Keyhole Effect Mitigation	62
3.3.1	Problem Statement	62
3.3.2	The Fixed Channel Case	62
3.3.3	The Stochastic Channel Case	65
3.4	Numerical Examples	68
3.5	Conclusion	70
4	Joint Scheduling and Feedback In MU-MIMO	71
4.1	Introduction	71
4.2	System Model	75

4.2.1	Traffic Model	75
4.2.2	Channel Model	75
4.3	A QoS-aware Channel Estimation and Feedback Scheme	76
4.3.1	Proposed Scheme	76
4.3.2	On The Optimal Feedback Threshold ϵ Under Gaussian Approximation	78
4.4	On The Required Number of Streams to Meet a Delay Constraint	80
4.4.1	In the Absence of Interference	81
4.4.2	In the Presence of Interference	82
4.5	Proposal of a delay-aware scheduling scheme	82
4.5.1	Need for a Delay-Aware Scheduling Scheme	83
4.5.2	Proposed Delay-Aware Scheduling Scheme	84
4.6	Simulation Results	86
4.6.1	On Incurred Delays	86
4.6.2	On Achievable Throughputs	86
4.7	Conclusion	87
5	Conclusions And Perspectives	89
	Appendix A Proof of Proposition 1	91
	References	95

List of Tables

1.1	Outline of Chapter 2	34
1.2	Outline of Chapter 3	35
1.3	Outline of Chapter 4	36
2.1	Main Simulation Parameters.	41
2.2	Precoding weights at the transmitter and the receiver for different precoding schemes.	53

List of Figures

1.1	Industry mobile data forecasts.	2
1.2	FCC estimate of mobile traffic by device type and forecasts.	2
1.3	FCC projected utilization of wireless spectrum.	3
1.4	A point-to-point communication system from an information-theoretical perspective.	4
1.5	An illustrative example of a wireless propagation environment and the typical attenuations that the wireless signal can be subject to.	5
1.6	MIMO system model.	13
1.7	Parallel decomposition of the MIMO channel	14
1.8	A general overview of MIMO systems.	16
1.9	Phases that make up closed-loop MIMO.	18
1.10	Various MU-MIMO scenarios.	22
1.11	An example of successive decoding with $K = 4$ inputs	27
1.12	An overview of the capacity-achieving power allocation strategies in MIMO systems.	30
1.13	Approaches towards MIMO capacity enhancement considered in this thesis.	32
1.14	Thesis main outline.	33
2.1	Phases brought into play in Closed-Loop MIMO during channel's coherence time	38
2.2	One second in the life of four wireless channels. The Doppler effect for different receive mobility patterns.	42
2.3	Comparison of the block-fading model and the true channel realizations	43
2.4	Proposed two-way communication scheme for CSI feedback in Closed-Loop MIMO	48
2.5	Comparison of the time slots of Echo-MIMO and TIF in terms of transmit power and number of transmitted pilots	50
2.6	Channel estimation accuracy in terms of NMSE (dB)	51

2.7	Spectral efficiency with perfect and estimated CSI. Different MIMO precoding schemes.	53
3.1	A MIMO (spatial) keyhole scenario	57
3.2	The general MIMO relay channel	59
3.3	Decode-and-forward (DF) relaying.	61
3.4	Comparison of the scaling of C_1, C_2 with N , the number of transmit antennas, for SNR = 5, 10 dB	69
3.5	Comparison of the scaling of C_1, C_2 , when $N = 4, 8$	70
4.1	Comparison of conventional and proposed channel feedback approaches in closed-loop multi-user MIMO.	74
4.2	Targeted number κ of B -users who are likely to feedback their CSI with a probability higher than p_0 , versus the optimal orthogonality threshold ϵ , for different values of p_0	80
4.3	Average number of schedules (time slots) a CBR user is scheduled for transmission within $T = 10$ time slots, if a capacity-maximizing scheduling is used	84
4.4	Proposed delay-aware scheduling algorithm	85
4.5	Comparison of instantaneous delays incurred by A -users' packets when using conventional and proposed scheduling policies.	86
4.6	Average delay versus total number of users (10 of which are A -users), for different numbers of transmit antennas	87
4.7	Average throughput versus SNR when using the conventional and the proposed scheduling approaches.	88

Capacity Enhancements to Wireless MIMO Systems: Channel Feedback, Keyhole Mitigation and Multi-User Support

Oussama Souihli
souihli@ohtsuki.ics.keio.ac.jp.
Keio University, 2011

Supervisor: Prof. Tomoaki Otsuki, Ph.D.
ohtsuki@ics.keio.ac.jp

Abstract

The advance of information and communication technology is perhaps among the most significant breakthroughs that have marked the past couple of decades. We are shifting towards an information society where access to and sharing of information in a timely and reliable manner, irrespective of the location, has become a daily necessity. Such progress has made the demand in data traffic ever increasing, especially with the upcoming of smart phones, tablet PCs and other non-PC networked devices who have further driven the demand trend.

Recently, multiple-input multiple-output (MIMO) technology, the use of multiple antennas at the transmitter and the receiver, has emerged as a potential solution to meet such demand. Early works on MIMO predicted a linear growth in capacity with the number of antennas, which could allow for unprecedented wireless data rates. Regrettably, some theoretical requirements and assumptions have challenged its implementation in practice.

In this thesis, we provide our (however humble) contributions towards this end by addressing some of these challenges. The study is divided into two parts based on whether the MIMO system is point-to-point (single-user MIMO, SU-MIMO) or multipoint (multi-user MIMO, MU-MIMO).

Following the first chapter where we provide a general introduction to this work, we address MIMO challenges pertaining to single-user systems.

Thus, chapter 2 considers the need for a fast channel estimation and feedback phase. Indeed, as MIMO capacity gains are conditional upon the availability of channel estimates at the transmitter's side, a channel feedback from the receiver ought to be made as quickly as possible (while

ensuring that the feedback channel estimates are sufficiently reliable). A recent work has suggested such delay-free channel feedback, Echo MIMO, where the receiver echoes the received signals *on the fly*. However, it came at a high signaling cost, as two feedback transmissions are required (one for the inward channel, one for the outward channel). Contrarily, we propose a feedback method that preserves the benefits of Echo-MIMO while requiring only one feedback transmission, by judiciously combining the two feedback transmissions by means of mutually-orthogonal precoding matrices.

Subsequently, we consider in chapter 3 the keyhole problem where the propagation environment has a single degree of freedom regardless of the number of transmit antennas, thereby reducing the capacity of a MIMO system to that of a single-input single-output system. Related literature seems to consider such degeneration hopeless. Contrary to this general belief, we show in this chapter that cooperative diversity can mitigate keyhole effects. Precisely, provided that the source-relay channel is keyhole-free, we show that there exists a “cutoff” relay transmit power above which keyhole effects can be mitigated *even when both the source-destination and the relay-destination channels incur keyhole effect*. We devise the closed form of this power threshold as function of the source transmit power and the channel matrices brought into play in the relay channel.

Chapter 4 focuses on MU-MIMO systems in which, compared with SU-MIMO systems, a new issue arise: user scheduling. Conventional scheduling approaches rely on the assumption that channel information related to all candidate users is available at the transmitter, so that the latter may pick the optimal users w.r.t. a given performance metric. Quite the opposite, we provide a more efficient feedback approach where *only likely-to-be-scheduled* users feed back their channel information to the transmitter, thereby reducing the number of required feedbacks and the computational burden of exhaustive search for best users at the transmitter’s side. Afterwards, we show that conventional capacity-maximizing scheduling policies fall short to meet the requisites of delay-sensitive applications. Consequently, we provide a QoS-aware scheduling scheme that allows to meet the demand of delay-constrained users.

Eventually, we conclude this dissertation in chapter 5 and provide perspectives and possible extensions to this work.

Acknowledgments

I would like to express my heartfelt gratitude to all those who made it possible for me to finalize this work.

I am deeply indebted to my supervisor, Prof. Tomoaki Ohtsuki, whose availability, support, continuous encouragement and assistance helped me through the ups and downs of my research. I am thankful to him for giving me the freedom to choose my research topics, target conferences/journals at will as well as define my own pace of work. I also owe to Prof. Ohtsuki that he recommended me to the IEICE Communication Society as a student member, and for introducing me to paper technical review, thereby allowing me to serve as technical reviewer for various notable conferences and peer-reviewed journals and have the privilege to sit on the technical program chair (TPC) the IEEE ICC'2011 conference.

Prof. Iwao Sasase, Prof. Yukitoshi Sanada and Prof. Hiroshi Shigeno deserve special thanks as my thesis jury members. I am very thankful to them for helpful comments and suggestions that helped improve the quality of this dissertation.

Further, I also would like to acknowledge the support of the following institutions of my research:

- The Japanese Government (MEXT), for a full PhD scholarship at Keio University. I also would like to thank the Japanese Embassy in Tunisia for recommending me for this scholarship, as well the faculty of the School of Communication of Tunis and Prof. Samuel Pierre from Ecole Polytechnique de Montreal for supporting my candidacy to this PhD program,
- The Global Center of Excellence (GCoE) Program at Keio University, for supporting my conference trips, funding my research and giving me the opportunity to attend joint workshops with sister universities overseas, attend special lectures by reputable researchers and professors and offering me exposure to the faculty of Keio and the chance to make new friendships and network with more colleagues,
- Keio Leading Edge Laboratory (KLL), for offering me research grants on two counts (2009, 2010) to support my research and conference trips,
- The NEC C&C Foundation: for offering me research grants on two counts (2009, 2010) to attend two conferences (IEEE Globecom'2009, IEEE Globecom'2010) to present my research work,
- The Hara Foundation: for offering me a research grant to support my conference trip to IEEE VTC'2009-Fall.

The contributions in this dissertation were further facilitated with the extraordinary support, genuine assistance and exemplary work ethics of the administrative staff of Keio University. Particularly, I would like to express my heartfelt gratitude to the International Center of Keio University, the administrative staff at G-CoE (particularly Ms Yuko Izuta and Ms Maki Adachi), Ms Chinatsu Ichikawa from the Office of Research Administration (ORA) and Ms Hiruzu Ikawa from KLL.

Many thanks go to my friends and colleagues at Ohtsuki lab, at G-CoE (P-03 program, in particular) and my dear friends Mohamed Amin Jabri, Maduranga Liyanague, Alex Fung, Chinnapat Serththin (Parn), Koichi Adachi, Marwan Jilani and Chaima Dhahri for their kind support and dear friendship.

Last but far from least, I would like to thank my family.

Oussama Souihli

Keio University

July 2011

Chapter 1

Introduction

The advance of information and communication technology is perhaps among the most significant breakthroughs that have marked the past couple of decades. We are shifting towards an information society where access to and sharing of information in a timely and reliable manner, irrespective of the location, has become a daily necessity. Such progress has made the demand in data traffic ever increasing. The upcoming of smart phones, tablet PCs and other non-PC networked devices are likely to further drive the trend, as shown in Figs. 1.1 and 1.2 where the forecast in data traffic by industry and the Federal Communications Commission are illustrated [1].

Because physical-layer capacity represents a theoretical upper-bound on the higher-layer capacity, the growth in demand in data traffic would, in all likelihood, drive the need for increased physical-layer capacity. In the particular case of wireless communications that are the focus of this thesis, such need in capacity is expected to be very pronounced owing to the limited capacity that current wireless systems do provide. As an illustrative example, Fig. 1.3 highlights such forecast in demand in wireless bandwidth. Wireless systems are, indeed, capacity-limited owing to a variety of challenges that are inherent to the wireless channel: user mobility, interference, weather conditions, path loss attenuation, shadowing (obstacles), to name a few. Bearing in mind the afore-mentioned demand forecast figures, the need for increased capacity in wireless systems is of the essence.

In this thesis, we provide our, -however humble-, contributions towards this end, by suggesting novel enhancements and approaches to multiple-input multiple-output (MIMO) wireless systems. These systems, as we shall explain later, are perhaps among the very few options (a.k.a. *degrees of freedom*) left to increase the capacity of wireless links by magnitudes large enough to meet the expected demand in wireless traffic.

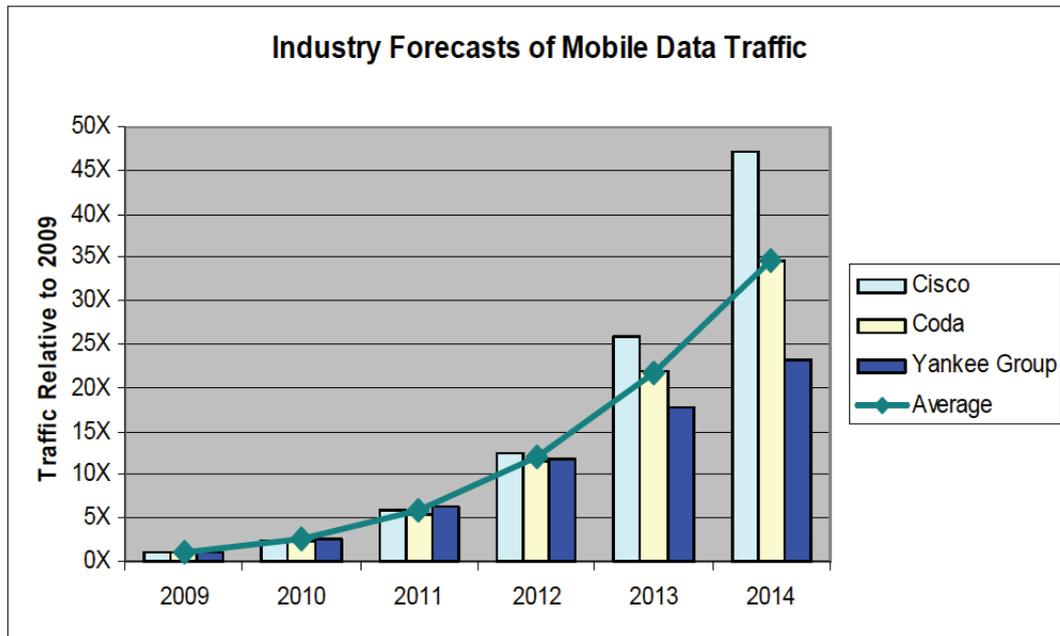


Figure 1.1: Industry mobile data forecasts.

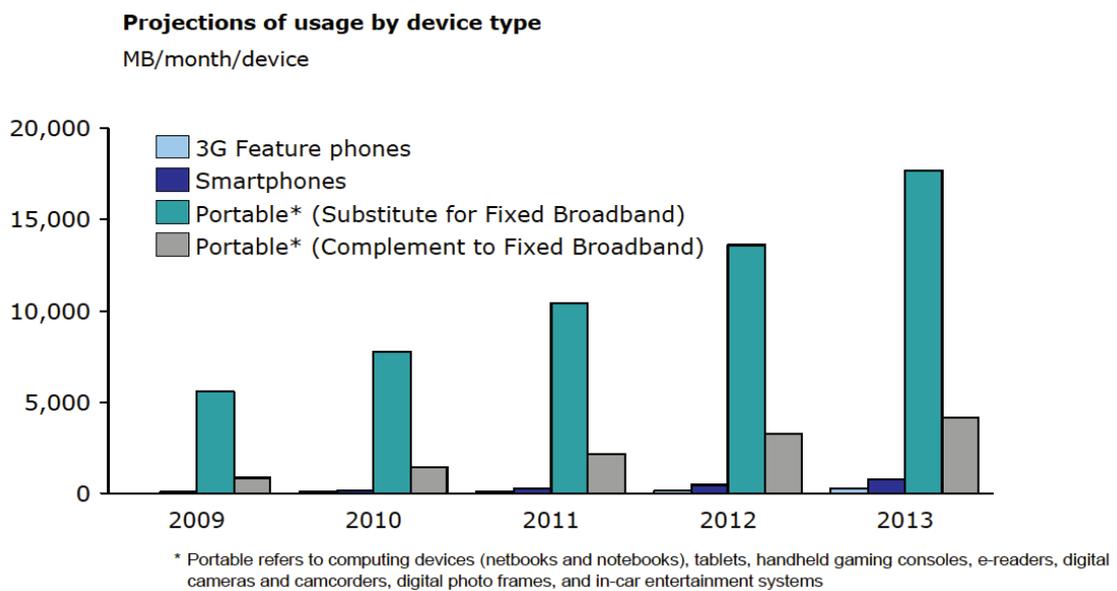


Figure 1.2: FCC estimate of mobile traffic by device type and forecasts.

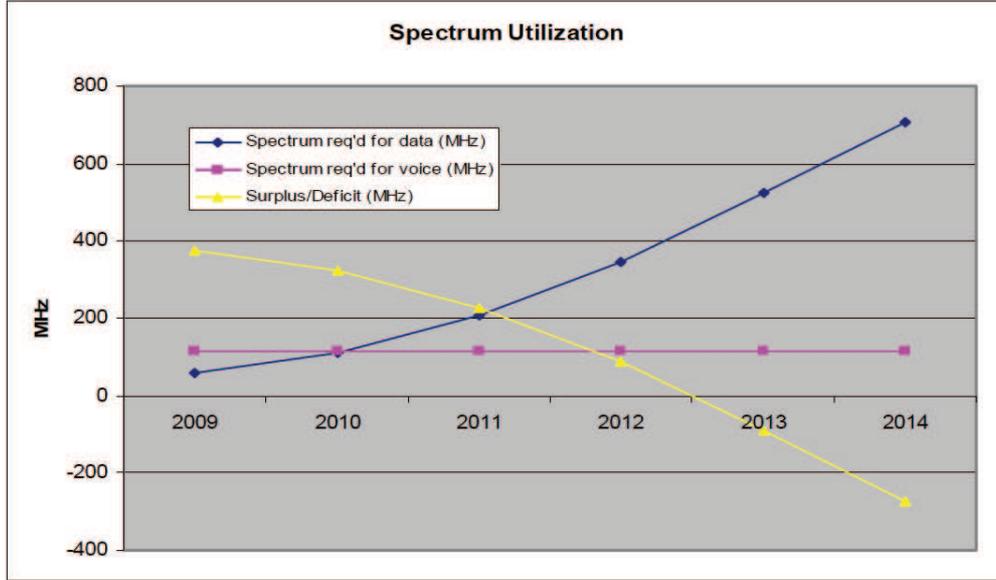


Figure 1.3: FCC projected utilization of wireless spectrum.

This chapter provides a general introduction and states the main purpose and outline of the thesis. We shall start by explaining in section 1.2 the communication problem from a capacity perspective, i.e. *why is it that we cannot transmit data at an arbitrarily large speed*, and *why is it that some communication systems offer larger data rates than others*. Then, we shall introduce in section 1.3 some key concepts relative to MIMO technology. Subsequently, we shall provide in section 1.4 the main capacity results that are available in the literature on MIMO systems. Ultimately, we shall summarize our contributions and outline this thesis in section 1.5.

1.1 Notations

Throughout this dissertation, the following notations shall be used. Vectors will be denoted in bold, and matrices in capital bold letters. \otimes denotes the outer product, T denotes the vector/matrix transpose, \dagger the Hermitian (conjugate transpose) operator, $\text{tr}(\cdot)$ the trace operator, $\mathbb{E}\{\cdot\}$ the mathematical expectation (expected value), $h(\mathbf{x})$ the Shannon entropy [2, 3] and $I(\mathbf{x}; \mathbf{y})$ the mutual information between input \mathbf{x} and output \mathbf{y} [2, 3].

When a is a real number, let $a^+ \triangleq \max(a, 0)$. When \mathbf{A} is a complex matrix, let $\|\mathbf{A}\|$ denote its Frobenius norm: $\|\mathbf{A}\| \triangleq \sqrt{\text{tr}\{\mathbf{A}\mathbf{A}^\dagger\}}$. When \mathbf{x}, \mathbf{y} are complex random vectors, let $\mathbf{Q}_{\mathbf{x}\mathbf{y}}$ denote their cross-covariance matrix:

$$\mathbf{Q}_{\mathbf{x}\mathbf{y}} \triangleq \mathbb{E}\{(\mathbf{x} - \mathbb{E}\{\mathbf{x}\})(\mathbf{y} - \mathbb{E}\{\mathbf{y}\})^\dagger\}, \quad (1.1)$$



Figure 1.4: A point-to-point communication system from an information-theoretical perspective.

which, when \mathbf{x} , \mathbf{y} are zero-mean, simplifies to:

$$\mathbf{Q}_{\mathbf{xy}} = \mathbb{E}\{\mathbf{xy}^\dagger\}. \quad (1.2)$$

1.2 The Communication Problem: A Capacity Perspective

A point-to-point communication system can be modeled as depicted in Fig. 1.4. It is made up by a transmitter, a receiver and a communication channel that lies in between.

The transmitter's goal is to convey, through the communication channel, a certain message to the receiver. It is generally¹ desired that the receiver receives the message promptly and reliably. In wireless communications, messages are carried over wireless signals (electromagnetic waves) that are carefully designed so as to match the propagation environment (i.e. the channel).

The communication speed is mainly determined at the transmitter's side. That is, the higher the cadence of bits or symbols generated at the transmitter, the higher the communication speed will be. As the electromagnetic waves travel at (almost) the speed of light, one might expect that, at least in theory, a message can be transmitted at any arbitrarily high communication speed.

If true, then *why*, -one may wonder-, *is it that point-to-point wireless systems have limited the communication rates?*

The reason is as follows. Surely, at any transmit rate, a message will arrive to the receiver. But a message that is not necessarily identical to the the transmitted one.

Indeed, some channels induce propagation effects that alter the transmitted signals. Thus, if say a message m_1 is transmitted through the channel, a message $m_2 \neq m_1$ may be received instead. The wireless channel is one such channels. A transmitted signal that propagates through a wireless channel may be altered owing to factors including, but not limited to, the following:

- **Power attenuation:** Various propagation losses may weaken the transmitted signal's power. These include path-loss attenuation, shadowing, scattering, etc. These factors are generally dependent on the propagation characteristic of the wireless channel, such as the presence

¹There may be situations where transmission reliability or speed is not of the essence, such as fault-tolerant and delay-tolerant networks, respectively.

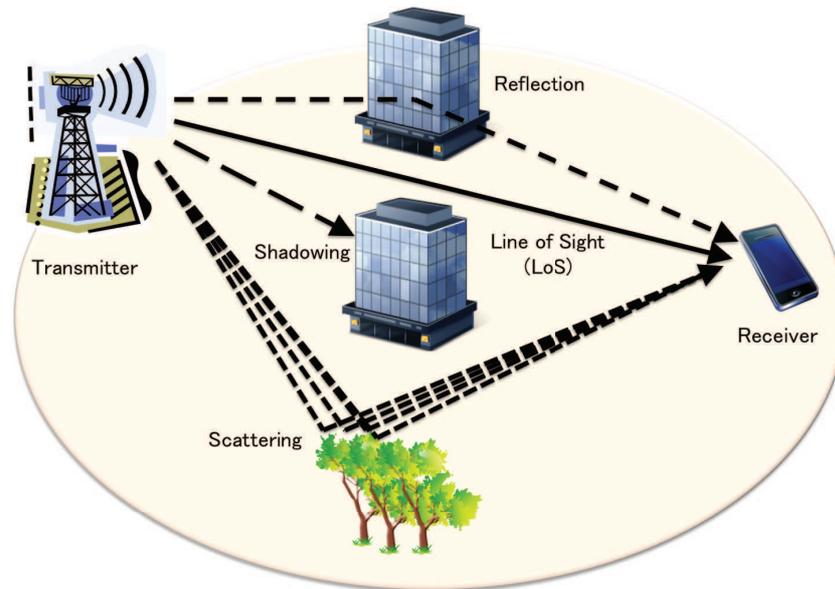


Figure 1.5: An illustrative example of a wireless propagation environment and the typical attenuations that the wireless signal can be subject to.

of obstacles, reflectors, scatterers, weather conditions (humidity), etc. See Fig. 1.5 for an illustrative example.

- **Signal incoherence:** Signal temporal/spatial incoherence may occur owing to factors such as transmit/receive desynchronization, frequency mismatch, mobility (Doppler effect), etc.
- **Thermal noise:** Additive thermal noise from electronic components adds up to the signal at the receiver's side.
- **Wireless interference:** Signals from equipments/infrastructure that transmit on the same channel interfere with the useful signal and add up at the receiver's side.

As the receiver can no longer trust the message m_2 delivered through the channel, it has to try to guess from the received message m_2 the message m_1 that was originally transmitted by the transmitter.

But can it? Claude Elwood Shannon, who pioneered the mathematical science of communication theory, demonstrated in his seminal work [2] a key correlation between the transmission rate (speed of communication) and the probability of error of the receiver (when guessing m_1 from m_2): If the transmission rate is below a certain threshold C , then this probability of error can be made as small as desired. However, beyond the threshold C , it is not possible for the receiver to reliably guess m_1 from m_2 , no matter the encoding/decoding strategy.

Thus, from a capacity perspective, the communication problem consists in determining, for a given communication channel, the upper bound C on the reliable transmission rates R , as well as the related communication strategies (a.k.a. *channel codes*) that allow the receiver to reliably guess the transmitted messages at such high transmission rates.

We shall now develop these ideas in more details, so as to gain a better understanding of the reasons behind the limits on the transmission data rate of a reliable communication.

1.2.1 The Channel Coding Theorem

Shannon introduced two key concepts, namely the entropy of a discrete random variable and the mutual information between two discrete random variables. These two notions have been crucial in determining the capacity limits of communication systems, as we shall explain momentarily.

Definition 1 (Entropy [3]) *The entropy $H(x)$ of a discrete random variable x with probability density function $p(x)$ is defined as:*

$$H(x) \triangleq -\mathbb{E} \{ \log_2 (p(x)) \}, \quad (1.3)$$

and is expressed in bits when the logarithm is to base 2.

The entropy is a measure of the average uncertainty in the random variable x . It is also the average number of bits required to describe the random variable x [3]. For instance, if x is a uniformly-distributed random variable that takes one value among 32 possibilities, then x can be described using a 5-bit sequence ($2^5 = 32$). Meanwhile, its entropy is:

$$H(x) \triangleq - \sum_{i=1}^{32} p(x=i) \log(p(x=i)) \quad (1.4)$$

$$= - \sum_{i=1}^{32} \frac{1}{32} \log \left(\frac{1}{32} \right) \quad (1.5)$$

$$= \log(32) \quad (1.6)$$

$$= 5 \text{ bits}, \quad (1.7)$$

which is in agreement with the number of bits needed to describe x .

Definition 2 (Conditional Entropy [3]) *The conditional entropy $H(x|y)$ is the entropy of a discrete random variable x conditional on the knowledge of another random variable y :*

$$H(x|y) \triangleq -\mathbb{E}\{\log_2(p(x|y))\}, \quad (1.8)$$

where $p(x|y)$ denotes the conditional probability density of x given y :

$$p(x|y) \triangleq \frac{p(x,y)}{p(y)}. \quad (1.9)$$

The knowledge of y provides additional information about x , thereby diminishing the uncertainty of x . Therefore, we always have:

$$H(x) \geq H(x|y). \quad (1.10)$$

Definition 3 (Mutual Information [3]) *The reduction in uncertainty of the random variable x due to the knowledge of the random variable y is called the mutual information between x and y :*

$$I(x; y) \triangleq H(x) - H(x|y) \geq 0. \quad (1.11)$$

It is therefore expressed in bits when the logarithm is to base 2.

Thus, the mutual information indeed provides a quantitative assessment of *how much information did we learn about x after receiving y .*

The Capacity of A Communication Channel

Shannon defined the capacity C of a discrete memoryless channel between a source x and a destination y (both modeled as discrete random variables) as the maximum mutual information between x and y , over the set of probability densities $p(x)$ [3]:

$$C = \max_{p(x)} I(x; y). \quad (1.12)$$

Denoting R the transmission rate over an arbitrary discrete memoryless channel, Shannon proved the following:

- If $R < C$, there exists a sequence of error-correcting codes such that the receiver's probability of error can be made as small as desired.
- If $R > C$, then the receiver's probability of error is bounded away from zero, regardless of the communication strategy.

Further, Shannon derived a single-letter expression for some classical communication channel models. One such model is the additive white Gaussian noise (AWGN) channel where, basically, only an AWGN adds up to the transmitted signal at the receiver's side. Such a model is widely used in many applications, including wireless communications.

For the AWGN channel, Shannon found the channel capacity to be² :

$$C_{\text{AWGN}} = B \log_2(1 + \text{SNR}) \quad (\text{in b/sec}), \quad (1.14)$$

where:

- C_{AWGN} : denotes the channel capacity of the AWGN channel.
- B : denotes the frequency bandwidth of the channel (in Hz).
- SNR: denotes the signal-to-noise ratio. Traditionally, $\text{SNR} = \frac{P}{N}$, where P , N respectively denote the powers of the transmitted signal at the reception and the power of the additive noise (in watts).

Thus if R denotes the rate at which the transmitter transmits the message m_1 , Shannon claims that:

- $R < C_{\text{AWGN}}$: there exists a communication strategy such that the receiver can reliably guess m_1 from the received message m_2 (with a probability of error that can be made as small as desired)
- $R > C_{\text{AWGN}}$: there exists no a communication strategy such that the receiver can reliably guess m_1 from the received message m_2 .

²This capacity result assumes the transmission to be carried over a frequency bandwidth of B Hz. An alternative definition of capacity is the so-called *spectral efficiency*:

$$C_{\text{AWGN}} = \log_2(1 + \text{SNR}) \quad (\text{in b/sec/Hz}), \quad (1.13)$$

which is the channel capacity per Hz of bandwidth. In this thesis, we shall make use of both definitions.

In the early years of publication of his work, Shannon's work received little recognition, in part as the channel capacity predicted by Shannon was deemed far larger than what was needed at the time. One reviewer of Shannon's work wrote [4]:

The author mentions computing machines, such as the recent ENIAC. Well, I guess one could connect such machines, but a recent IBM memo stated that a dozen or so such machines will be sufficient for all the computing that we'll ever need in the foreseeable future, so there won't be a whole lot of connecting going on with only a dozen ENIACs! IBM has decided to stay out of the electronic computing business, and this journal should probably do the same!

But the extraordinary advances in communication and information technology created a huge demand in transmission rates that called for data rates much larger than the Shannon bound itself.

1.2.2 Traditional Ways of Increasing Channel Capacity

From the capacity formula in (1.14), there are limited ways through which the channel capacity can be increased. So far, the following solutions have been suggested to increase the channel capacity:

- **Increasing the SNR:** According to (1.14), the channel capacity is increasing with SNR. The latter could be increased by using either of the following:
 - Increasing the signal's power at the reception P : this could be achieved by increasing the transmit power, shortening the distance between the transmitter and the receiver (e.g. by reducing the cell size from a few kilometers in the 70's to a few meters (so-called femto-cells) nowadays, bringing the transmitter closer to the base station, relaying, etc.).
 - Lowering the noise effect by amplifying the signals at the reception through low-noise amplifiers (LNAs).

While these solutions have improved the performance of communication systems, the increase in capacity has been insufficient to meet the huge demand in capacity. Reasons include:

- Limited resources: Battery-equipped devices such as cellular/mobile phones and wireless sensor nodes have stringent power consumption limits that prohibit them from transmitting at large transmit powers.
 - Mutual interference: If a transmitter T_1 increases its power, while it improves the reception of its signal at a receiver R_1 's side, it causes additional interference on receivers that are receiving on the same channel. A transmitter T_2 serving such receivers would have to increase, in turn, its own transmit power, thereby causing interference to R_2 , and forcing T_1 to increase its transmit power again, etc. This would cause an endless vicious circle of power increase.
 - According to (1.14), the channel capacity scales logarithmically with the transmit power. That is, for every 10 watts of power that we spend on the transmission, we roughly get $\log(10)$ worth of capacity gains. As $\log(x) < x$, increasing the transmit power could be too costly to be worthwhile.
 - Health concerns and regulatory requirements: As of today, whether wireless communications pose health risks on humans is a matter of argument. As a precautionary measure, the transmit power of signals emitting from wireless infrastructures is traditionally standardized and a cap is usually defined. For instance, in cellular communications, the Federal Communications Commission (FCC) of the U.S. Government sets the Specific Absorption Rate (SAR, the exposure standard for wireless devices) in GSM to 1.6 W/kg.
- **Increasing the signal's frequency bandwidth:** Increasing the frequency bandwidth has an advantage over increasing the transmit power that it generates a linear increase in capacity. This led to the emergence of ultra wideband (UWB) communications. Unfortunately, bandwidth is scarce. Further, in the licensed spectrum, it is very expensive.
 - **Multiplexing:** Multiplexing consists in exploiting certain channel characteristics such that the channel can unfold into multiple sub-channels, thereby increasing the channel capacity. Such characteristics are called *degrees of freedom*. They include:
 - *Frequency:* Frequency Division Multiplexing (FDM) systems are communication systems where different signals are simultaneously transmitted over different frequencies. Examples include orthogonal frequency division multiplex (OFDM) systems. A requirement is that signals be perfectly synchronized in time.

- *Time*: Time Division Multiplexing (TDM) systems are communication systems where different signals are transmitted over same frequencies but on different time slots, as in GSM. A requirement is that signals be perfectly synchronized in frequency.
- *Codes*: Code Division Multiplexing (CDM) systems are communication systems where different signals are transmitted over the same frequencies and time slots but are spectrum-spread using orthogonal chip codes, as in CDMA2000. Owing to the orthogonality requirement, there is usually a limit on the number of chip codes that can be simultaneously used in a single transmission.
- *Polarization*: Electromagnetic wavelengths may have vertical, horizontal or elliptic polarizations. Polarization diversity systems allow to multiplex two signals on the same time slot and frequency but using different polarizations. A requirement is that propagation environment preserves to a sufficient extent the polarization of each of the transmitted signals. This, for instance, is not the case of rain fields where rain drops alter the polarization of polarized signals, as in satellite communications.

1.2.3 Emergence of MIMO Technology: A Brief History

The previous solutions, though helpful in achieving further capacity gains, are unable to meet the future demand in bandwidth. As one channel between a transmitter and a receiver inherently has an insufficient channel capacity, came the intuition to use multiple channels between them.

In wireless communications, a channel is defined as the propagation environment between one transmit antenna and one receive antenna. Therefore, having multiple channels between the transmitter and the receiver amounts to deploying multiple antennas at the transmitter and the receiver. That is multiple-input multiple-output, or simply: MIMO. Other variants include: using multiple antennas at the transmitter but a single antenna at the receiver (multiple-input single-output, or MISO) and vice versa (single-input multiple-output or SIMO). Finally, the conventional single antenna case is traditionally termed single-input single-output or SISO.

Although MIMO has gained in popularity from the late 90's, the early ideas trace back to the 70's, with the works of A.R. Kaye and D.A. George (1970) as well as W. van Etten (1975, 1976). In the 80's, Bell Laboratories started to gain interest in MIMO, with Jack Winters and Jack Salz publishing seminal works on beamforming [5]. The concept of spatial multiplexing using MIMO was first introduced by A. Paulraj in 1993 [6] and in a subsequent patent in 1994, emphasizing applications to wireless broadcasting.

But only in 1998 has MIMO become a popular research topic when Bell Labs first demonstrated a laboratory prototype of spatial multiplexing called Bell Labs Layered Space-Time Architecture (BLAST) [7] where fantastic capacity gains have been demonstrated in practice using MIMO-assisted spatial multiplexing.

Today, MIMO technology is an essential physical layer component of next generation wireless standards, such as WiMAX, IEEE 802.11n and 4G systems. Novel concepts such as MIMO radars [8] and MIMO-over-satellite [9] have also recently emerged.

1.3 MIMO Technology: Key Concepts

The following is intended to introduce key concepts behind MIMO technology that will be later needed to understand the huge potential of MIMO to provide fantastic wireless capacity gains.

1.3.1 System Model

A MIMO system is depicted in Fig. 1.6. Assume the transmitter is equipped with $N \geq 1$ transmit antennas and the receiver with $M \geq 1$ receive antennas, (N and M being arbitrarily defined). The following input-output model is traditionally adopted in the MIMO literature [10]:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} h_{11} & \dots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{M1} & \dots & h_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} z_1 \\ \vdots \\ z_M \end{bmatrix}, \quad (1.15)$$

or equivalently:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad (1.16)$$

where:

- $\mathbf{x} = [x_1, \dots, x_N]^T$ denotes the vector of N inputs (signals) transmitted on the N antennas (one input per transmit antenna);
- $\mathbf{y} = [y_1, \dots, y_M]^T$ denotes the vector of M outputs (signals) received on the M antennas (one output per receive antenna);
- $\mathbf{z} = [z_1, \dots, z_M]^T$ denotes the vector of M AWGN (noise) terms that corrupt the signal received on each receive antenna: $\mathbf{z} \sim (0, \sigma^2 \mathbf{I}_M)$, \mathbf{I}_M being the identity matrix;

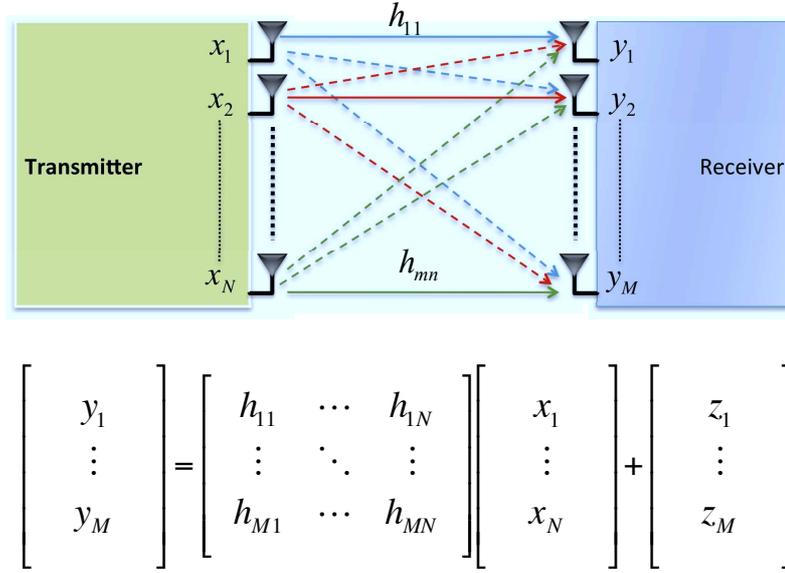


Figure 1.6: MIMO system model.

- $\mathbf{H} = (h_{ij})_{1 \leq i \leq M, 1 \leq j \leq N}$ denotes the matrix of channel gains (a.k.a. the channel matrix). Precisely, h_{ij} denotes the channel gain coefficient between the i th transmit antenna and the j th receive antenna. In a way, h_{ij} reflects the channel effect on the signal transmitted from i th transmit antenna and received by the j th receive antenna.

Traditionally, the transmitted multi-dimensional signal is subject to a transmit power constraint such as:

$$\text{tr}(\mathbf{Q}_x) \leq P, \quad (1.17)$$

where P , \mathbf{Q}_x respectively denote the maximum transmit power and the input covariance matrix. Further, it is common in wireless communications to transmit signals that have zero mean, in which case the previous equation (1.17) simplifies to:

$$\text{tr}(\mathbf{Q}_x) = \sum_{i=1}^N \mathbb{E} \{ x_i x_i^\dagger \} \leq P. \quad (1.18)$$

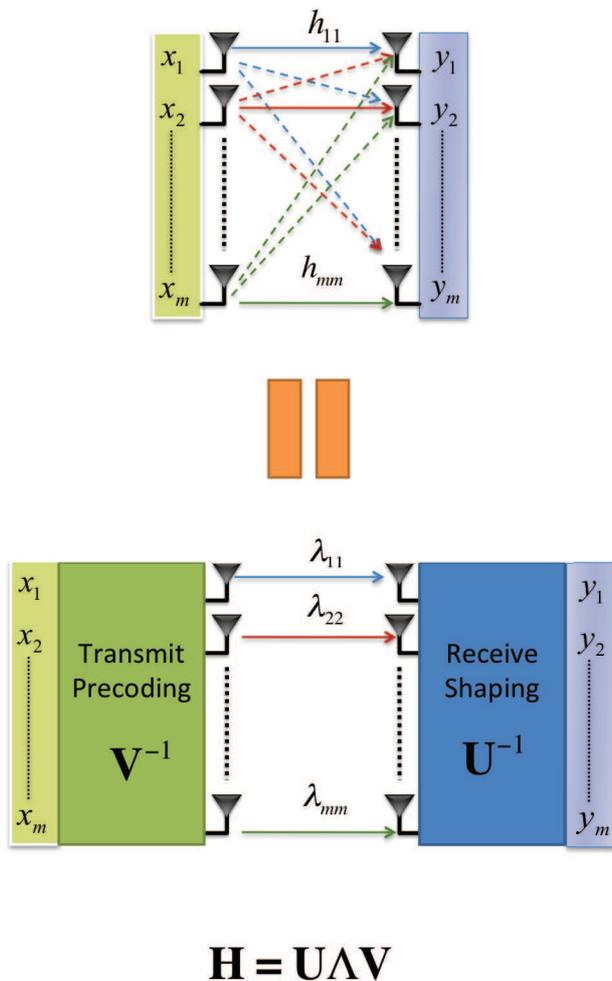


Figure 1.7: Parallel decomposition of the MIMO channel

Of significant importance is how the input \mathbf{x} is designed at the transmitter’s side and how the received signal \mathbf{y} processed at the receiver’s side for reliable communication. We shall explain this in the subsequent paragraph.

Throughout this thesis, unless otherwise mentioned, we shall assume the same number of transmit and receive antennas, for simplicity.

Parallel Decomposition of The MIMO Channel

Let us consider the system model described by (1.25). Without loss of generality, let us focus on the signal y_1 at the 1st receive antenna, given by:

$$y_1 = h_{11}x_1 + \sum_{i=2}^N h_{1i}x_i + z_1. \tag{1.19}$$

Suppose the receiver wants to detect (guess) the most-likely signal x_1 that was transmitted, given the received signal y_1 . Then, one approach would be to treat the whole term $\sum_{i=2}^N h_{1i}x_i + z_1$ as noise. However, this term could have a very large power (norm) compared with the term $h_{11}x_1$. Thus appears the need to cancel the term $\sum_{i=2}^N h_{1i}x_i$ prior to detecting x_1 . One approach towards this end is to use the *singular-value decomposition* (SVD). It is known that any matrix \mathbf{H} can be SVD-decomposed as follows:

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}, \quad (1.20)$$

where:

- \mathbf{U} , \mathbf{V} are unitary matrices (i.e. unit-norm matrices whose columns are mutually-orthogonal)
- $\mathbf{\Lambda}$ is a diagonal matrix (i.e. a matrix where all non-diagonal entries are zero).

Therefore, if the respective inverses \mathbf{U}^{-1} and \mathbf{V}^{-1} of \mathbf{U} and \mathbf{V} are resp. applied to the received and transmitted signals \mathbf{x} and \mathbf{y} as depicted in Fig. 1.7, we get:

$$\mathbf{y} = \mathbf{U}^{-1}(\mathbf{H}\mathbf{V}^{-1}\mathbf{x} + \mathbf{z}) \quad (1.21)$$

$$= \mathbf{U}^{-1}(\mathbf{U}\mathbf{\Lambda}\mathbf{V}\mathbf{V}^{-1}\mathbf{x} + \mathbf{z}) \quad (1.22)$$

$$= \lambda\mathbf{x} + \tilde{\mathbf{z}}, \quad (1.23)$$

where $\tilde{\mathbf{z}}$ is a modified noise but nonetheless white, Gaussian and having the same power as the original noise \mathbf{z} (owing to the fact that the matrices \mathbf{U} , \mathbf{V} are unitary). It follows that the received signal at e.g. receive antenna 1 becomes:

$$y_1 = \lambda_{11}x_1 + z_1, \quad (1.24)$$

thereby making the detection of x_1 at the receiver's side possible.

The procedure of computing the matrix \mathbf{V} and applying it at the transmitter's side is called *transmit precoding*.

Likewise, the procedure of computing the matrix \mathbf{U} and applying it at the receiver's side is called *receive shaping*.

Finally, the eigenvalues $\lambda_{11}, \dots, \lambda_{NN}$ are called the *eigenmodes* of the channel³.

³The number of eigenvalues of the channel is $\min N, M$. By denoting λ_{NN} the last eigenvalue, we are implicitly assuming (without loss of generality) that $N \leq M$

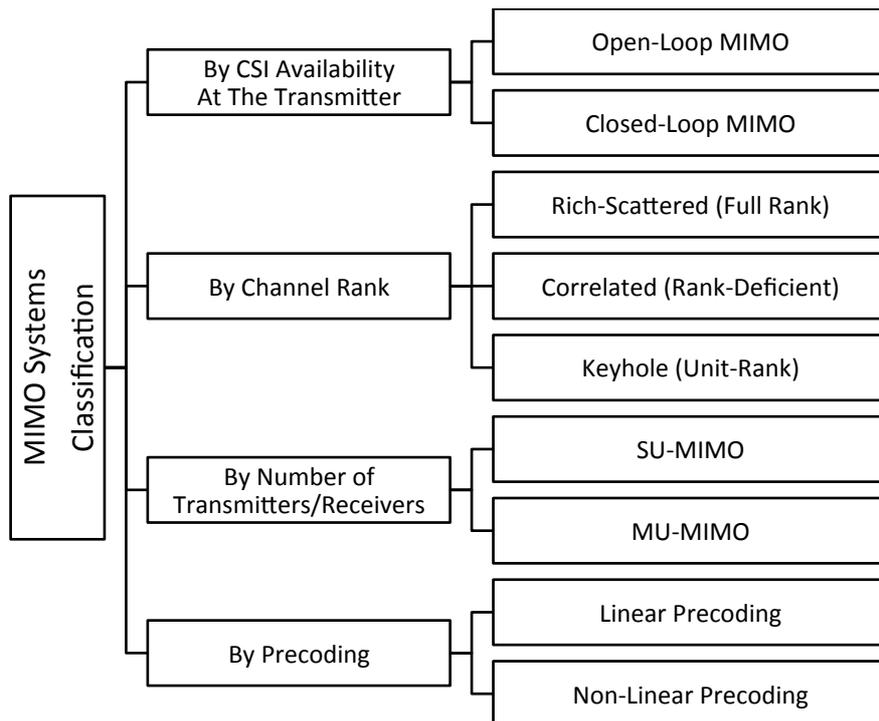


Figure 1.8: A general overview of MIMO systems.

Depending on the number $1 \leq n \leq N$ of non-zero eigenmodes, we can distinguish:

- **A rich-scattered channel**, when $n = N$, or equivalently when the channel matrix \mathbf{H} is full-rank.
- **A correlated channel**, when $n < N$, or equivalently when the channel matrix \mathbf{H} is rank-deficient.
- **A keyhole channel** (which is a particular kind of correlated channels), when $n = 1$ and the correlation is purely owing to the propagation environment (a.k.a. *spatial correlation*).

1.3.2 Open-Loop MIMO Vs Closed-Loop MIMO

As we have just seen, performing SVD decomposition to determine the transmit precoder allows for a reliable detection of MIMO signals at the receiver's side. The availability of the CSI \mathbf{H} at the transmitter and the receiver is a requisite to achieve the parallel decomposition of the channel and allow for a reliable transmission. While usually CSI is somewhat easily made available at the receiver's side, there may be situations where CSI is unavailable to the transmitter.

Based on CSI availability to the transmitter, we may distinguish two MIMO approaches:

- **Open-loop transmissions**, where CSI is not available to the transmitter. In such a case, the transmitter blindly transmits its signals without adapting the transmission to the channel realizations.
- **Closed-loop transmissions**, where CSI is available to the transmitter. In such a case, the transmitter can adapt its transmission to the channel, thereby increasing its transmission rate (e.g. SVD decomposition).

Open-Loop MIMO

In open-loop MIMO systems, the transmitter does not know the instantaneous realization of the channel matrix \mathbf{H} . This is the case for instance when the channel fades are too fast to track. As such, it cannot use adaptive transmissions to adapt its transmission to the channel. In such a case, rather than adapting to the instantaneous channel realizations, it is possible to adapt the channel statistics instead. Indeed, channel statistics usually change less rapidly than the channel realizations themselves. Thus, even though the channel fades may be rapidly-varying (e.g. high mobility scenario), the channel statistics can still be tracked. Subsequently, it is possible to increase the system *average capacity*, commonly known as the *ergodic capacity*.

Closed-Loop MIMO

In closed-loop MIMO systems, the instantaneous realization of the channel matrix \mathbf{H} , or some related information, is available to the transmitter. In practice, it may be difficult for the transmitter to determine such channel matrix by itself. Rather, it is common that the transmitter sends an *a priori* known message \mathbf{X}_p , called a *pilot sequence*, to the receiver. The latter receives:

$$\mathbf{Y} = \mathbf{H}\mathbf{X}_p + \mathbf{Z}, \quad (1.25)$$

from which it attempts to estimate the channel \mathbf{H} under the uncertainty caused by the additive noise \mathbf{Z} . The following estimators are traditionally used to estimate the channel [11]:

- The Least Squares Estimator (LS)

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{X}_p^\dagger \left(\mathbf{X}_p\mathbf{X}_p^\dagger \right)^{-1} \quad (1.26)$$

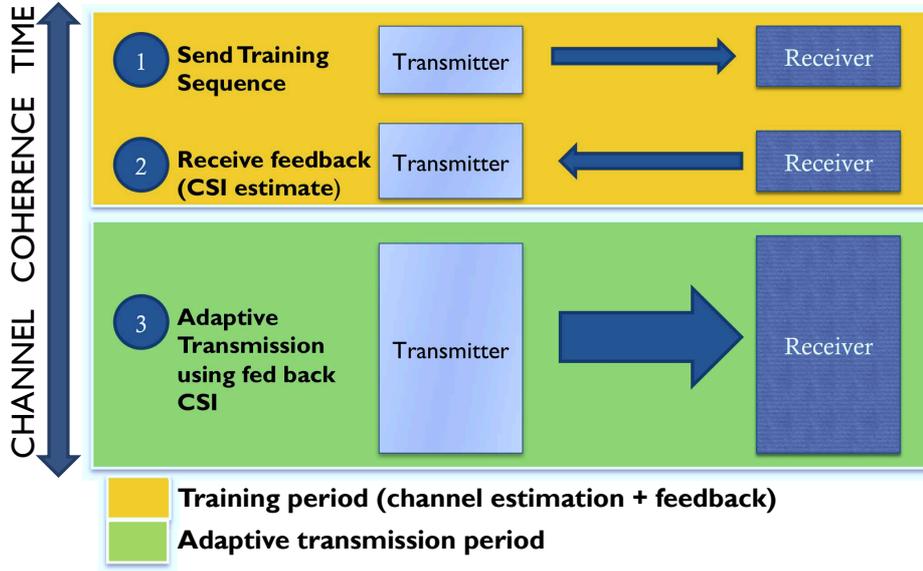


Figure 1.9: Phases that make up closed-loop MIMO.

- The Minimum Mean Squared Error Estimator (MMSE)

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{X}_p^\dagger \left(\left(\mathbf{X}_p\mathbf{X}_p^\dagger \right)^{-1} + \sigma^2\mathbf{I}_M \right)^{-1}. \quad (1.27)$$

Subsequently, the receiver feeds back the channel state information (either the channel matrix estimate $\hat{\mathbf{H}}$ or function of it) to the transmitter. The latter uses this channel information to optimally transmit its data, e.g. to perform the SVD decomposition and determine the optimal precoder. Fig. 1.9 illustrates the afore-described procedure.

Alas, the channel is not always constant and may change owing to a variety of reasons. It has, however, a characteristic period of time during which it remains, on the average, almost constant. Such period is called the *channel's coherence time*. Thus, a key requisite for closed-loop MIMO to be efficient is that the channel estimation phase uses as little time as possible from the total coherence time, so that much of the latter be used for transmitting useful data before the channel changes into a new realization.

Throughout this thesis, unless otherwise mentioned, we shall assume that MIMO is operating in the closed-loop mode.

1.3.3 Full / Limited / Quantized CSI

Depending on the content of the signaling information that is transmitted from the receiver back to the transmitter about the channel condition, three types of CSI can be distinguished, namely: *Full CSI*, *limited* (a.k.a. *imperfect*, *partial*) *CSI* and *quantized CSI*.

The Full CSI Case

Originally, the term CSI referred to the channel matrix \mathbf{H} . However, in closed-loop MIMO systems, feeding back \mathbf{H} as the CSI appeared to be a time-and-resource costly signaling procedure. For instance, in a 4×4 MIMO system, the channel matrix \mathbf{H} has 16 complex-valued channel gains. Thus, if the receiver is to feed back \mathbf{H} to the transmitter, then 32 floating numbers (half referring to the real parts and the other half to the imaginary parts) have to be transmitted.

This led to the suggestion of alternative forms of CSI in the literature, as we shall see now.

The Limited CSI Case

The limited CSI category itself can be divided into 3 sub-categories, depending on how the channel is modeled [10]:

- **The zero mean spatially white (ZMSW) channel model:** under such model, the channel is assumed to have zero mean and white covariance (i.e. i.i.d. channel entries):

$$\mathbb{E} \{ \mathbf{H} \} = \mathbf{0} \quad (1.28)$$

$$\mathbf{H} = \mathbf{H}^w, \quad (1.29)$$

where \mathbf{H}^w denotes a matrix with i.i.d. entries ($\mathbf{H}^w \sim \mathcal{CN}(0, \mathbf{I}_{NM})$). In such a case, feedback from the receiver is not needed at all, since the channel distribution is perfectly known ($\mathcal{CN}(0, \mathbf{I}_{NM})$).

- **The channel mean information (CMI) model:** the channel is assumed to have a non-zero mean and a white covariance:

$$\mathbb{E} \{ \mathbf{H} \} = \bar{\mathbf{H}} \quad (1.30)$$

$$\mathbf{H} = \bar{\mathbf{H}} + \sqrt{\alpha} \mathbf{H}^w, \quad (1.31)$$

where \mathbf{H}^w denotes a matrix with i.i.d. entries ($\mathbf{H}^w \sim \mathcal{CN}(0, \mathbf{I}_{NM})$), α denotes a parameter and $\bar{\mathbf{H}}$ denotes the channel mean. In such a case, the channel mean has to be computed by the receiver and fed back to the transmitter.

- **The channel covariance information (CCI) model:** the channel is assumed to have a zero mean but a non-white covariance:

$$\mathbb{E}\{\mathbf{H}\} = 0 \quad (1.32)$$

$$\mathbf{H} = (\mathbf{R}^r)^{1/2} \mathbf{H}^w (\mathbf{R}^t)^{1/2}, \quad (1.33)$$

where \mathbf{H}^w denotes a matrix with i.i.d. entries ($\mathbf{H}^w \sim \mathcal{CN}(0, \mathbf{I}_{NM})$) and $\mathbf{R}^t, \mathbf{R}^r$ respectively denote the correlation matrices at the transmit and receive array antennas. In such a case, the channel covariance has to be computed by the receiver and fed back to the transmitter.

In [12], [10], it has been shown that adaptive transmission (beamforming) may not be optimal (i.e. may not be the capacity-achieving strategy) when using limited CSI. Precisely, a condition on the largest two eigenvalues of the channel has been required in order to achieve the channel capacity with only the channel mean or the channel covariance available at the transmitter. As there are no ways to determine (in advance) whether a channel may or may not satisfy such conditions, limited feedback can not guarantee a capacity-achieving MIMO data transmission.

The Quantized CSI Case

When full CSI is available to the transmitter, the latter can perform SVD decomposition and obtain the optimal beamformer (precoder). This, as we said earlier, is the capacity-achieving power allocation strategy. Albeit optimal, this solution comes at a high signaling cost, as the entire channel matrix \mathbf{H} has to be fed back to the transmitter every time the channel changes.

Quantized feedback is an alternative approach where the goal, rather than maximizing the capacity, is to minimize the signaling cost while achieving a capacity as large as possible.

For this sake, the transmitter shall no longer use the SVD decomposition to determine the optimal precoder.

Instead, it will use a precoder ω^* from within a predefined set of precoders $\Omega \triangleq \{\omega_1, \dots, \omega_{|\Omega|}\}$.

A precoder $\omega \in \Omega$ is called a *codeword*.

The set Ω of codewords shall be called a *codebook*.

In that case, the full CSI \mathbf{H} need not be fed back to the transmitter. Rather, the receiver can select from within Ω the optimal precoder ω^* as follows [13]:

$$\omega^* \triangleq \arg \max_{\omega \in \Omega} \|\mathbf{H}\omega\|^2 \quad (1.34)$$

$$\triangleq \arg \max_{\omega \in \Omega} \left| \omega^\dagger \mathbf{H}^\dagger \mathbf{H} \omega \right| \quad (1.35)$$

Then, the receiver can feed back the vector ω^* to the transmitter, thereby reducing the signaling burden from one matrix down to one vector. Even better, if the set Ω is *a priori* known to the transmitter, then it suffices that the receiver only sends the index of the optimal codeword within Ω , rather the sending codeword itself. Thus, if the codebook Ω has $|\Omega| = n$ codewords, then only $m = \log_2(n)$ bits need be transmitted in the feedback phase.

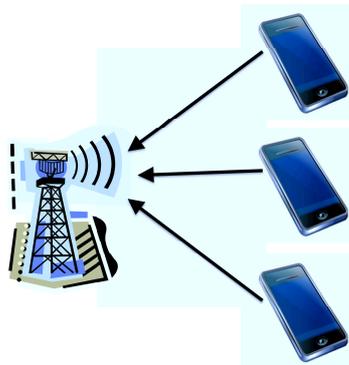
3GPP LTE and beyond wireless standards uses quantized feedback, with codebooks having up to 16 codewords [14], thereby requiring $\log_2(16) = 4$ bits for CSI feedback, at the most.

1.3.4 Single-User MIMO Vs Multi-User MIMO

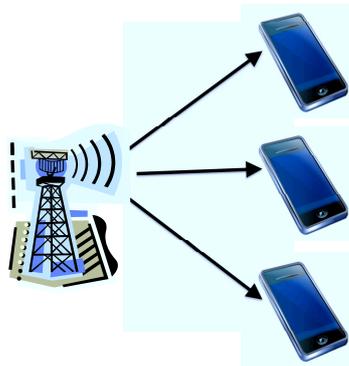
Single-User MIMO (SU-MIMO) refers to scenarios where only one transmitter is sending data to only one receiver on a given channel.

Contrarily, multi-user MIMO (MU-MIMO) refers to settings where MIMO communications involve more than one transmitter and/or more than one receiver, see Fig. 1.10:

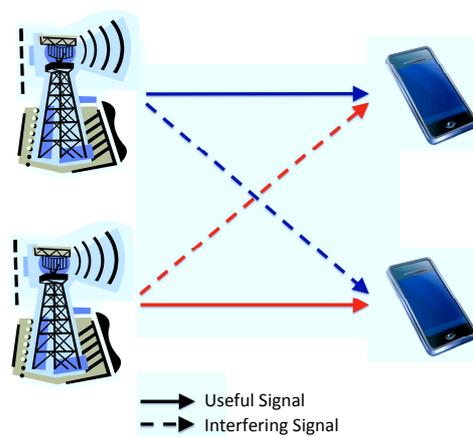
- When multiple transmitters are simultaneously transmitting over the same channel to the same (single) receiver, the channel is called a *multiple-access channel* (MAC). In cellular networks, this may correspond to the uplink channel where multiple cell phones may simultaneously transmit data to the same base station.
- When only one transmitter is transmitting over the same channel to the multiple receivers, the channel is called a *broadcast channel* (BC). In cellular networks, this may correspond to the downlink channel where multiple cell phones may simultaneously receive data (over the same channel) from a single base station.
- When multiple transmitters are transmitting to their respective receiver(s) over the same channel, the channel is called an *interference channel* (IC).



(a) A multiple-access channel (MAC, many-to-one)



(b) A broadcast channel (BC, one-to-many)



(c) An interference channel (IC)

Figure 1.10: Various MU-MIMO scenarios.

1.4 On The Channel Capacity of MIMO Systems

The potential benefits from using multiple antennas can be assessed using two performance metrics:

1. **Capacity gains:** How much capacity increase do we get when we use $N > 1$ antennas at the transmit side and/or $M > 1$ antennas at the receive side w.r.t. the single antenna case (SISO) ?
2. **Capacity scaling gains (also known as the MIMO *degrees of freedom*):** How does capacity *scale* with the number of transmit/receive antennas ? For instance, if we use a 4-by-4 MIMO system ($N = 4$ transmit antennas and $M = 4$ receive antennas), do we get 4 times the channel capacity of a SISO (1-by-1) system?

1.4.1 The Single-User Case

Main Capacity Result

The channel capacity, as defined by Shannon [2], is given by the mutual information $I(\mathbf{x}, \mathbf{y})$ between the channel input \mathbf{x} and the output \mathbf{y} :

$$C \triangleq \max_{\mathbf{Q}_x} I(\mathbf{x}, \mathbf{y}) \quad (1.36)$$

$$= \max_{\mathbf{Q}_x} (H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})). \quad (1.37)$$

It can be shown [15] that the capacity formula (1.37) can be re-written as:

$$C = \max_{\mathbf{Q}_x} \log_2 \det \left(\mathbf{I}_M + \frac{1}{\sigma^2} \mathbf{H} \mathbf{Q}_x \mathbf{H}^\dagger \right). \quad (1.38)$$

When Full CSI Is Available At The Transmitter

In order for the afore-mentioned channel parallel decomposition to be feasible, the channel \mathbf{H} has to be known at the transmitter and the receiver, so that each can compute its respective precoding/shaping matrix. In such a case, the capacity formula in (1.38) is shown to be [10]:

$$C = \sum_{i=1}^{\text{rank}(\mathbf{H})} \log_2 \left(1 + \lambda_i \frac{p_i}{\sigma^2} \right), \quad (1.39)$$

where p_i denotes the transmit power allotted to the i th transmit signal x_i . This capacity result leads to an important conclusion: optimal transmit powers $(p_i)_{1 \leq i \leq N}$ are most-likely different from each other. Indeed, the channel gains λ_i are, most-likely, unequal. Therefore, applying a given transmit power to one antenna/eigenmode would yield a different capacity from applying this very same power on a different antenna/eigenmode.

For instance, Let us assume $\sigma^2 = 1$, for simplicity. Then if $\lambda_1 = 0.5$ and $\lambda_2 = 1$, then using $p_1 = 5$ watt to transmit on the first eigenmode would contribute to the total channel capacity C by $C_1 = \log_2(1 + 0.5 \times 5) \approx 1.8$ bit/sec/Hz, whereas using the same transmit power for the second eigenmode would contribute to the total channel capacity C by $C_2 = \log_2(1 + 1 \times 5) \approx 2.6$ bit/sec/Hz, i.e. 45 % more capacity.

Channel eigenvalues can differ by a significant order of magnitude. So, it pays more to use most of transmit power on the good eigenmodes. The optimal power allocation scheme under perfect CSI is called the *waterfilling algorithm* [10].

When CSI Is Not Available At The Transmitter

When CSI is not available at the transmitter, the transmitter can no longer apply the waterfilling algorithm as it cannot perform the SVD decomposition of the channel. In such a case, it has been shown that uniform power allocation (applying the same transmit power to all antennas: $\mathbf{Q}_x \triangleq \frac{P}{\sigma^2} \mathbf{I}_M$) can be optimal when the number of antennas is sufficiently large [15].

Further, as the channel is unknown to the transmitter, the notion of instantaneous capacity becomes irrelevant. Instead, an ergodic capacity is computed, which is the average of the instantaneous capacity over a sufficiently large number of channel realizations (i.e. a period of time sufficiently larger than the channel's coherence time):

$$C_{\text{erg}} = \mathbb{E}_{\mathbf{H}} \left\{ \log_2 \det \left(\mathbf{I}_M + \frac{P}{\sigma^2} \mathbf{H}\mathbf{H}^\dagger \right) \right\}. \quad (1.40)$$

On MIMO Capacity Scaling: The Keyhole Problem

The previous capacity formulae predict not only the capacity gains (in bit/sec/Hz) but also the capacity scaling gains with the number of antennas. Assuming channel state information (CSI) availability, it can be inferred from (1.39) that the capacity of a MIMO channel between an N -antenna transmitter and an M -antenna receiver is n times that of a single-input-single-output (SISO) channel, where n is the rank of the MIMO channel gain matrix, less than or equal to $\min\{N, M\}$.

In particular, if the channel matrix has full rank (a so-called *rich-scattered* environment), channel capacity scales with $\min\{N, M\}$, i.e. $n = \min\{N, M\}$.

Clearly, the higher the channel rank, the more beneficial MIMO will be (from a capacity perspective) compared with a SISO system. Thus, from this perspective, it is interesting to determine conditions under which channel matrix has full rank, and to ensure that these conditions are met so that channel capacity is maximized.

While it has been long believed that decorrelating transmit antennas (e.g. by sufficiently spacing them) amply ensures a full-rank channel matrix [16, 17], recent works [16–18] demonstrated that in a so-called *keyhole* scenario, channel matrix has unit rank, *even when its entries have zero correlations between each other*. Subsequently, the benefits of rich scattering are suppressed, and channel capacity scales as that of a SISO channel (i.e. $n = 1$).

Such a frustrating result was first theoretically predicted in [16, 17], and later verified through an experimental testbed in [18]. MIMO keyholes occur when radio waves come across metal obstacles with small holes only through which they can propagate (*spatial keyholes*, see Fig. 3.1). They are also encountered in urban environments with so-called *street canyons* (narrow streets bordered by tall buildings), and in some indoor environments such as corridors, hallways and subway tunnels, settings which may act as single-moded waveguides at large distance from the source, thereby allowing only a single electromagnetic mode to pass through (*modal keyholes*). Finally, outdoor keyholes may also occur owing to a diffraction at rooftop edges (*diffraction-induced keyholes*).

1.4.2 The Multi-User Case

Unlike the single-user case, the channel capacity is *not defined by a single capacity value*, but rather by a set of combinations of capacity values called *the capacity region*.

The MAC Capacity Region

In the MAC channel, $K > 1$ transmitters simultaneously transmit data to a single receiver.

- *System model*: $\forall k \in \{1, \dots, K\}$, let:
 - \mathbf{x}_k denote the signal vector transmitted by the k th transmitter to the receiver
 - \mathbf{H}_k denote the MIMO channel gain matrix relative to the channel between the k th transmitter and the receiver

Then, the received signal \mathbf{y} at the receiver's side could be modeled as follows:

$$\mathbf{y} = \sum_{k=1}^K \mathbf{H}_k \mathbf{x}_k + \mathbf{z}, \quad (1.41)$$

where \mathbf{z} denotes the additive noise at the receiver's side.

- *Optimal signal detection strategy:* The optimal detection strategy is called *successive decoding* [10] and goes as follows:
 - Signals are detected in a decreasing order of signal strength. Without loss of generality, assume the signals $\mathbf{x}_1, \dots, \mathbf{x}_K$ are indexed such that \mathbf{x}_K has the strongest signal.
 - First, \mathbf{x}_K is detected from the received signal \mathbf{y} by treating $\mathbf{x}_1, \dots, \mathbf{x}_{K-1}$ as additive noise.
 - Then, the estimate $\hat{\mathbf{x}}_K$ is subtracted from the signal \mathbf{y} to yield the filtered signal $\mathbf{y}_{1, \dots, k-1}$.
 - Then \mathbf{x}_{K-1} is detected from the signal $\mathbf{y}_{1, \dots, k-1}$ by treating $\mathbf{x}_1, \dots, \mathbf{x}_{K-2}$ as additive noise.
 - Then, the estimate $\hat{\mathbf{x}}_{K-1}$ is subtracted from the signal $\mathbf{y}_{1, \dots, k-1}$ to yield the filtered signal $\mathbf{y}_{1, \dots, k-2}$.
 - The procedure is reiterated until all signals are detected.

For clarity, an illustrative example of successive decoding is provided in Fig. 1.11 with 4 inputs $\mathbf{x}_1, \dots, \mathbf{x}_4$ and one output \mathbf{y}_1 . The receiver's goal is to ultimately detect \mathbf{x}_1 . For that sake, it has to successively detect the other inputs, then cancel them from the received signal \mathbf{y}_1 , then finally detect the useful signal \mathbf{x}_1 .

- *Achievable capacity region by successive decoding:* Let R_k denote the rate of the transmission of each transmitter. In order for the receiver to reliably understand all the received messages, the transmission rates R_k must satisfy 2 conditions:
 - Each transmission rate R_k should not exceed the capacity of the sub-channel between transmitter k and the receiver:

$$\forall k \in \{1, \dots, K\}, R_k \leq I(\mathbf{x}_k; \mathbf{y}). \quad (1.42)$$

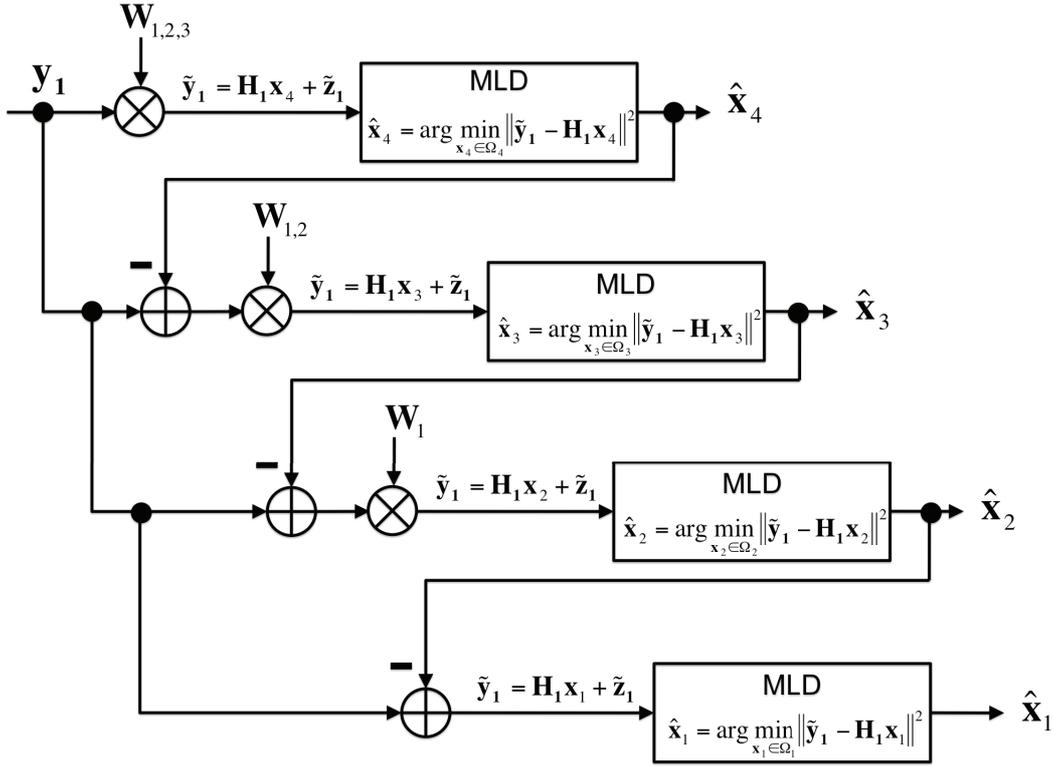


Figure 1.11: An example of successive decoding with $K = 4$ inputs. The receiver's goal is to ultimately guess \mathbf{x}_1 . $\mathbf{W}_{1,2,3}$: filter (e.g. MMSE) that cancels $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$.

- The sum of all transmission rates should not exceed the capacity of the channel between $\mathbf{x}_1, \dots, \mathbf{x}_K$ on one side and \mathbf{y} on the other:

$$\sum_{k=1}^K R_k \leq I(\mathbf{x}_1, \dots, \mathbf{x}_K; \mathbf{y}). \quad (1.43)$$

Denoting by $\mathbf{Q}_1, \dots, \mathbf{Q}_K$ the covariance matrices of the inputs $\mathbf{x}_1, \dots, \mathbf{x}_K$, the MAC capacity region Γ_{MAC} is shown to be [10]:

$$\Gamma_{\text{MAC}} = \left\{ (R_1, \dots, R_K) : \sum_{k=1}^K R_k \leq \log \det \left(\mathbf{I}_M + \sum_{k=1}^K \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger \right) \right\}. \quad (1.44)$$

The BC Capacity Region

In the BC channel, a single transmitter transmits to K receivers, with $K > 1$.

- *System model*: $\forall k \in \{1, \dots, K\}$, let:

- \mathbf{x}_k denote the signal vector transmitted by the transmitter to the k th receiver

- \mathbf{H}_k denote the MIMO channel gain matrix relative to the channel between the k th transmitter and the receiver

Then, the received signal \mathbf{y}_k at the k th receiver's side could be modeled as follows:

$$\mathbf{y}_k = \mathbf{H}_k \sum_{j=1}^K \mathbf{x}_j + \mathbf{z}_k, \quad (1.45)$$

where \mathbf{z}_k denotes the additive noise at the k th receiver's side.

- *Optimal precoding strategy*: the optimal precoding strategy for the broadcast channel is *dirty-paper coding* [10]. This transmission strategy relies on Costa's original scheme [19] for interference channels where the interference is non-causally known by the transmitter. For instance, in a SISO system, suppose that the received signal is :

$$y = x + s + z, \quad (1.46)$$

where x , s , z resp. denote the useful signal, the additive interference and the additive noise. Thus, if the transmitter non-causally knows the interference, it can subtract in advance from the useful signal and send instead:

$$\tilde{x} = x - s, \quad (1.47)$$

in which case the received signal would be:

$$y = \tilde{x} + s + z \quad (1.48)$$

$$= (x - s) + s + z \quad (1.49)$$

$$= x + z, \quad (1.50)$$

thereby achieving the capacity of an interference-free channel. The application to the MIMO case is a bit more involved, though:

- First, perform an LQ decomposition of the channel matrix \mathbf{H} . That is, find matrices \mathbf{L} , \mathbf{Q} such that:
 - * $\mathbf{H} = \mathbf{LQ}$
 - * \mathbf{L} is lower triangular

* \mathbf{Q} is quadratic.

– Define $\mathbf{u} \triangleq [u_1, \dots, u_N]$ such that:

$$u_1 = x_1 \quad (1.51)$$

$$u_2 = x_2 - \frac{l_{21}}{l_{22}}x_1 \quad (1.52)$$

$$u_3 = x_3 - \frac{l_{32}}{l_{33}}x_2 - \frac{l_{31}}{l_{33}}x_1 \quad (1.53)$$

$$\vdots \quad (1.54)$$

$$u_N = x_N - \frac{l_{N(N-1)}}{l_{NN}}x_{N-1} - \dots - \frac{l_{N2}}{l_{NN}}x_2 - \frac{l_{N1}}{l_{NN}}x_1 \quad (1.55)$$

– Finally, send:

$$\mathbf{v} = \mathbf{Q}^{-1}\mathbf{u} \quad (1.56)$$

– Subsequently, all computations done, the received signal \mathbf{y} is such that:

$$y_1 = l_{11}x_1 + z_1 \quad (1.57)$$

$$y_2 = l_{22}x_2 + z_2 \quad (1.58)$$

$$\vdots \quad (1.59)$$

$$y_N = l_{NN}x_N + z_N \quad (1.60)$$

- *On the DPC achievable capacity region:* Let R_k denote the rate of the transmission intended for receiver k , $1 \leq k \leq K$. Then, it can be shown that :

$$R_k = \log \frac{\det \left(\mathbf{I}_M + \mathbf{H}_k \left(\sum_{i=k}^K \mathbf{Q}_{\mathbf{x}_i} \right) \mathbf{H}_k^\dagger \right)}{\det \left(\mathbf{I}_M + \mathbf{H}_k \left(\sum_{i=k+1}^K \mathbf{Q}_{\mathbf{x}_i} \right) \mathbf{H}_k^\dagger \right)} \quad (1.61)$$

where $\mathbf{Q}_{\mathbf{x}_i}$ denotes the covariance matrix relative to the signal \mathbf{x}_i intended for receiver i . In the previous example, we have applied the DPC algorithm by starting with \mathbf{x}_1 . We could have started with any other input, though. Clearly, for every permutation of user input indices, we could get a different capacity value. The union of all rates R_k over the all permutations is the DPC capacity region.

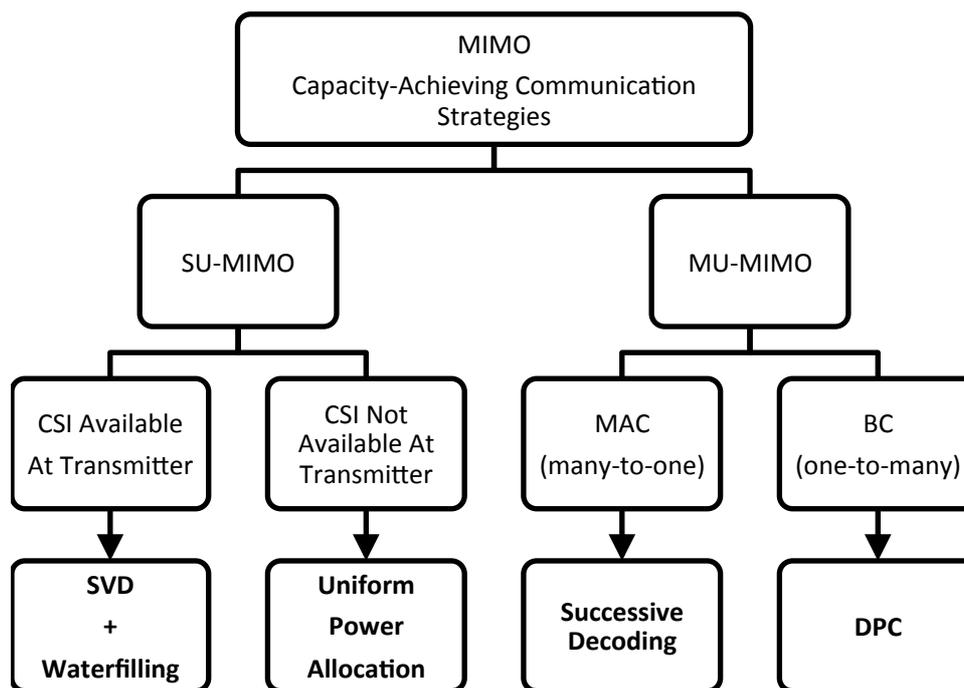


Figure 1.12: An overview of the capacity-achieving power allocation strategies in MIMO systems.

The Scheduling Issue

The scheduling issue arises when the number of transmit antennas is smaller than the number of users.

Suppose, for instance that a cellular site has, on the average, 20 active users that are served by a 4-transmit-antenna base station (BS). Evidently, at a given time, no more than 4 users can be served.

The scheduling problem consists in determining the best selection strategy that maximizes the system's total capacity, possibly subject to given constraints (ensuring fairness among users, minimizing the average queuing delay, etc.).

For instance, back to the afore-mentioned example, the BS, -out of concern for fairness-, may schedule the users' transmissions in a Round-Robin fashion, periodically serving each group of 4 users every 5 time slots.

Nonetheless, although fair, such is not an optimal solution from a capacity perspective. Indeed, if we are to maximize capacity, we had better opportunistically pick the users who have the 4 best channel conditions while having minimum interference between each other (i.e. 4 largest-SNR mutually-orthogonal users) [20].

But then again, some users may have good channel conditions all the time (e.g. owing to proximity to the BS), while others may go into deep fades and have bad channel conditions most of the time. The capacity-maximizing scheduling approach is less likely to schedule users with bad channels, and as such it is far from perfect, too.

The computational cost of the scheduling procedure is also an important design factor as well for scheduling policies. As the set of candidates becomes larger, the computational cost of the search for optimal candidate set becomes prohibitively expensive. With just a population of $K = 20$ users, there are $\binom{20}{4} \approx 5000$ possible scheduling combinations. With a population of $K = 100$ users, this figure jumps to roughly 4 million combinations.

As always in telecommunications, the signaling cost is important in deciding the worthiness of a communication scheme. In order to determine the optimal to-be-scheduled 4 users, all 20 users have to send their CSI to the transmitter. But why should 20 receivers feed back their CSI when only 4 will be scheduled for transmission ?

Finally, there is the cross-layer design issue. Physical-layer scheduling only schedules users based on their channel conditions. But packets intended for the different users have different orders of arrivals from higher layers. If, the queuing delay is disregarded in the scheduling policy, some packets may end up incurring very large delays at the transmitter's buffer (or worse, be dropped). Hence, failure to take the queuing delay into account when scheduling packets would most-likely result in a failure to satisfy the requirements of delay-sensitive applications (Video on Demand (VoD), circuit-emulated voice calls and networked gaming, to name a few) for which some packet delay constraints should be met (see our contribution in Chapter 4).

1.5 Thesis Outline and Summary of Results

For convenience in exposition, a thesis outline is illustrated in Figs. 1.13, 1.14 and a summary of results is provided in Tables 1.1, 1.2, 1.3 and ??.

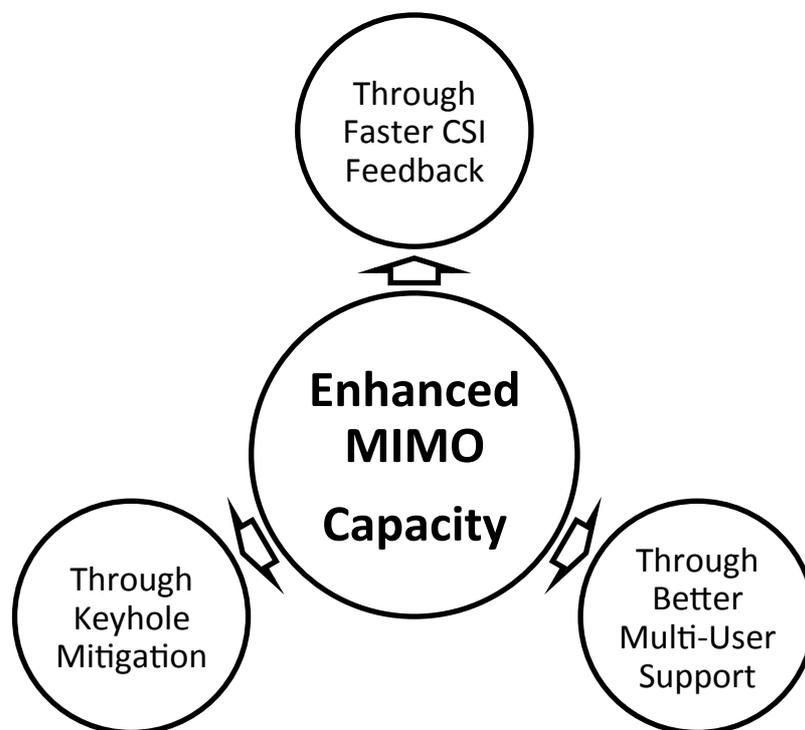


Figure 1.13: Approaches towards MIMO capacity enhancement considered in this thesis.

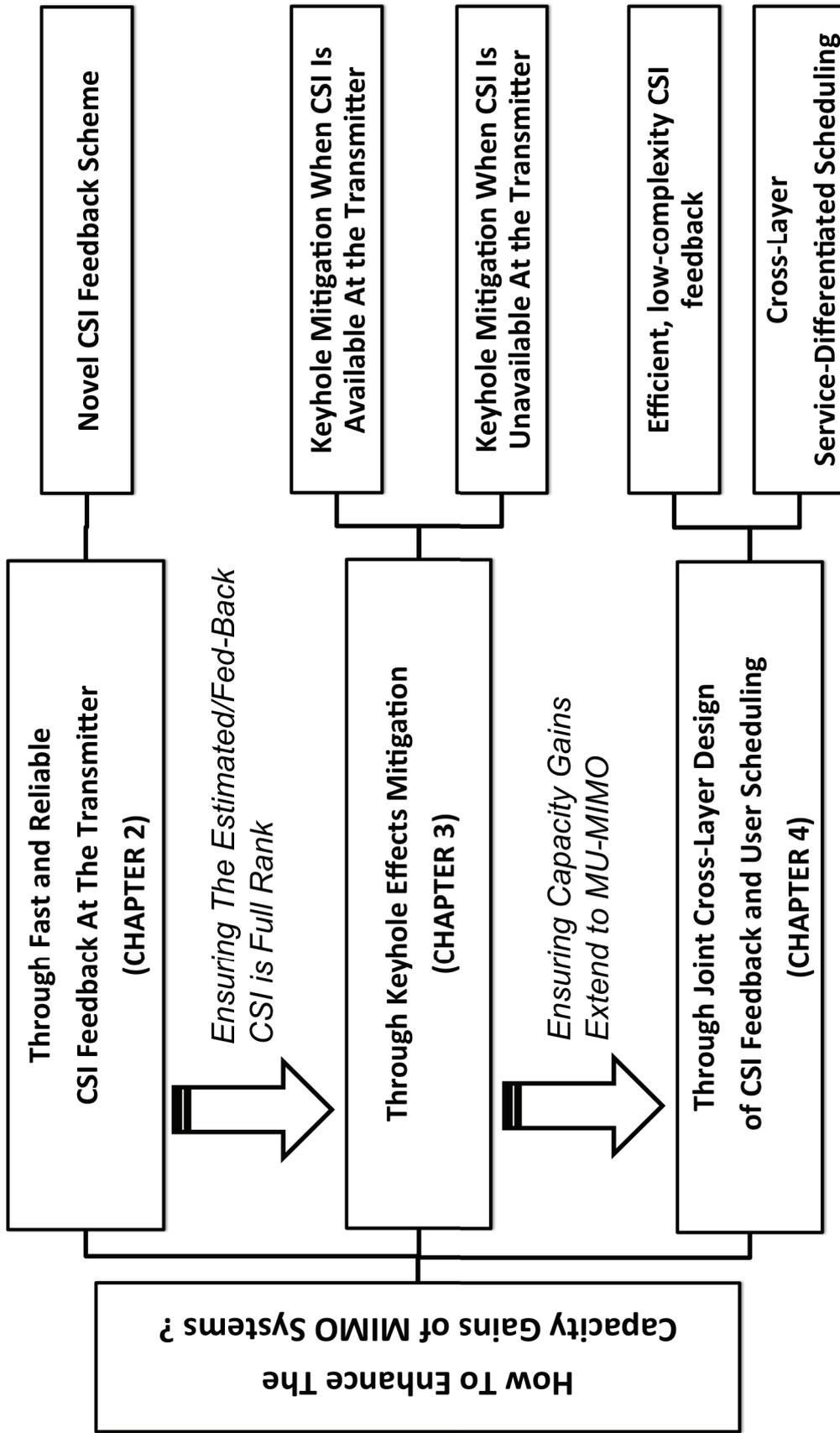


Figure 1.14: Thesis main outline.

Table 1.1: Outline of Chapter 2

<i>Research Problem</i>	<ul style="list-style-type: none"> • In Closed-loop MIMO, it is crucial to minimize the duration of the CSI estimation and feedback phase. • Thus, our aim is to achieve a <i>delay-free</i> CSI estimation and feedback phase.
<i>Conventional Approach</i>	<ul style="list-style-type: none"> • In [21], a delay-free feedback scheme called Echo-MIMO has been proposed for Closed-Loop MIMO systems, where the receiver echoes the received signal <i>on the fly</i> to the transmitter without any processing.
<i>Limitations</i>	<ul style="list-style-type: none"> • Though Echo-MIMO allows for a reduced feedback latency, it comes at high power-and-bandwidth costs, as two MIMO transmissions are required in the feedback phase to send two full CSIs related to the inward (BS \rightarrow MS) and the outward (MS \rightarrow BS) channels.
<i>Proposed Solution</i>	<ul style="list-style-type: none"> • In [22], we present a feedback scheme that preserves the advantages of Echo-MIMO while requiring only one feedback transmission to transmit both CSIs, at no extra transmit power or bandwidth costs.
<i>Summary of Results</i>	<ul style="list-style-type: none"> • A delay-free feedback with only 1 feedback transmission (unlike the conventional scheme where 2 feedback transmissions are required) • Enhanced CSI estimation reliability, as in the proposed scheme we do not echo the noise back to the transmitter.

Table 1.2: Outline of Chapter 3

<i>Research Problem</i>	<ul style="list-style-type: none"> • A MIMO <i>keyhole</i> is a propagation environment such that the channel gain matrix has unit rank (single degree of freedom), <i>irrespective</i> of the number of deployed antennas or their spacing, thereby reducing the MIMO channel capacity to that of a SISO channel.
<i>Conventional Approach</i>	<ul style="list-style-type: none"> • None. To the best of our knowledge, the related literature seems to consider such degeneration hopeless. Related works are limited to performance analysis under the keyhole effect: [23–31], to name a few.
<i>Limitations</i>	<ul style="list-style-type: none"> • Not Applicable (As there are no conventional schemes to compare with).
<i>Proposed Solution</i>	<ul style="list-style-type: none"> • In [32], we demonstrates that relay-assisted MIMO systems can mitigate keyhole effects. • Precisely, provided that the source-relay channel is keyhole-free, we show that there exists a “cutoff” relay transmit power above which keyhole effects can be mitigated <i>even when both the source-destination and the relay-destination channels incur keyhole effect</i>. • We devise the closed form of this power threshold as function of the source transmit power and the channel matrices brought into play in the relay channel. • Later in [33], we provide enhancements to the previous idea by providing a power allocation scheme that does not require any CSI (channel matrix) knowledge.
<i>Summary of Results</i>	<ul style="list-style-type: none"> • Keyhole effect is properly dealt with as we can achieve a linear scaling growth of the capacity with the number of transmit antennas. • Further, the optimized solution in [33] does not require any CSI knowledge, thereby making it very fit for practical scenarios as no signaling costs are incurred.

Table 1.3: Outline of Chapter 4

<i>Research Problems</i>	<ul style="list-style-type: none"> • On joint scheduling and feedback in MU-MIMO.
<i>Conventional Approach</i>	<ul style="list-style-type: none"> • In order for the transmitter to perform user selection, conventional works [20, 34] assume the knowledge of the CSI related to all candidate users to be available at the transmitter. • Traditional scheduling approaches aim at maximizing capacity [20] or ensuring fairness among users [34]. By doing so, they disregard higher-layer QoS requirements, such as packet delays.
<i>Limitations</i>	<ul style="list-style-type: none"> • Costly and inefficient feedback procedure: Why would, say, 20 users feed back their CSI when only 4 will be scheduled for transmission? • Further, the conventional physical-layer scheduling do not consider the order at which packets arrive, nor their individual delay constraints (depending on their respective guaranteed quality of service (QoS) requirements).
<i>Proposed Solution</i>	<ul style="list-style-type: none"> • We provide in [35, 36] a power-and-bandwidth efficient feedback scheme in which <i>only likely-to-be-scheduled users</i> feed back their CSI, thereby reducing the number of required feedbacks and the computational burden of exhaustive search for best users at the transmitter's side. • We show in [35, 36] that conventional sum-capacity maximizing scheduling policies fall short to meet the requisites of delay-sensitive applications, and we provide appropriate scheduling scheme for such-constrained users.
<i>Summary of Results</i>	<ul style="list-style-type: none"> • Signaling and complexity burden reduction owing to the proposed feedback scheme. • By-order-of-magnitude delay reduction owing to the proposed QoS-aware scheduling scheme. • The proposed scheme, however, incurs a slight capacity decrease owing to the fact that the search for to-be-scheduled users is only a subset of the total available users. This decrease in capacity trades for the cost and complexity reduction owing to the proposed feedback scheme.

Chapter 2

On Full CSI Feedback in Closed-Loop MIMO

2.1 Introduction

Multiple-input-multiple-output (MIMO) systems provide high capacity gains when channel state information (CSI) is available to both channel endpoints¹ [15], [10], by means of adaptive transmission (e.g. water-filling [37]). Perhaps the main challenge of an adaptive bidirectional communication is for both endpoints to acquire *accurate* and *timely* CSI estimates of their respective inward² and outward³ channels. While, traditionally, inward CSI is somewhat easily estimable (e.g. from a received training sequence [11]), estimating outward CSI is challenging at best.

In practice, if the variations of transmitter's outward channel are slow enough, the receiver (for which this channel is inward, thus estimable) may feed its CSI estimation back to the transmitter, a scheme commonly-known as *Closed-Loop MIMO*. Fig. 2.1 illustrates this concept. A popular feedback approach in Closed-Loop MIMO systems is *Quantized Feedback* (QF) [13], where the receiver determines from the estimated CSI the best beamforming weight from a codebook and sends to the transmitter the index of the corresponding codeword. While such a processing tremendously reduces the feedback rate, its computational complexity brings forth prohibitive delays at the receiver's side.

¹An endpoint being a transmitter or a receiver

²An inward channel is a channel where the endpoint is on the receiving side

³An outward channel is a channel where the endpoint is on the transmitting side

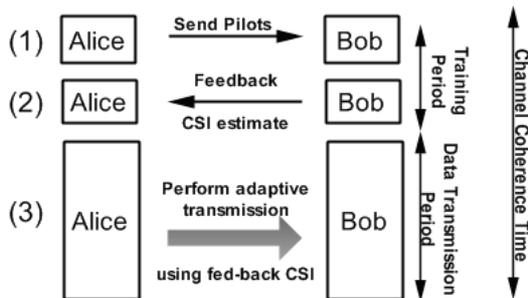


Figure 2.1: Phases brought into play in Closed-Loop MIMO during channel's coherence time

Recently, “Echo-MIMO”, a delay-free feedback scheme, has been proposed [21], where the receiver echoes the received signals *on the fly* back to the transmitter without any significant processing. Then, the receiver is required to transmit its own training symbols (these are needed to estimate the transmitter's inward channel). By efficiently exploiting the two received snapshots, the transmitter is capable of estimating the CSI related to both its inward and outward channels. Such a method virtually takes out any processing delays at the receiver's side. Further, it provides the transmitter with full CSI (rather than a quantized version), thereby achieving higher multiplexing gains during the adaptive transmission phase [21]. Yet, Echo-MIMO has drawbacks of its own, too. It comes at high power-and-bandwidth costs, since the fed-back CSI is retransmitted using as many antennas as for “regular” MIMO transmissions. Besides, two transmissions are required from the receiver so that the transmitter can estimate the channels. Finally, the noise of the inward channel is also echoed to the transmitter, which - as will be explained later-, may degrade the channel estimation accuracy.

In this chapter, based on the concept of Echo-MIMO, we introduce *Transparent Inband Feedback* (TIF), a two-way MIMO scheme that overcomes the shortcomings of the former while preserving its advantages. The proposed feedback scheme is labeled *Inband* as the echoed signals and the receiver's are combined, therefore no dedicated feedback channel is required. It is labeled *Transparent* because the signals are projected onto orthogonal signal subspaces prior to their combination (thereby making their separation lossless at the transmitter's side), and because it does not require any extra bandwidth nor extra transmit power.

2.2 On The Coherence Time of A Wireless MIMO Channel

In this section, we explain the notion of channel coherence and its relation to the feedback frequency in MIMO systems.

2.2.1 The Doppler Effect

The state of a communication channel may change in time owing to a variety of reasons. Such variations are particularly significant in wireless channels, owing to the mobility of the transmitter and/or the receiver.

The time variations of the channel that arise from the transmitter's/receiver's motion are called the *Doppler effect* or the *Doppler shift*. Basically, a shift in frequency is observed as follows: the signal's frequency is increased when the receiver approaches the transmitter and is decreased otherwise.

The Doppler shift can be intuitively explained as follows. When the receiver is getting closer to the transmitter, a transmitted electromagnetic wave would take slightly less time (then the previous one) to reach the receiver. Such time difference creates an increase in the frequency. Thus, even though the frequency at which the signals are emitted is unchanged, the frequency of the signals at the reception is increased. A similar justification can be provided for the case where the receiver is recedes from the transmitter.

2.2.2 Why Is It Necessary That The Fed-Back CSI Matches The True Channel In Closed-Loop MIMO ?

Throughout thesis, we have reiterated the importance of CSI availability at the transmitter's side. But, as we explained earlier, the wireless channel changes frequently. Therefore, if the CSI feed-back is late, there is a chance the CSI would not match the true channel.

One might wonder what impact would have a CSI mismatch on the channel capacity.

Recall that the capacity of a point-to-point closed-loop MIMO system is given by:

$$C = \max_{\mathbf{Q}_x} \log_2 \det \left(\mathbf{I}_M + \frac{1}{\sigma^2} \mathbf{H} \mathbf{Q}_x \mathbf{H}^\dagger \right), \quad (2.1)$$

where:

- M denotes the number of receive antennas,
- \mathbf{H} denotes the channel matrix,
- \mathbf{Q}_x denotes the power allocation (input covariance matrix) at the transmitter,
- σ^2 denotes the noise power.

Let $\mathbf{e} \triangleq \mathbf{H} - \tilde{\mathbf{H}}$, where $\tilde{\mathbf{H}}$ denotes the outdated CSI value. Then, the channel feedback delay will impact the channel capacity as follows:

$$C_{\text{Delay}} = \log_2 \det \left(\mathbf{I}_M + \frac{1}{\sigma^2 + \|\mathbf{e}\|^2} \tilde{\mathbf{H}} \mathbf{Q}_x \tilde{\mathbf{H}}^\dagger \right). \quad (2.2)$$

In other words, the feedback delay will decrease the channel capacity by:

- increasing the noise power owing to the channel error \mathbf{e} ,
- possibly making the optimal power allocation sub-optimal: The covariance input \mathbf{Q}_x is optimal for \mathbf{H} , but it may not necessarily be optimal for channel $\tilde{\mathbf{H}}$.

2.2.3 Then, How Often Should The CSI Be Fed Back To The Transmitter ?

In theory, the channel could change into a new realization at any moment. If the channel has to be tracked at the slightest change, then clearly closed-loop MIMO may become too difficult to realize in practice.

However, even though the channel realizations may change instantaneously, it is rather unlikely that the channel realizations at times t and $t + \delta t$ (where δt is an infinitesimal time period) may be significantly different. Therefore, there exists a period of time during which the channel may be deemed constant. Such period is the afore-described *channel coherence time*.

2.2.4 Motivation of Our Research: The Need For A Fast CSI Feedback

From this observation, emerged a new channel model that is widely used in closed-loop MIMO: *The block-fading model*. In a block-fading channel of coherence time T_c , the channel is modeled as constant during T_c seconds after which it changes into a new realization.

T_c is given by:

$$T_c \triangleq \frac{c}{8f_c\nu}, \quad (2.3)$$

where c , f_c and ν denote the speed of light (3×10^8 m/sec), the carrier frequency in Hz and the node speed in m/sec.

Let us take an example. Say the carrier frequency is 2.4 GHz and receiver's speed is 20 km/hr. Under such assumptions, the channel's coherence time is $T_c = 2.8$ ms.

Table 2.1: Main Simulation Parameters.

Number of transmit antennas (M)	2
Number of receive antennas (N)	2
Distance between transmit antennas (in wavelengths)	6
Distance between receive antennas (in wavelengths)	0.4
Carrier frequency (f_c , in GHz)	6
Number of Paths	20
Mobile node's velocity (in km/hr)	5, 10, 20, 50

Therefore, if the transmission rate is $R = 128$ Kbps, at most 358 bits can be transmitted before the channel changes into a new realization⁴.

In closed-loop MIMO, this upper bound of transmittable bits includes *both* the pilot symbols transmitted during the channel estimation phase and the data symbols transmitted during the data transmission phase. More to the point, this upper bound does not account for the processing delay at the receiver's side. Therefore, it is imperative that as little time as possible should be allotted to the channel estimation and feedback phase, so that most of the coherence time T_c be used in the data transmission phase.

For illustrative purposes, we provide below in Figs. 2.2, blockfading, the evolution of the power gain a channel coefficient h_{11} in time for different mobility patterns, as well as the corresponding block-fading model. The channel coefficients were generated in accordance with the 3GPP spatially-correlated model (SCM) with simulations parameters as specified in Table 2.1.

2.3 System Model And Preliminaries

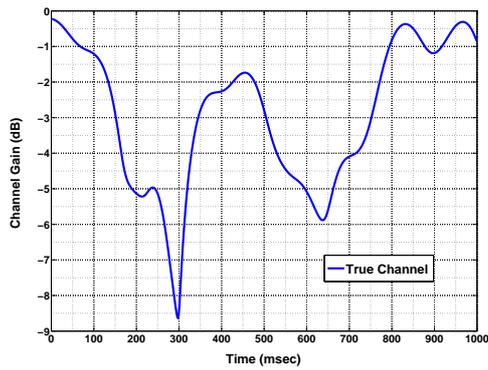
We start by defining the system model, then we briefly outline the conventional scheme and we point out some of its limitations that will be tackled in this work.

2.3.1 System Model

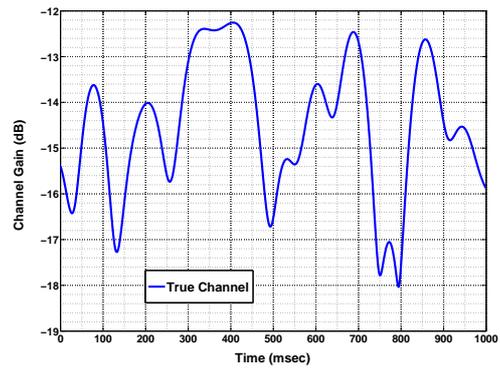
We consider the conventional point-to-point MIMO system model of [21], which we shall briefly outline. Because we are considering a bidirectional communication, the notions of “transmitter” and “receiver” may lead to confusion. Thus, we denote by Alice and Bob the two terminals brought in play by our model, having M and N antennas, respectively. W.r.t. Alice, inward (\mathbf{H}_{ba}) and outward (\mathbf{H}_{ab}) channels are assumed to be Rayleigh-faded, frequency-flat and time-varying, obeying the conventional block-fading law of coherence time⁵ T . Two-way additive

⁴ $(2.8 \times 10^{-3}) \times (128 \times 10^3) = 358.4$.

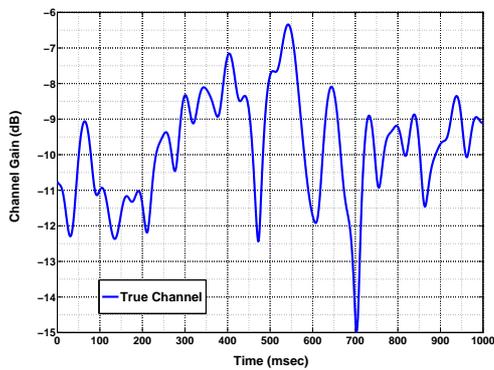
⁵ T is the time interval during which the channel remains constant, before changing to a new independent realization.



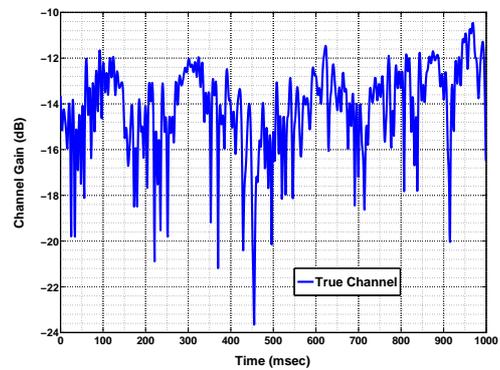
(a) The channel at 5 km/hr



(b) The channel at 10 km/hr

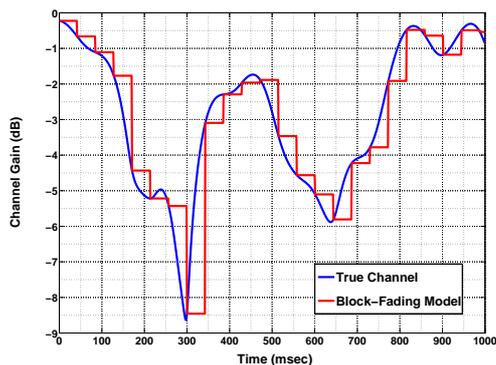


(c) The channel at 20 km/hr

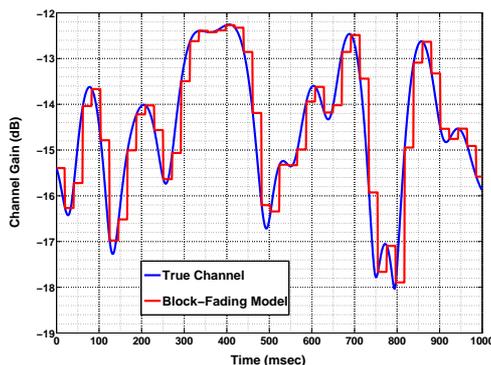


(d) The channel at 50 km/hr

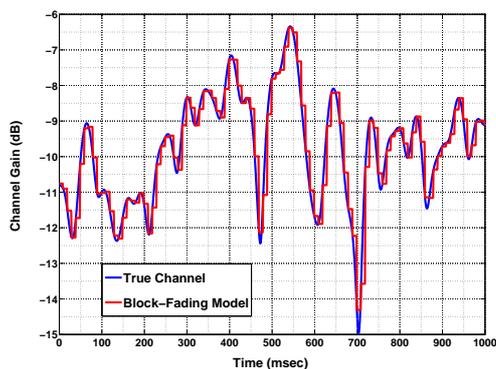
Figure 2.2: One second in the life of four wireless channels. The Doppler effect for different receive mobility patterns.



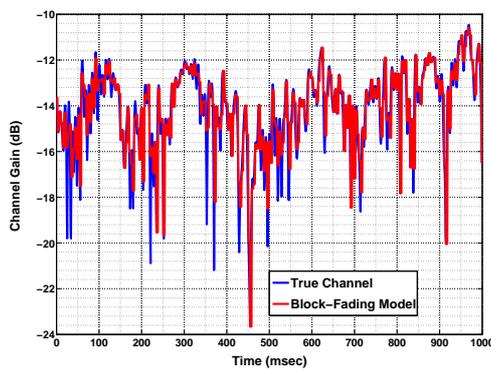
(a) The channel at 5 km/hr



(b) The channel at 10 km/hr



(c) The channel at 20 km/hr



(d) The channel at 50 km/hr

Figure 2.3: Comparison of the block-fading model and the true channel realizations. Observe how the coherence time is decreased with mobility. At high speed, the block-faded channel has the same realization as the true model, thereby making closed-loop MIMO very impractical (the channel variations being too fast to track).

noises ($\mathbf{Z}_a, \mathbf{Z}_b$) are modeled as Zero Mean White Gaussian (ZMWG), and are assumed to have roughly equal noise variances, i.e. $\sigma_a^2 \approx \sigma_b^2$. Transmission of a training block \mathbf{X} (matrix of pilot symbols) of L signal vectors using an $M \times N$ MIMO system can be modeled as:

$$\mathbf{Y} = \sqrt{\frac{\rho}{M}} \mathbf{H} \mathbf{X} + \mathbf{Z} \quad (2.4)$$

where ρ is the average transmit power, $\mathbf{X} \in \mathbb{C}^{M \times L}$ is the transmitted training matrix, $\mathbf{H} \in \mathbb{C}^{N \times M}$ is the complex channel gain matrix, $\mathbf{Y} \in \mathbb{C}^{N \times L}$ is the received signal matrix and $\mathbf{Z} \in \mathbb{C}^{N \times L}$ denotes the noise matrix.

At the reception, the following estimators may be used to estimate the channel [11]:

- The Least Squares Estimator (LS)

$$\hat{\mathbf{H}} = \sqrt{\frac{M}{\rho}} \mathbf{Y} \mathbf{X}^H (\mathbf{X} \mathbf{X}^H)^{-1} \quad (2.5)$$

- The Minimum Mean Squared Error Estimator (MMSE)

$$\hat{\mathbf{H}} = \sqrt{\frac{M}{\rho}} \mathbf{Y} \mathbf{X}^H \left((\mathbf{X} \mathbf{X}^H)^{-1} + \frac{M}{\rho \sigma^2} \mathbf{I}_M \right)^{-1} \quad (2.6)$$

The coherence time T is decomposed into two phases: a phase of duration T_p where channel is estimated by Bob and fed back to Alice, and a phase of duration $T_d = T - T_p$ where Alice uses the fed-back CSI to send a data block of length T_d via capacity-achieving adaptive transmission. Obviously, we are interested in having $T_d \gg T_p$ so as to perform as many adaptive transmissions as possible with the fed-back CSI, before the channel changes from state. Fig. 2.1 illustrates the considered scenario.

2.3.2 Echo-MIMO: An Overview

We briefly overview a part of the feedback protocol proposed in [21] in relation with the stated problem:

1. Alice sends her training matrix, \mathbf{X}_a , to Bob. For channel identifiability, \mathbf{X}_a should be full-rank (M) and such that ($L \geq M$) [11], [38]. Following (1), Bob receives:

$$\mathbf{Y}_{ab} = \sqrt{\frac{\rho}{M}} \mathbf{H}_{ab} \mathbf{X}_a + \mathbf{Z}_b \quad (2.7)$$

2. Bob amplifies the signal to palliate the pathloss and shadowing effects, then echoes it back to Alice. Alice receives:

$$\mathbf{Y}_{aba} = \sqrt{\frac{\rho}{N}} \mathbf{H}_{ba} \mathbf{H}_{ab} \mathbf{X}_a + \overbrace{\sqrt{\frac{\rho}{N}} \mathbf{H}_{ba} \mathbf{Z}_b + \mathbf{Z}_{a_1}}^{\mathbf{Z}_1} \quad (2.8)$$

(We assume same transmit powers for Bob and Alice, for simplicity.)

3. Bob sends his own training matrix \mathbf{X}_b . Alice receives

$$\mathbf{Y}_{ba} = \sqrt{\frac{\rho}{N}} \mathbf{H}_{ba} \mathbf{X}_b + \mathbf{Z}_{a_2} \quad (2.9)$$

4. Alice, knowing the training matrix \mathbf{X}_b , estimates \mathbf{H}_{ba} (e.g. using (3)). Then, she plugs the estimate ($\hat{\mathbf{H}}_{ba}$) in (3) and, knowing her own training matrix \mathbf{X}_a , she estimates the fed-back CSI \mathbf{H}_{ab} from (5).

2.3.3 Comments on Echo-MIMO

Certainly, Echo-MIMO provides significant gains -compared with Quantized Feedback- in terms of processing delays at Bob's side, as the received signal is echoed on the fly. However, it has at least the following shortcomings:

- Two transmissions ((5), (6)) are required from Bob before Alice can estimate any channel.
- The noise \mathbf{Z}_b is echoed as well, therefore \mathbf{Z}_1 in (5) involves noises of both Alice's inward and outward channels. Thus, there is a legitimate concern on whether the echoed noise significantly affects the estimation accuracy of channel \mathbf{H}_{ab} (and in turn the achievable capacity during the adaptive transmission).

These limitations will be further investigated in the following section.

2.4 Proposed Scheme

In what follows, we present a feedback scheme that overcomes the previously mentioned limitations while preserving the advantages of Echo-MIMO.

2.4.1 Transparent Inband Feedback (TIF)

To reduce the number of feedback transmissions from two to one, we suggest that the two signals be combined together after being projected on subspaces spanned by two orthogonal matrices \mathbf{P} and \mathbf{Q} . These matrices are required to be full rank (for channel identifiability) and to lie in the null space of each other, i.e. $\mathbf{P}\mathbf{Q} = \mathbf{0}_{L \times L}$. Owing to the rank-nullity theorem, a requisite for such matrices to exist is that $K \geq L \geq M + N$, K being Bob's training sequence length. Besides, we require these matrices to be unit norm, so that noise be not enhanced by the processing at Alice's side. The proposed feedback scheme consists of the following steps:

- **As in Echo-MIMO**, Alice sends her signal matrix \mathbf{X}_a to Bob through the channel \mathbf{H}_{ab} . Bob receives:

$$\mathbf{Y}_{ab} = \mathbf{H}_{ab}\mathbf{X}_a + \mathbf{Z}_b \quad (2.10)$$

- **Unlike Echo-MIMO**, instead of echoing \mathbf{Y}_{ab} as received, Bob estimates the channel \mathbf{H}_{ab} (knowing \mathbf{X}_a) and uses this estimate to reproduce a less-noisier replica of \mathbf{Y}_{ab} :

$$\tilde{\mathbf{Y}}_{ab} = (\mathbf{H}_{ab} + \Delta\mathbf{H}_{ab})\mathbf{X}_a \quad (2.11)$$

where $\Delta\mathbf{H}_{ab}$ denotes Bob's estimation error. It will be shown, later, that this operation significantly enhances the estimation accuracy at Alice's side.

- **Unlike Echo-MIMO**, instead of sending \mathbf{Y}_{ab} and \mathbf{X}_b in two transmissions, Bob sends the following mixture:

$$\mathbf{V} = \sqrt{\frac{\rho_a}{N}}\tilde{\mathbf{Y}}_{ab}\mathbf{P} + \sqrt{\frac{\rho_b}{N}}\mathbf{X}_b\mathbf{Q}^H \quad (2.12)$$

where $\mathbf{P} \in \mathbb{C}^{L \times K}$, $\mathbf{Q} \in \mathbb{C}^{K \times L}$, ρ_a and ρ_b denote the average transmit powers dedicated to the echo and Bob's training signal, respectively and H denotes the Hermitian (conjugate transpose) operator.

- Alice receives:

$$\mathbf{Y}_{ba} = \mathbf{H}_{ba}\mathbf{V} + \mathbf{Z}_a \quad (2.13)$$

$$= \sqrt{\frac{\rho_a}{N}} \mathbf{H}_{ba} (\mathbf{H}_{ab} + \Delta \mathbf{H}_{ab}) \mathbf{X}_a \mathbf{P} + \sqrt{\frac{\rho_b}{N}} \mathbf{H}_{ba} \mathbf{X}_b \mathbf{Q}^H + \mathbf{Z}_a$$

Now we show how Alice can estimate both unknown channels \mathbf{H}_{ab} and \mathbf{H}_{ba} without requiring any further transmissions from Bob. This estimation is achieved in two steps:

- **Multiplying the received signal by $\mathbf{Q}(\mathbf{Q}^H \mathbf{Q})^{-1}$ zero-forces the term in \mathbf{P} :**

$$\mathbf{Y}_{ba} \mathbf{Q}(\mathbf{Q}^H \mathbf{Q})^{-1} = \sqrt{\frac{\rho_b}{N}} \mathbf{H}_{ba} \mathbf{X}_b + \mathbf{Z}_a \mathbf{Q}(\mathbf{Q}^H \mathbf{Q})^{-1} \quad (2.14)$$

Knowing Bob's training matrix \mathbf{X}_b , Alice now can estimate the channel \mathbf{H}_{ba} .

- **Multiplying the received signal by $\mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1}$ zero-forces the term in \mathbf{Q}^H :**

$$\mathbf{Y}_{ba} \mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1} = \sqrt{\frac{\rho_a}{N}} \mathbf{H}_{ba} \mathbf{H}_{ab} \mathbf{X}_a + \mathbf{Z}_2 \quad (2.15)$$

where

$$\mathbf{Z}_2 = \sqrt{\frac{\rho_a}{N}} \mathbf{H}_{ba} \Delta \mathbf{H}_{ab} \mathbf{X}_a + \mathbf{Z}_a \mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1} \quad (2.16)$$

Using the channel estimate $\hat{\mathbf{H}}_{ba}$ obtained in the previous step and the training sequence \mathbf{X}_a , Alice can estimate the channel \mathbf{H}_{ab} . Fig. 2.4 summarizes the proposed scheme.

2.4.2 TIF vs Echo-MIMO, A Comparative Study

Feedback in TIF is less-noisier than that in Echo-MIMO

We demonstrate the following:

Lemma 1 *The noise matrix \mathbf{Z}_1 in (5) is zero-mean, and its variance is given by:*

$$\sigma_{echo}^2 = \frac{\rho_a \sigma_a^2}{NMK} \sum_{i=1}^r \lambda_i^2 + \frac{\sigma_b^2}{MK}$$

where σ_a^2, σ_b^2 denote the noise variances of Alice's outward and inward channels, respectively, and $(\lambda_i)_{1 \leq i \leq r}$ denote the eigenvalues of matrix $\mathbf{H}_{ba} \mathbf{H}_{ba}^H \in \mathbb{C}^{N \times N}$ of rank $r \leq N$.

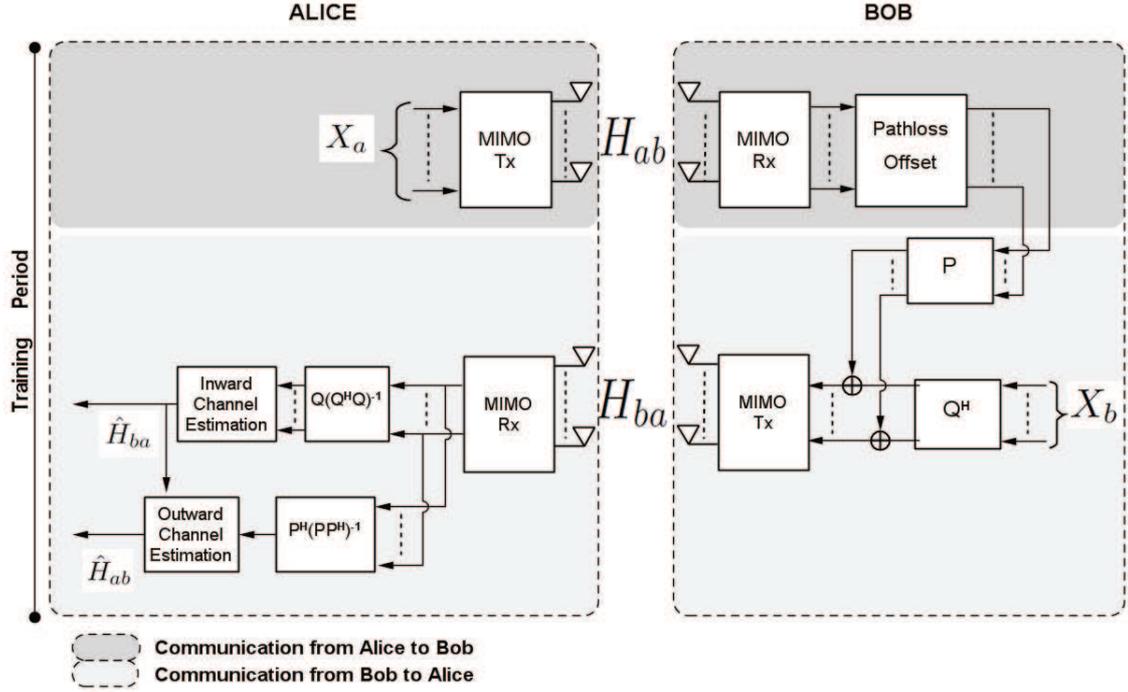


Figure 2.4: Proposed two-way communication scheme for CSI feedback in Closed-Loop MIMO

Proof The fact that \mathbf{Z}_1 is zero-mean is straightforward (direct application of the expectation, channel matrices and \mathbf{P} are constants and noises \mathbf{Z}_b and \mathbf{Z}_{a_1} are zero-mean). The noise variance is given by:

$$\sigma_{echo}^2 \triangleq \frac{1}{MK} \text{Tr} (E \{ \mathbf{Z}_1 \mathbf{Z}_1^H \}) \quad (2.17)$$

$$= \frac{\rho_a \sigma_a^2}{NMK} \text{Tr} (\mathbf{H}_{ba} \mathbf{H}_{ba}^H) + \frac{\sigma_b^2}{MK} \quad (2.18)$$

$$= \frac{\rho_a \sigma_a^2}{NMK} \sum_{i=1}^r \lambda_i^2 + \frac{\sigma_b^2}{MK} \quad (2.19)$$

Q.E.D.

Lemma 2 Assume that the estimation error $\Delta \mathbf{H}_{ab}$ has zero mean and variance σ_H^2 [11], and that \mathbf{X}_a is unitary. Then, the noise matrix \mathbf{Z}_2 in (12), (13) is zero-mean and, using the same notations as Lemma 1, its variance is given by:

$$\sigma_{TIF}^2 = \frac{\rho_a \sigma_H^2}{NMK} \sum_{i=1}^r \lambda_i^2 + \frac{\sigma_a^2}{MK}$$

Proof Similar to that of Lemma 1.

Theorem 1 Assume a unit transmit power, i.e. $\rho_a = 1$. Then, irrespective of the channel \mathbf{H}_{ba} , feedback in TIF is always less-noisier than that in Echo-MIMO, i.e.:

$$\sigma_{TIF}^2 < \sigma_{echo}^2, \forall \mathbf{H}_{ba}$$

Proof In Section II, we have already assumed noises to have similar noise variances. This yields:

$$\sigma_{TIF}^2 - \sigma_{echo}^2 = \frac{\rho_a (\sigma_H^2 - \sigma_a^2)}{NMK} \sum_{i=1}^r \lambda_i^2 \quad (2.20)$$

From [11], we have:

$$\sigma_H^2 = \frac{1}{1 + \frac{\rho_a}{M\sigma_a^2}L} \quad (2.21)$$

Therefore:

$$\sigma_H^2 - \sigma_a^2 = \frac{\sigma_a^2 (1 - \sigma_a^2 - \frac{L}{M})}{\sigma_a^2 + \frac{L}{M}} \quad (2.22)$$

In TIF, owing to channel identifiability requirement, we have $L \geq N + M > M$, therefore $\frac{L}{M} > 1$ and the fact that $\sigma_a^2 > 0$ concludes the proof.

Hence, we can see that the noise power in the proposed feedback scheme is less than that in Echo MIMO. This improves the estimation reliability of both channels \mathbf{H}_{ba} and \mathbf{H}_{ab} , as will be later observed in the numerical examples.

TIF Is No Less Power-Efficient Than Echo-MIMO

We express power efficiency in terms of how many symbols are transmitted in both cases for the same transmit power. Fig. 2.5 compares the time slots in both TIF and Echo-MIMO and the involved transmit powers. It is self-evident that allotting the same power for both schemes during a time slot of length $N + M$ symbols implies that $\rho = \rho_a + \rho_b$. In Echo-MIMO, this power is used to transmit N pilots of Bob and M pilots of Alice, i.e., a total $N + M$ pilots, while in TIF, it is used to transmit $N + M$ pilots of Alice and $N + M$ pilots of Bob, i.e., twice as much as in Echo-MIMO ($2 \times (N + M)$ pilots). In theory, only N pilots (resp. M pilots) are required to fully identify the channel \mathbf{H}_{ba} (resp. \mathbf{H}_{ab}). In such a case, both TIF and Echo-MIMO have the same power efficiency as the total pilot duration from Bob to Alice is $N + M$ symbols. In

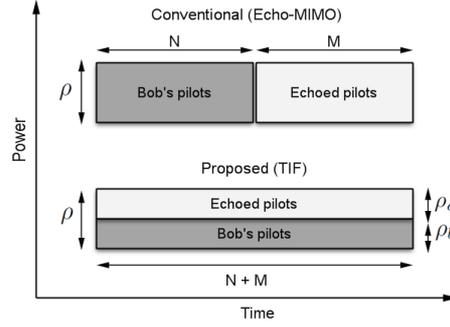


Figure 2.5: Comparison of the time slots of Echo-MIMO and TIF in terms of transmit power and number of transmitted pilots

practice, however, more pilots $N' > N$ (resp. $M' > M$) may be required to ensure reliable channel identifiability. In such a case, Echo-MIMO based systems are required to increase their power to transmit $N' + M'$ pilots during $N' + M'$ -symbol durations. In TIF, however, $N' + M'$ pilots may be transmitted during $N + M$ -symbol durations only, owing to channel orthogonality (provided that $L \geq N' + M'$, the channel identifiability requirement). Thus, if more pilots are needed than the theoretical minimum, TIF is more power efficient than Echo-MIMO.

Cost of TIF

Unlike Echo-MIMO (where Bob echoes Alice's signal on the fly), TIF requires some processing at Bob's side. Channel needs to be estimated at Bob's side (linear estimation, $\mathcal{O}(N)$), and overall three additional matrix multiplications ($\mathcal{O}(N^3)$) and one matrix addition ($\mathcal{O}(N)$) are required, vis-à-vis Echo-MIMO. However, we believe this extra processing can be tolerated as it trades for a significant increase in the estimation accuracy of Alice's both inward at outward channels.

2.5 Numerical Examples

In this section, we report results averaging 10^5 runs performed through computer simulation. A 4×4 MIMO system was considered, and channels were modeled following Dent's Rayleigh block fading model [39]. By definition, the channel's coherence time, T_c , is given by:

$$T_c \triangleq \frac{c}{8f_c\nu}, \quad (2.23)$$

where c , f_c and ν denote the speed of light (3×10^8 m/sec), the carrier frequency in Hz and the node speed in m/sec. In our simulations, the carrier frequency was set to 2.4 GHz and Bob's

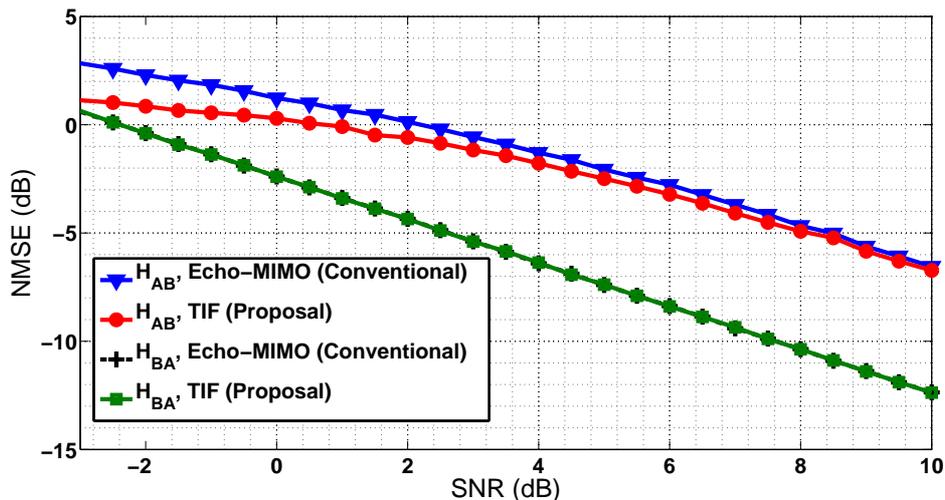


Figure 2.6: Channel estimation accuracy in terms of NMSE (dB)

speed to 20 km/hr. Furthermore, the transmission rate was set to 128 Kilo (BPSK) symbols per second (Ksps). Therefore, channel coherence time allows for transmitting roughly 358 BPSK symbols, see section 2.2.4. Finally, orthonormal sets of pilots were obtained from a Hadamard matrix and 4 pilot symbols per antenna were transmitted every channel coherence time (thereby leaving roughly 350-symbol duration for data transmission, as a 4-symbol duration is dedicated to sending pilots from Alice to Bob and another 4-symbol duration is dedicated to sending pilots from Bob to Alice). Besides, out of concern for fairness (vis-à-vis Echo-MIMO), equal powers were allocated between the echoed pilots and Bob's pilots.

2.5.1 Evaluation of the Estimation Accuracy

Fig. 2.6 illustrates the estimation accuracy of Alice's both inward and outward channels in terms of the Frobenius-norm-based Normalized Mean Square Error (NMSE) in dB. Regarding the channel H_{ba} , we observe that both TIF and Echo-MIMO have similar estimation accuracies. This result confirms the previous intuition that the orthogonal projection operations by Alice are practically seamless and do not cause any information loss. As for the inward channel H_{ab} , we observe a better estimation accuracy of the proposed scheme than that of the conventional one. This is owing to the noise reduction at Bob's side in the proposed scheme. These results confirm the intuition that a proactive action to mitigate the effects of the noise in the received signal (echoing an almost noise-free signal) is better than a reactive action (on-the-fly echoing the received signal with its noise, then accounting for its variance in the weight of the LMMSE estimator at Alice's side).

However, it is noteworthy that the accuracy gains of the proposed compared with Echo-MIMO decrease as the SNR increases. We shall provide an intuitive and an analytical interpretation of this observation.

- Intuitively, we have mentioned that TIF out-performs Echo-MIMO owing to the noise reduction operation at Bob's side, which allows the latter to echo Alice's signal practically noise-free. However, when SNR increases, the noise degradation is less significant, in which case TIF and Echo-MIMO achieve similar performance.
- Analytically, we have shown in Theorem 1 that TIF outperforms Echo-MIMO whenever the estimation error at Bob's side, σ_H^2 , is smaller than the noise variance at Bob's side, σ_a^2 . Further, we have:

$$\frac{\sigma_a^2}{\sigma_H^2} = \frac{\sigma_a^2}{\frac{1}{1 + \frac{L\rho}{M\sigma_a^2}}} \quad (2.24)$$

$$= \sigma_a^2 + \frac{L\rho_a}{M} \quad (2.25)$$

Clearly, when the noise σ_a^2 decreases, so do the benefits of TIF w.r.t. to those of Echo-MIMO.

Thus, we may conclude the following: while in theory TIF always outperforms Echo-MIMO, the estimation accuracy gains are interesting when the noise at Bob's side is non-negligible.

2.5.2 Evaluation of The Spectral Efficiency

In the absence of estimation error, it is known that the spectral efficiency (channel capacity per Hertz of bandwidth) of a MIMO system with channel matrix \mathbf{H} , where the transmitter employs a linear precoding \mathbf{G} and the receiver employs a linear precoding \mathbf{F} , is given by:

$$C_{SVD|H} = \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} (\mathbf{FHG}) (\mathbf{FHG})^H \right), \quad (2.26)$$

where σ^2 denotes the noise variance. Whereas, in the absence of estimation error, the spectral efficiency is given by:

$$C_{SVD|\hat{H}} = \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2 + \sigma_H^2} (\hat{\mathbf{FHG}}) (\hat{\mathbf{FHG}})^H \right) \quad (2.27)$$

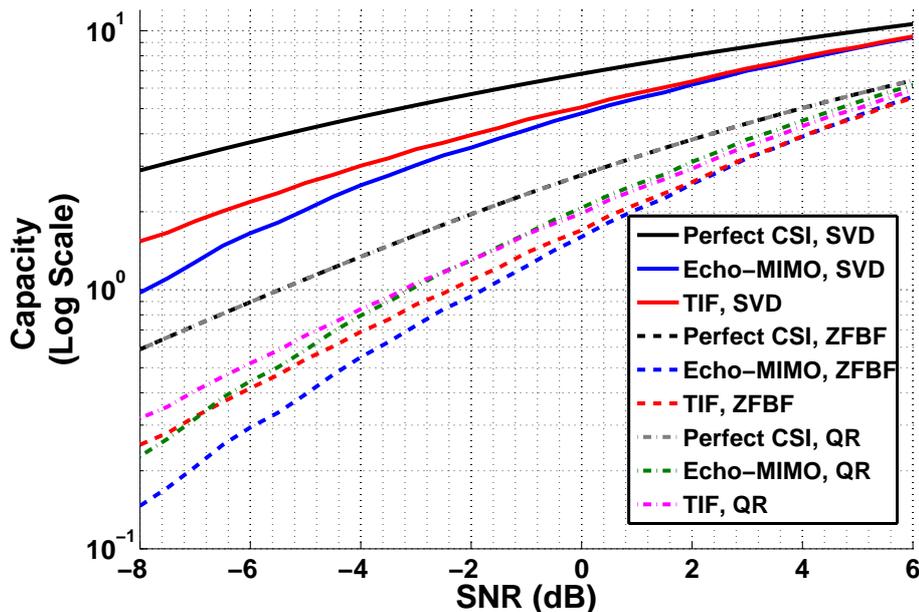


Figure 2.7: Spectral efficiency with perfect and estimated CSI. Different MIMO precoding schemes.

Table 2.2: Precoding weights at the transmitter and the receiver for different precoding schemes.

Precoding	Matrix Decomposition	Precoding At The Receiver (\mathbf{F})	Precoding At The Transmitter (\mathbf{G})
SVD	$\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}$	\mathbf{U}^\dagger	\mathbf{V}^\dagger
ZFBF	Not applicable	None	\mathbf{H}^\dagger
QR	$\mathbf{H} = \mathbf{Q}\mathbf{R}$	\mathbf{Q}^\dagger	\mathbf{R}^\dagger

From (2.27), as pointed out in [40], it appears that channel estimation error has two negative effects on the spectral efficiency formula:

1. The precoding matrices $\hat{\mathbf{F}}$ and $\hat{\mathbf{G}}$ become ill-matched with the channel matrix \mathbf{H}
2. The noise term is increased by the variance of the estimation power, thereby reducing the effective SNR.

In this paragraph, we evaluate the achievable spectral efficiency of the adaptive transmission using perfect and estimated CSI under different precoding schemes: SVD-based, Zero-Forcing Beamforming (ZFBF) and QR-decomposition-based precoding. For the afore-mentioned precoding schemes, the precoders at the transmitter and the receiver are summarized in Table 1, where \dagger denotes the matrix Moore-Penrose pseudo-inverse. It follows that the spectral efficiencies achieved

by such precoders in the absence of estimation error are given by:

$$C_{SVD} = \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \left(\mathbf{U}^\dagger \mathbf{H} \mathbf{V}^\dagger \right) \left(\mathbf{U}^\dagger \mathbf{H} \mathbf{V}^\dagger \right)^H \right) \quad (2.28)$$

$$C_{ZFBF} = \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \left(\mathbf{H}^\dagger \mathbf{H} \right) \left(\mathbf{H}^\dagger \mathbf{H} \right)^H \right) \quad (2.29)$$

$$C_{QR} = \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \left(\mathbf{Q}^\dagger \mathbf{H} \mathbf{R}^\dagger \right) \left(\mathbf{Q}^\dagger \mathbf{H} \mathbf{R}^\dagger \right)^H \right) \quad (2.30)$$

while the formulas in the presence of estimation error can be trivially derived from (2.27).

Fig. 2.7 illustrates the achievable spectral efficiency of the afore-mentioned schemes. The spectral efficiency has been plotted in logarithmic scale because these schemes achieve capacities with different orders of magnitudes. First, we observe that the proposed scheme (TIF) outperforms the conventional scheme (Echo-MIMO) for all three precoding schemes. Indeed, it has already been shown that TIF has a better estimation accuracy compared with Echo-MIMO. Such increased estimation accuracy yields a lower estimation error σ_H^2 and precoders that are better matched with the real channel \mathbf{H} . However, such capacity gains are more pronounced in the low SNR regime. As the SNR increases, we have seen that the estimation accuracy gains of TIF are decreased, because echoing the noise does not hurt Echo-MIMO much. Hence, we conclude once again that the proposed scheme is more suited for situations where the noise at Bob's side is non-negligible. Finally, we observe that the highest capacity gains are obtained with the SVD-based precoding, since the latter is the capacity achieving precoding [15] and is very sensitive to channel estimation accuracy. The smallest capacity gains are achieved with ZFBF as the latter requires no precoding at the transmitter, thereby becoming less sensitive to CSI inaccuracies.

2.6 Conclusion

In this chapter, based on the concept of Echo-MIMO, a low-latency power-and-bandwidth efficient feedback scheme was presented. The echoed signals are judiciously combined with the receiver's signals such that their separation at the transmitter is lossless, and that no extra transmit power nor bandwidth be required. In addition, we confirmed the intuition that feeding back an estimate of the received signal is better than echoing the received noisy signal on the fly and dealing with the noise effect upon feedback reception.

Chapter 3

On Keyhole Effect Mitigation Through Relay-Assisted Communications

3.1 Introduction

One reason behind the popularity of multiple-input-multiple-output (MIMO) systems is their high spectral efficiency [15], [7]. Assuming channel state information (CSI) availability, it is known that the capacity of a MIMO channel between an N_T -antenna transmitter and an N_R -antenna receiver is n times that of a single-input-single-output (SISO) channel, where n is the rank of the MIMO channel gain matrix, less than or equal to $\min\{N_T, N_R\}$ [7, 10, 15]. In particular, if the channel matrix has full rank (a so-called *rich-scattered* environment), channel capacity scales with $\min\{N_T, N_R\}$, i.e. $n = \min\{N_T, N_R\}$. From this perspective, it is interesting to determine conditions under which channel matrix has full rank, and to ensure that these conditions are met so that channel capacity is maximized.

While it has been long believed that decorrelating transmit antennas (e.g. by sufficiently spacing them) amply ensures a full-rank channel matrix [16, 17], recent works [16–18] demonstrated that in a so-called keyhole/pinhole scenario, channel matrix has unit rank, *even when its entries have zero correlations between each other*. Subsequently, the benefits of rich scattering are suppressed, and channel capacity scales as that of a SISO channel (i.e. $n = 1$).

Such a frustrating result was first theoretically predicted in [16, 17], and later verified through an experimental testbed in [18]. MIMO keyholes occur when radio waves come across metal obstacles with small holes only through which they can propagate (*spatial keyholes*, see Fig. 3.1). They are also encountered in urban environments with so-called *street canyons* (narrow streets

bordered by tall buildings), and in some indoor environments such as corridors, hallways and subway tunnels, settings which may act as single-moded waveguides at large distance from the source, thereby allowing only a single electromagnetic mode to pass through (*modal keyholes*). Finally, outdoor keyholes may also occur owing to a diffraction at rooftop edges (*diffraction-induced keyholes*) [17].

Because the keyhole channel has unit rank irrespective of fading correlations or the number of deployed antennas, such degeneration seems to have been thought of as irremediable. Indeed, related literature has been limited to system performance analyses in the presence of keyholes: achievable outage capacity regions [23, 24], performance analyses in multiple keyhole scenarios [25, 26], performance of STBC codes in keyhole environments [27–29], evaluation of level crossing rate (LCR) [30] and pairwise error probability [31], to name a few. There seems to be no related work that has attempted to provide a solution to such issue, perhaps with the exception of [17] where the authors pointed out that a horizontal arrangement of rooftop antenna arrays mitigates diffraction-induced keyholes. Regrettably, this solution proves inadequate to combat other kinds of keyholes (e.g. spatial keyholes or modal keyholes). To the best of our knowledge, there has been no universal solution for the MIMO keyhole problem, to date.

In this chapter, we investigate whether cooperative diversity (relay deployment) can mitigate keyhole effects. A MIMO relay channel is depicted in Fig. 3.2. It brings into play three nodes (source, relay and destination) and three channel gain matrices denoted by \mathbf{F} , \mathbf{G} and \mathbf{H} . In downlink channels of cellular networks, source and relay nodes are fixed base stations (BS) arbitrarily positioned by the network operator. Therefore, we assume that the source and the relay nodes are positioned such that the channel between them, \mathbf{F} , to enjoy rich scattering. Cases where **either** of the channels involving the destination (i.e. \mathbf{G} or \mathbf{H}) is keyhole-free are trivial, as either the direct link (source-destination) or the relayed link (source-relay-destination) are keyhole-free. Hence, we focus in this work on the more challenging scenario where **both** channels involving the destination (i.e. \mathbf{G} and \mathbf{H}) suffer from keyhole effects (i.e. have unit rank). This challenging assumption is also in line with practical scenarios as it makes no assumptions on the destination's mobility pattern or localization (in downlink channels of cellular networks, the destination is traditionally a mobile station (MS)). Finally, we only consider *degraded* relay channels, i.e. channels where the source-relay signal is better than the source-destination signal, as it is often the case [41].

Under such framework, we take aim at determining necessary and sufficient conditions for keyhole mitigation to be feasible.

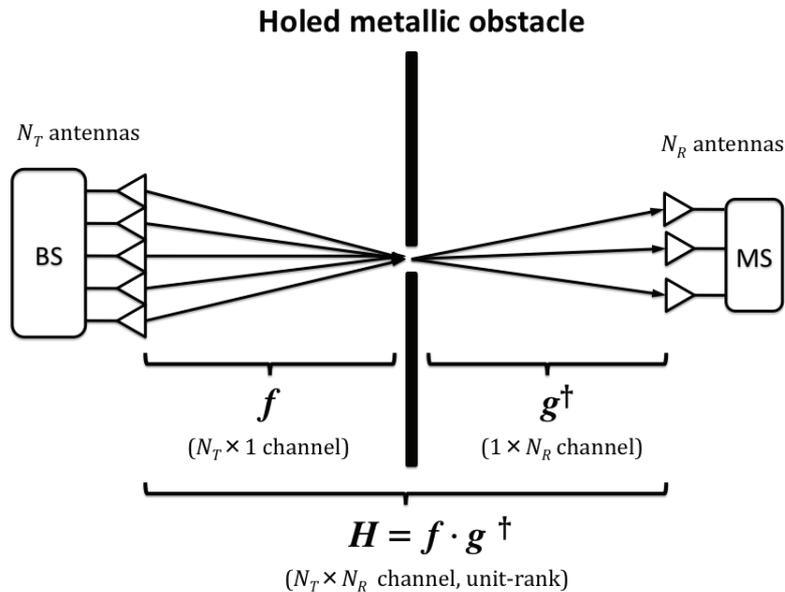


Figure 3.1: A MIMO (spatial) keyhole scenario

We make the following findings:

1. There exists a “cutoff” relay transmit power P_1^* above which the keyhole effects can be mitigated (i.e. MIMO capacity is $n = \min\{N_T, N_R\}$ times that of a SISO system).
2. Decreasing the relay transmit power below P_1^* makes the keyhole effect unresolvable.

Besides, we provide a closed form expression for the relay transmit power threshold P_1^* , which we find to be function of the source transmit power and the channel matrices brought into play in the considered relaying scenario.

Hence, assuming appropriate power allocation, cooperative diversity is put forward as an efficient way to mitigate MIMO keyhole effects, even when *both* the source-destination and the relay-destination channels are unit-rank. Furthermore, as we tackle the problem from an information-theoretic perspective, the proposed solution is universal, being indifferent to the physical origins behind the keyhole phenomenon.

The remainder is organized as follows. We start by presenting in Section 2 the system model and preliminaries related to our study, then we derive a closed form for the capacity of the MIMO degraded relay channel. From this closed form, we infer in Section 3 necessary and sufficient conditions for keyhole mitigation to be possible under various CSI assumptions. Ultimately, we provide numerical examples in Section 4 and we conclude our work in Section 5.

3.2 System Model and Preliminaries

3.2.1 System Model

A generic MIMO relay channel is presented in Fig. 3.2, where the transmitter, the receiver and the relay are equipped with multiple antennas. For simplicity, we assume same number N of transmit/receive antennas for all three entities, i.e. $N_T = N_R = N$. Source-relay, relay-destination and source-destination channels are denoted by \mathbf{F} , \mathbf{G} , \mathbf{H} , respectively. These are assumed to be frequency-flat, complex-valued and subject to Rayleigh fading. Signals transmitted by the source and the relay are respectively denoted by \mathbf{x} , \mathbf{x}_1 and are assumed to be zero-mean complex random vectors. Signals received by the relay and the destination are respectively denoted by \mathbf{y}_1 , \mathbf{y} . These are corrupted by additive zero mean white Gaussian (ZMWG) noises \mathbf{z}_1 , \mathbf{z} of variances $\sigma_{z_1}^2, \sigma_z^2$, respectively. Thus, the MIMO relay channel can be modeled by the following equations:

$$\begin{cases} \mathbf{y}_1 = \mathbf{F}\mathbf{x} + \mathbf{z}_1 \\ \mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{x}_1 + \mathbf{z} \end{cases} \quad (3.1)$$

Finally, source and relay transmissions are subject to power constraints that can be modeled as follows:

$$\begin{cases} \text{tr}(\mathbf{Q}_{\mathbf{x}\mathbf{x}}) \triangleq \text{tr}(\mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}) \leq P \\ \text{tr}(\mathbf{Q}_{\mathbf{x}_1\mathbf{x}_1}) \triangleq \text{tr}(\mathbb{E}\{\mathbf{x}_1\mathbf{x}_1^\dagger\}) \leq P_1 \end{cases} \quad (3.2)$$

where $\mathbf{Q}_{\mathbf{x}\mathbf{x}}$, $\mathbf{Q}_{\mathbf{x}_1\mathbf{x}_1}$ denote the covariance matrices of signals \mathbf{x} , \mathbf{x}_1 , respectively.

3.2.2 Preliminaries

On The Keyhole Effect

Consider a point-to-point MIMO scenario made up by an N -antenna transmitter and an N -antenna receiver, and assume perfect CSI to be available to both channel endpoints. Then, it is widely acknowledged that the achievable capacity is given by [10, 15]:

$$C(\mu) = \sum_{i=1}^N (\log(\mu\lambda_i))^+ \quad (3.3)$$

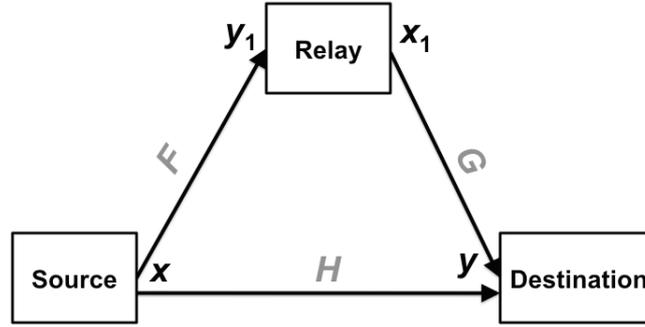


Figure 3.2: The general MIMO relay channel

where μ denotes the *waterfilling level*, a parameter chosen to meet a given power constraint and $(\lambda_i)_{1 \leq i \leq N}$ denote the squared eigenvalues of the channel matrix \mathbf{H} . From (5), one may infer that the higher the channel rank (i.e. the more non-null eigenvalues the channel matrix has), the greater the channel capacity.

In a keyhole environment, however, channel matrix \mathbf{H} is the outer product of two random vectors \mathbf{f} , \mathbf{g} [16–18] (see Fig. 3.1):

$$\mathbf{H} = \mathbf{f} \otimes \mathbf{g} \quad (3.4)$$

$$\triangleq \mathbf{f} \cdot \mathbf{g}^\dagger \quad (3.5)$$

$$= \begin{pmatrix} f_1 g_1^\dagger & \cdots & f_1 g_N^\dagger \\ \vdots & \ddots & \vdots \\ f_M g_1^\dagger & \cdots & f_M g_N^\dagger \end{pmatrix} \quad (3.6)$$

which clearly has unit rank¹. Therefore, under such circumstances, MIMO channel capacity is no better than that of a SISO channel, as only one channel eigenvalue is non-null.

The Degraded Relay Channel

The relay channel of Fig. 3.2 is completely defined by specifying the probability density function $p(\mathbf{y}, \mathbf{y}_1 | \mathbf{x}, \mathbf{x}_1)$ [41]. We start by recalling the definition and capacity theorem relative to the general degraded relay channel.

¹Observe, for instance, that any two rows are linearly dependent, since: $\forall 1 \leq i, j \leq M, \text{row}_j = \frac{f_j}{f_i} \text{row}_i$

Definition 4 ([41]) *The relay channel is said to be degraded if:*

$$p(\mathbf{y}, \mathbf{y}_1 | \mathbf{x}, \mathbf{x}_1) = p(\mathbf{y}_1 | \mathbf{x}, \mathbf{x}_1) p(\mathbf{y} | \mathbf{y}_1, \mathbf{x}_1) \quad (3.7)$$

Theorem 2 ([41]) *The capacity C of the degraded relay channel is given by:*

$$C = \max_{p(\mathbf{x}, \mathbf{x}_1)} \min \{I(\mathbf{x}, \mathbf{x}_1; \mathbf{y}), I(\mathbf{x}; \mathbf{y}_1 | \mathbf{x}_1)\} \quad (3.8)$$

In [41], it is proved that the relaying strategy that achieves the aforementioned capacity region is *Decode-and-Forward* (DF). Therefore, this strategy will be considered here. Though we shall briefly and intuitively explain such strategy, the interested reader is kindly referred to [41] for further details.

3.2.3 On Decode-and-Forward (DF) Relaying

Let us consider the following situation: A wireless BS - MS link as depicted in Fig. ??, with a channel capacity $C = I(\mathbf{x}; \mathbf{y})$. The BS wishes to transmit at a rate R higher than C . If we only rely on this link, then Shannon's coding theorem (stated in Chapter 1) informs us that the MS will not be able to reliably understand the messages transmitted by the BS, as $R > C$. Therefore, we would like to use the relay to increase the capacity to a value $C' > R > C$.

Outline

DF relaying is made of two time-slots, as follows:

1. In the first time slot (Fig. 3.3(a)), BS broadcasts the intended message \mathbf{x} to the MS and the relay, at rate R such that $C' > R > C_1$. In the degraded relay channel, the relay is expected to have better understanding (reception quality) of the transmitted message (only a better-informed relay would be able to help the MS) than the MS. This can be achieved by placing the relay closer to the BS than the MS is. In other words, the BS - relay channel enjoys a channel capacity $C_R = I(\mathbf{x}; \mathbf{y}_1)$ greater than or equal to C' .
2. In the second time slot (Fig. 3.3(b)), the MS would have some uncertainty about the transmitted message, given that $R > C$. Contrarily, the relay is assumed to have understood the message \mathbf{x} , as the transmit rate R falls below the capacity C_R of the channel between the BS and the relay. Thus, the relay sends *some* signaling information \mathbf{x}_1 to help the MS resolve

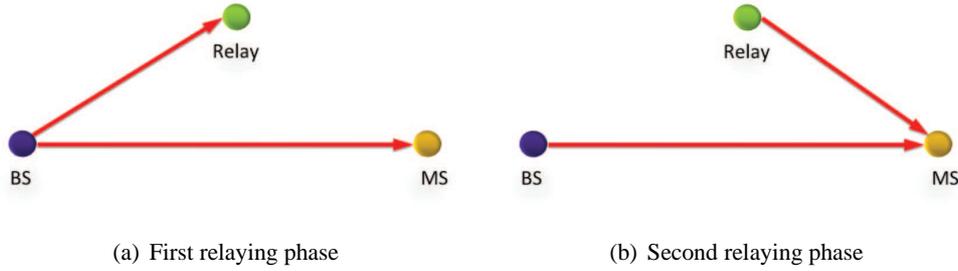


Figure 3.3: Decode-and-forward (DF) relaying.

its uncertainty about the message \mathbf{x} . In practice, such signaling could be, for instance, extra parity bits in case the BS's transmission is LDPC-encoded. During the same time slot, it is suggested that the BS may send a subsequent message for further spectral efficiency.

A Hint On The Capacity Derivation

Next, we hint on how the capacity bound for the degraded relay channel was derived. We shall only explain the converse, while we refer the reader to [41] for the achievability proof:

- We have 2 transmitters (BS, Relay) and 1 receiver (MS). Therefore, by definition, the capacity C is lower than (or equal to) the mutual information of the channel between BS and the relay on one side and the MS on the other:

$$C \leq I(\mathbf{x}, \mathbf{x}_1; \mathbf{y}). \quad (3.9)$$

- On the other hand, according to the relaying scheme that we mentioned earlier, in order for the relay to help the BS, the relay must understand the message in the first time slot. Therefore, the BS should transmit at a rate lower than the channel capacity of the BS Relay channel, i.e.:

$$C \leq I(\mathbf{x}; \mathbf{y} | \mathbf{x}_1). \quad (3.10)$$

- Combining the two results, we get:

$$C \leq \min \{I(\mathbf{x}; \mathbf{y} | \mathbf{x}_1)\}. \quad (3.11)$$

- In the literature [41], it can be shown that the afore-mentioned relaying scheme achieves the upper-bound on capacity for the degraded case, i.e.:

$$C = \min \{I(\mathbf{x}; \mathbf{y}|\mathbf{x}_1)\}. \quad (3.12)$$

3.3 On Keyhole Effect Mitigation

3.3.1 Problem Statement

As explained in the introduction, we focus on the more challenging case where *both* the source-destination channel, \mathbf{H} , and the relay-destination channel, \mathbf{G} , incur keyhole effect (i.e. have unit rank). From previous capacity results, we would like to infer necessary and/or sufficient conditions under which the keyhole effect can be mitigated (i.e. capacity scales linearly with $n = N$), assuming the source-relay channel, \mathbf{F} , is rich-scattered (i.e. full-rank).

3.3.2 The Fixed Channel Case

Capacity Results For The MIMO Degraded Relay Channel

Proposition 1 *If the afore-described MIMO relay channel is degraded, i.e.:*

$$p(\mathbf{y}, \mathbf{y}_1|\mathbf{x}, \mathbf{x}_1) = p(\mathbf{y}_1|\mathbf{x}, \mathbf{x}_1)p(\mathbf{y}|\mathbf{y}_1, \mathbf{x}_1), \quad (3.13)$$

channel capacity is as follows:

$$C^d = \min \{C_1, C_2\} \quad (3.14)$$

where:

$$\begin{cases} C_1 = \log \left| \mathbf{I}_N + \frac{1}{\sigma_z^2} \mathbf{G} \mathbf{Q}_{\mathbf{x}_1 \mathbf{x}_1} \mathbf{G}^\dagger + \frac{1}{\sigma_z^2} \mathbf{H} \mathbf{Q}_{\mathbf{x} \mathbf{x}} \mathbf{H}^\dagger \right| \\ C_2 = \log \left| \mathbf{I}_N + \frac{1}{\sigma_{z_1}^2} \mathbf{F} \mathbf{Q}_{\mathbf{x} \mathbf{x}} \mathbf{F}^\dagger \right| \end{cases} \quad (3.15)$$

Proof See Appendix A

Mitigating The Keyhole Effect Under Uniform Power Allocation Constraint

Theorem 3 Consider a degraded keyhole channel where the source relay-channel \mathbf{F} is full-rank and both the source-destination \mathbf{H} and the relay-destination \mathbf{G} channels are unit-rank. Then, there exists a relay transmit power threshold P_1^* such that:

- For any relay transmit power $P_1 \geq P_1^*$, keyhole effect can be mitigated, i.e. capacity scales linearly with N ;
- For any relay transmit power $P_1 < P_1^*$, keyhole effect cannot be mitigated.

Moreover, when the source equally allocates its transmit power P among its transmit antennas, P_1^* is given by:

$$P_1^* = \frac{\sigma_z^2}{\gamma} \left(\prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{z_1}^2} \phi_i \right) - \frac{P}{\sigma_z^2} \eta - 1 \right) \quad (3.16)$$

where γ, η resp. denote the only non-null squared eigenvalues of \mathbf{G}, \mathbf{H} and, $\forall 1 \leq i \leq N$, ϕ_i denotes the i th squared non-null eigenvalue of \mathbf{F} .

Proof From Theorem 1, we know that the capacity of the general degraded relay channel is given by:

$$C^d = \min \{C_1, C_2\} \quad (3.17)$$

From (3.15), C_1 and C_2 are given by:

$$\begin{cases} C_1 = \log \left| \mathbf{I}_N + \frac{1}{\sigma_z^2} \mathbf{G} \mathbf{Q}_{x_1 x_1} \mathbf{G}^\dagger + \frac{1}{\sigma_z^2} \mathbf{H} \mathbf{Q}_{xx} \mathbf{H}^\dagger \right| \\ C_2 = \log \left| \mathbf{I}_N + \frac{1}{\sigma_{z_1}^2} \mathbf{F} \mathbf{Q}_{xx} \mathbf{F}^\dagger \right| \end{cases} \quad (3.18)$$

Of all three channel matrices, *only* \mathbf{F} is full-rank. Therefore, only C_2 scale linearly with N . It follows that channel capacity C^d scales linearly with the number of antennas *iff* $\min \{C_1, C_2\} = C_2$, which is equivalent to:

$$\left| \mathbf{I}_N + \frac{1}{\sigma_z^2} \mathbf{G} \mathbf{Q}_{x_1 x_1} \mathbf{G}^\dagger + \frac{1}{\sigma_z^2} \mathbf{H} \mathbf{Q}_{xx} \mathbf{H}^\dagger \right| \geq \left| \mathbf{I}_N + \frac{1}{\sigma_{z_1}^2} \mathbf{F} \mathbf{Q}_{xx} \mathbf{F}^\dagger \right| \quad (3.19)$$

As we assumed equal power allocation, the right expression in (3.19) reads [10, 15]:

$$\left| \mathbf{I}_N + \frac{1}{\sigma_{\mathbf{z}_1}^2} \mathbf{F} \mathbf{Q}_{\mathbf{x}\mathbf{x}} \mathbf{F}^\dagger \right| = \prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{\mathbf{z}_1}^2} \phi_i \right) \quad (3.20)$$

where ϕ_i denotes the i th non-null eigenvalue of $\mathbf{F} \mathbf{F}^\dagger$, which is the i th non-null squared eigenvalue of \mathbf{F} .

On the other hand, because channels \mathbf{H} and \mathbf{G} are unit-rank, the left expression in (3.19) simplifies to:

$$\left| \mathbf{I}_N + \frac{1}{\sigma_{\mathbf{z}}^2} \mathbf{G} \mathbf{Q}_{\mathbf{x}_1 \mathbf{x}_1} \mathbf{G}^\dagger + \frac{1}{\sigma_{\mathbf{z}}^2} \mathbf{H} \mathbf{Q}_{\mathbf{x}\mathbf{x}} \mathbf{H}^\dagger \right| = 1 + \frac{P_1}{\sigma_{\mathbf{z}}^2} \gamma + \frac{P}{\sigma_{\mathbf{z}}^2} \eta \quad (3.21)$$

where γ, η are the only non-null squared eigenvalues of \mathbf{G}, \mathbf{H} , respectively. Therefore:

$$C_1 \geq C_2 \Leftrightarrow 1 + \frac{P_1}{\sigma_{\mathbf{z}}^2} \gamma + \frac{P}{\sigma_{\mathbf{z}}^2} \eta \geq \prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{\mathbf{z}_1}^2} \phi_i \right) \quad (3.22)$$

$$\Leftrightarrow \boxed{P_1 \geq \frac{\sigma_{\mathbf{z}}^2}{\gamma} \left(\prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{\mathbf{z}_1}^2} \phi_i \right) - \frac{P}{\sigma_{\mathbf{z}}^2} \eta - 1 \right)} \quad (3.23)$$

Q.E.D.

Mitigating The Keyhole Effect Under Waterfilling

Waterfilling is achieved by allocating power to each eigenmode depending on the value of each non-null channel eigenvalue, with the evident assumption that at least two eigenvalues be non null². In the setting considered in this work, only the channel between the BS and the relay, \mathbf{F} , has more than one non-null eigenvalue, therefore BS can apply waterfilling only to that channel. Denote by μ the waterfilling levels relative to the power allocation to the eigenvalues ϕ_1, \dots, ϕ_N . Then, the Shannon capacity C_2 of the channel \mathbf{F} between the BS and the relay reads [15]:

$$C_2 = \sum_{i=1}^N (\log(\mu \phi_i))^+ \quad (3.24)$$

subject to the total power constraint [15]:

$$\sum_{i=1}^N \left(\mu - \frac{1}{\phi_i} \right)^+ = P \quad (3.25)$$

²if only one eigenvalue is non-null, then necessarily all the power will be allocated to that non-null eigenmode.

Recall that, owing to (15), the keyhole effect can be mitigated *iff* $C_1 \geq C_2$. This implies:

$$1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \geq \sum_{i=1}^N (\log(\mu \lambda_i))^+ \quad (3.26)$$

To conclude:

Theorem 4 *If the BS allocates its power by means of waterfilling, then, under the assumptions and notations of Theorem 2, the keyhole effect can be mitigated iff the relay transmit power is above a power threshold P_1^* given by:*

$$P_1^* = \frac{\sigma_z^2}{\gamma} \left(\sum_{i=1}^N (\log(\mu \lambda_i))^+ - \frac{P}{\sigma_z^2} \eta - 1 \right) \quad (3.27)$$

where μ denotes the MIMO waterfilling level.

3.3.3 The Stochastic Channel Case

In 3.3.2, we provided necessary and sufficient conditions and a downlink transmission protocol that ensure that the keyhole effect is mitigated *at any time*, under the degradedness assumption. In particular, we found out that the relay transmit power, P_1 needs to be adjusted w.r.t. the BS transmit power P every time any of the channels change to a new realization, such that P_1 is always above a certain threshold P_1^* that we found out to be function of the channels' eigenvalues and the SNR parameters of each channel.

This proposal, though it mitigates the keyhole effect, may however have some limits. In some situations, this procedure may be regarded as time-consuming or as computationally-expensive, as it requires the power threshold to be recomputed every time any of the channel matrices is changed. Besides, the accuracy of the computed power threshold P_1^* is sensitive to the accuracy of the computation of the channels' eigenvalues, thereby requiring a sufficiently accurate channel estimation. Though this may be achieved through longer training sequences in slowly-varying MIMO channels, it may be unfeasible in other cases, for instance in transmissions based on frequency hopping or in high-mobility scenarios where the fading block length is too short to allow sufficiently long training sequences³. Even more, CSI may be unavailable to the transmitters (BS

³In such situations, the channel realizations are constant only for a short period of time (i.e. $T \ll \llcorner$), thus if too long training sequences are used (i.e. if $T_p \gg \gg$), not enough time is left for data transmission (i.e. $T_d = T - T_p \ll \ll$)

and/or relay), e.g. in open-loop MIMO settings. Finally, cellular systems usually obey average transmit power constraints rather than instantaneous ones.

Given all the above, it would be desirable to determine a similar necessary and/or sufficient condition on the *average* relay transmit power, that ensures the keyhole effect mitigation, yet without requiring the *a priori* knowledge of any channel state information. This raises interest in the notion of *ergodic capacity*: rather than determining a relay *instantaneous* power threshold that ensures that the *instantaneous capacity* at a given time t scales linearly with the number of antennas, we will attempt to determine a threshold on the relay *average* transmit power that ensures that the *ergodic capacity* (i.e. capacity averaged on the channel realizations [10]) scales linearly with the number of antennas.

By definition, we have:

$$C_{1,avg} = \mathbb{E}_{\mathbf{H},\mathbf{G}} \left\{ \log \left(1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right) \right\} \quad (3.28)$$

$$C_{2,avg} = \mathbb{E}_{\mathbf{F}} \left\{ \log \left(\prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{\mathbf{z}_1}^2} \phi_i \right) \right) \right\} \quad (3.29)$$

Following the reasoning in 3.3.2, the keyhole effect can be mitigated if and only if $C_{1,avg} \geq C_{2,avg}$, or equivalently:

$$\mathbb{E}_{\mathbf{H},\mathbf{G}} \left\{ \log \left(1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right) \right\} \geq \mathbb{E}_{\mathbf{F}} \left\{ \log \left(\prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{\mathbf{z}_1}^2} \phi_i \right) \right) \right\} \quad (3.30)$$

As \log is a concave function, applying Jensen's inequality [42] yields the following *sufficient* condition:

$$\log \left(\mathbb{E}_{\mathbf{H},\mathbf{G}} \left\{ 1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right\} \right) \geq \mathbb{E}_{\mathbf{F}} \left\{ \log \left(\prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{\mathbf{z}_1}^2} \phi_i \right) \right) \right\} \quad (3.31)$$

The expectation averaged on channels \mathbf{H} , \mathbf{G} is given by:

$$\mathbb{E}_{\mathbf{H},\mathbf{G}} \left\{ 1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right\} \triangleq \int_0^\infty \int_0^\infty \left(1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right) p_{\gamma,\eta}(\gamma, \eta) d\gamma d\eta \quad (3.32)$$

where $p_{\gamma,\eta}(\gamma, \eta)$ denotes the joint distribution (probability density function, pdf) of (γ, η) . In principle, channels \mathbf{G} and \mathbf{H} are decorrelated since they involve two different transmitters (the BS and the relay, resp.) that are located apart from each other by a distance far larger than half the wavelength (which is of the order of the cm for carrier frequencies of the order of the GHz).

Hence, the joint pdf satisfies:

$$p_{\gamma,\eta}(\gamma, \eta) = p_{\gamma}(\gamma)p_{\eta}(\eta) \quad (3.33)$$

thereby simplifying (3.32) into the following:

$$\mathbb{E}_{\mathbf{H},\mathbf{G}} \left\{ 1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right\} = \int_0^\infty \left(\int_0^\infty \left(1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right) p_{\gamma}(\gamma) d\gamma \right) p_{\eta}(\eta) d\eta \quad (3.34)$$

It is well-known that the matrices $\mathbf{F}\mathbf{F}^\dagger$, $\mathbf{G}\mathbf{G}^\dagger$ and $\mathbf{H}\mathbf{H}^\dagger$ follow a Wishart distribution [15], which is basically an extension of the Chi-Squared distribution to random matrices. The joint pdf of the unordered eigenvalues $\lambda_1, \dots, \lambda_N$ of a Wishart-distributed random matrix is given by [15]:

$$p_{\lambda}(\lambda_1, \dots, \lambda_N) = \frac{1}{N!K} e^{-\sum_i \lambda_i} \prod_{i < j} (\lambda_i - \lambda_j)^2 \quad (3.35)$$

where K is a normalizing factor.

As γ, η are the only non-null eigenvalues of their respective matrices, it follows that:

$$p_{\gamma}(\gamma) = \frac{1}{N!K_{\gamma}} \gamma^2 e^{-\gamma} \quad (3.36)$$

$$p_{\eta}(\eta) = \frac{1}{N!K_{\eta}} \eta^2 e^{-\eta} \quad (3.37)$$

$p_{\gamma}(\gamma), p_{\eta}(\eta)$ are pdfs, thus they must sum (integrate) to 1. Since $\int_0^\infty t^2 e^{-t} dt \triangleq \Gamma(2) = 1$ (where Γ denotes the Gamma function), it follows that $\frac{1}{N!K_{\gamma}} = \frac{1}{N!K_{\eta}} = 1$.

Bearing in mind that $\int_0^\infty t^n e^{-t} dt = \Gamma(n) = (n-1)!$, we get:

$$\mathbb{E}_{\mathbf{H},\mathbf{G}} \left\{ 1 + \frac{P_1}{\sigma_z^2} \gamma + \frac{P}{\sigma_z^2} \eta \right\} = 1 + \frac{2P_1}{\sigma_z^2} + \frac{2P}{\sigma_z^2} \quad (3.38)$$

On the other hand, $\mathbb{E}_{\mathbf{F}} \left\{ \log \left(\prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{z_1}^2} \phi_i \right) \right) \right\}$ is the ergodic capacity of a traditional point-to-point MIMO channel, whose expression has already been developed in literature. This was found to be [15]:

$$\mathbb{E}_{\mathbf{F}} \left\{ \log \left(\prod_{i=1}^N \left(1 + \frac{P}{N\sigma_{z_1}^2} \phi_i \right) \right) \right\} = \int_0^\infty \log \left(1 + \frac{P}{N\sigma_{z_1}^2} \phi \right) \sum_{k=0}^{N-1} L_k(\phi)^2 e^{-\phi} d\phi \quad (3.39)$$

where L_k is the Laguerre polynomial of order k . Thus, owing to (3.38) and (3.39), the sufficient condition for keyhole mitigation becomes:

$$\log \left(1 + \frac{2P_1}{\sigma_z^2} + \frac{2P}{\sigma_z^2} \right) \geq \int_0^\infty \log \left(1 + \frac{P}{N\sigma_{z_1}^2} \phi \right) \sum_{k=0}^{N-1} L_k(\phi)^2 e^{-\phi} d\phi, \quad (3.40)$$

from which the following theorem can be easily inferred:

Theorem 5 *From an ergodic capacity perspective, under the assumptions and notations of Theorem 2, a sufficient condition for the degraded MIMO relay channel to mitigate the keyhole effect is that the relay average transmit power $P_{1,avg}$ be above a threshold $P_{1,avg}^*$ given by:*

$$P_{1,avg}^* = \frac{\sigma_z^2}{2} \left[\exp \left(\int_0^\infty \log \left(1 + \frac{P_{avg}}{N\sigma_{z_1}^2} \phi \right) \sum_{k=0}^{N-1} L_k(\phi)^2 e^{-\phi} d\phi \right) - \frac{2P_{avg}}{\sigma_z^2} - 1 \right] \quad (3.41)$$

where P_{avg} is the average transmit power of the BS and L_k denotes the Laguerre polynomial of order k .

3.4 Numerical Examples

In this Section, we numerically evaluate the analytical results of Section 3 relative to keyhole effects mitigation, when DF is the relaying strategy implemented by the relay node. Reported simulation results have been averaged over 10,000 channel realizations. Our aim is to confirm that, *owing to our transmit power requirement, capacity scales with the number of antennas, N , when \mathbf{F} is full rank, even if \mathbf{H}, \mathbf{G} are unit-rank*. For this sake, unit-rank channels \mathbf{H}, \mathbf{G} were generated as the normalized outer product of random complex vectors uniformly distributed over the unit sphere, while channel \mathbf{F} was generated with i.i.d. entries such that it has full rank.

Fig. 3.4 portrays the capacity scaling with the number of antennas, for different SNR values. First, we observe that C_1 , -the capacity when the proposed power constraint is not met-, does not scale linearly with the number of antennas. This is owing to the fact that both channels \mathbf{H}, \mathbf{G} incur keyhole effect: No matter the number of antennas, \mathbf{H}, \mathbf{G} have only one non-null eigenvalue each, that affect the scaling of C_1 . Contrarily, C_2 , -the capacity when the proposed power constraint is met-, scales linearly with N , thereby exceeding C_1 when $N > 2$. Hence, under such settings, we clearly observe that satisfying the proposed transmit power requirement ensures that the capacity

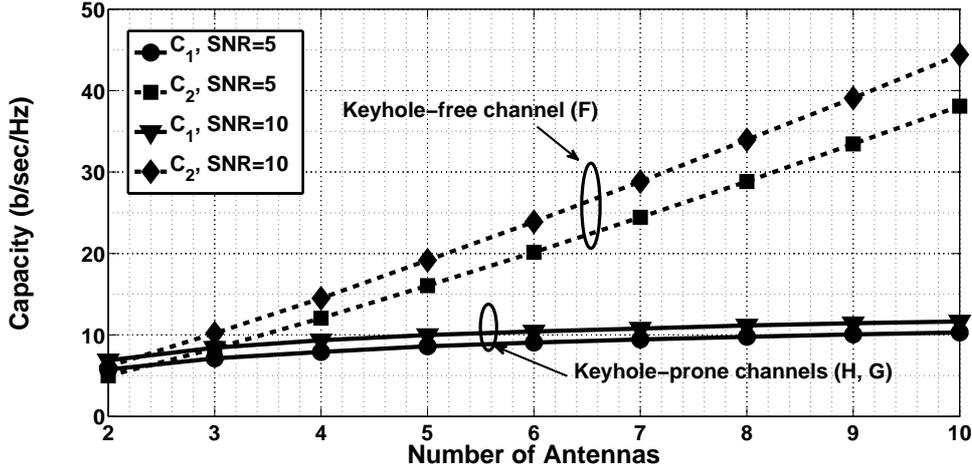


Figure 3.4: Comparison of the scaling of C_1 , C_2 with N , the number of transmit antennas, for SNR = 5, 10 dB. In this example, \mathbf{H} , \mathbf{G} incur keyhole effects (i.e. have unit rank), while \mathbf{F} is full-rank.

scales linearly with the number of antennas, despite having both channels \mathbf{H} , \mathbf{G} incur keyhole effects. Thus, the relevance of our claim in Theorem 2 is validated.

Furthermore, Fig. 3.5 portrays the capacity scaling with SNR, for different numbers of transmit antennas. Out of concern for fairness, same SNR values were set for the different channels involved. Again, we observe that, under the aforementioned assumptions, C_2 scales with SNR better than C_1 . Indeed, in the expression of C_1 (3.21), SNR is multiplied only by 2 eigenvalues, irrespective of N , whereas in the expression of C_2 (3.20), SNR is multiplied by N non-null eigenvalues. Hence, C_2 would scale with SNR better than C_1 whenever $N > 2$, again another impediment owing to the keyhole effects underwent by channels \mathbf{H} , \mathbf{G} .

As explained earlier throughout the previous sections of the chapter, the capacity gains are owing to the relaying strategy as well as the suggested power allocation. Indeed, had we used a different relaying strategy, say amplify-and-forward (AF), the relay's *data* transmission would also not help deal with the keyhole effect. Contrarily, the DF strategy is such that the relay is sending signaling information, hence the modified capacity formula. Further, the use of the proposed power allocation ensures that the total system capacity is only reliant upon the capacity of the BS - Relay portion by ensuring that this portion always has a smaller capacity than the remainder of the channel.

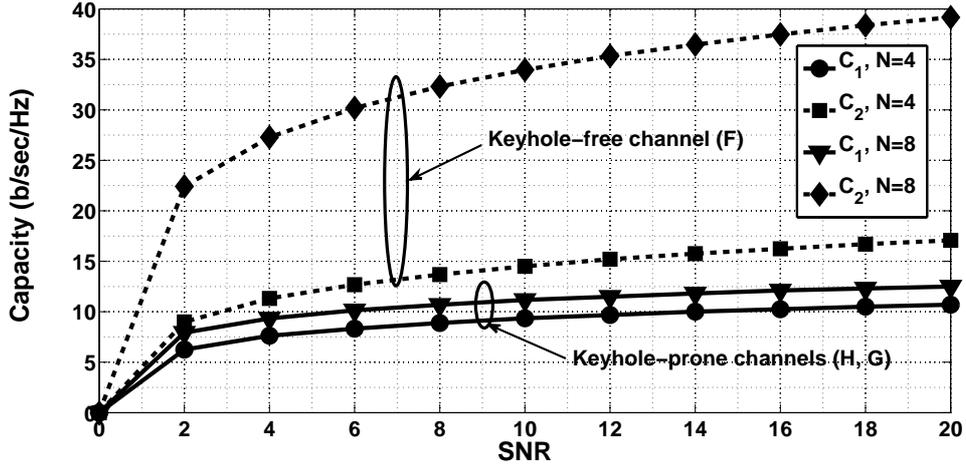


Figure 3.5: Comparison of the scaling of C_1 , C_2 , when $N = 4, 8$. In this example, \mathbf{H} , \mathbf{G} incur keyhole effects (i.e. have unit rank), while \mathbf{F} is full-rank.

3.5 Conclusion

In this chapter, we considered the problem of ensuring MIMO capacity linear scaling with the number of transmit antennas when the destination suffers from a keyhole effect. We demonstrated that cooperative diversity (relay deployment) can mitigate such phenomenon. Precisely, under degradedness assumption, if the source-relay channel is full rank, we proved that there exists a “cutoff” relay transmit power above which keyhole effects can be mitigated *irrespective of the ranks of the source-destination and relay-destination channels*. We devised the closed form of this power threshold as function the source transmit power and the channels brought into play in the relaying scenario. Furthermore, we also provided a sufficient condition for keyhole effect mitigation in wireless MIMO systems in the absence of CSI at the BS and the relay, by investigating the ergodic capacity of the MIMO degraded relay channel. Numerical examples confirmed the relevance of our claims.

Chapter 4

On Joint User Scheduling and CSI Feedback in Multi-user Closed-Loop MIMO

4.1 Introduction

Multiple-Input-Multiple-Output (MIMO) systems can achieve high-throughput wireless communications [7, 15], provided that Channel State Information (CSI) be available to both channel endpoints. Precisely, the absence of CSI at the receiver makes the system capacity not scale linearly with the number of antennas [10]. Likewise, the absence of CSI at the transmitter limits the achievable capacity [10]. This highlights the significance of CSI availability to the performance of MIMO systems.

In practical systems, however, CSI may be available to the receiver (e.g. by estimation from received pilot symbols) but not to the transmitter, in which cases a CSI feedback is often required from the receiver to the transmitter.

Several works have focused on MIMO feedback issues in the single user case. These include CSI quantization and codebook design [13, 43], feedback delay and error mitigation [44], and capacity analysis under different channel assumptions (for further details, see [45] and references therein). Most of these issues seem to be well covered in literature.

Multi-user MIMO, on the other hand, seems to have been less investigated, and more issues remain challenging, particularly in the non-trivial case where the number of transmit antennas is smaller than the number of users, which will be the focus of this work.

The situation is particularly incurred when Zero-Forcing Beamforming (ZFBF) [46] is used at the transmitter. This is a linear precoding technique that, albeit sub-optimal, achieves DPC-like (*Dirty Paper Coding* [19]) close-to-capacity performance in multi-user MIMO systems [10, 46], provided the receivers have mutually-orthogonal channels. Given a large number of candidate receivers with independent fades, it is likely to find a subset of users having channels in the null space of each other's. Hence, from this perspective, scenarios ideal for ZFBF are those where the number of users N is much larger than the number of available transmit antennas.

Nevertheless, the existence of a large set of candidate receivers yields numerous impediments. Suppose, for instance, an average population of 20 active users served by a 4-transmit-antenna Base Station (BS). Evidently, no more than 4 users can be served at a given time. Out of concern for fairness, the BS may schedule the users' transmissions in a Round-Robin fashion, periodically serving each group of 4 users every 5 time slots. Yet, it has been shown in [47] that a judicious design of the scheduling policy may provide significant throughput gains while preserving some fairness among users. Specifically, spectral efficiency is increased if, based on the received feedbacks, the BS picks 4 mutually-orthogonal users with largest SNRs (such as *Zero-Forcing Beamforming with Semi-orthogonal User Selection* (ZFBF-SUS) [20]). We shall refer to such scheduling approaches as sum-capacity maximizing scheduling schemes. Though such schemes achieve high average downlink throughputs, in our view, they are burdened with, at least, the following three shortcomings.

First, multi-user feedback in this generic solution is very resource hungry. Indeed, in order to select best users, the BS has to receive CSI from all users (so as to compare them to each other). Simply put, why would 20 users feedback their CSI when only 4, among them, will be scheduled for next transmission? Some attempts have been made to remedy this issue. For instance, in [48], it was suggested that only users with SINR larger than a certain threshold feedback their CSI. However, though it may reduce the number of feedbacks, such solution does not necessarily guarantee that the users who feedback their CSI are indeed optimal (in the ZFBF sense) as they may not necessarily have mutually-orthogonal channels.

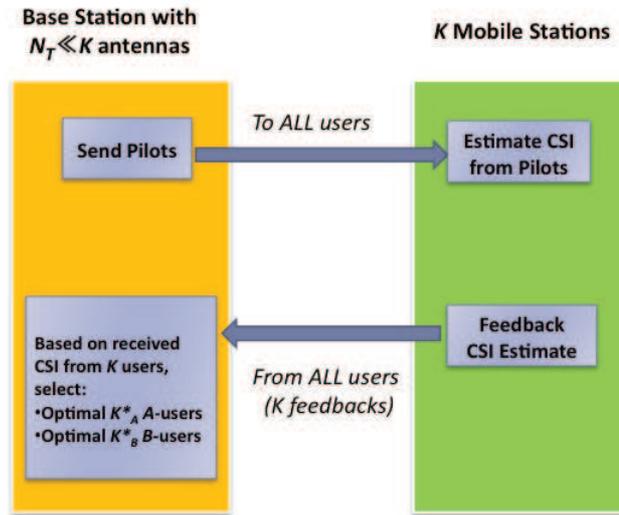
Second, if all users feedback their channels, then exhaustively searching for a subset of users with orthogonal channels may become computationally prohibitive at the BS's side. Though works such as [20] suggested some suboptimal search heuristics, it still remains preferable if the number of feedbacks is kept to a minimum, so as to reduce this computational burden.

Finally, as queuing delay is disregarded when scheduling packets (only channel condition is considered), our intuition is that sum-capacity maximizing scheduling approaches fail to satisfy the requirements of delay-sensitive applications (Video on Demand (VoD), circuit-emulated voice calls and networked gaming, to name a few) for which some delay constraints should be met. In this chapter, we attempt to provide a solution to the aforementioned problems. We shall see that these problems are not necessarily independent, and that they may be jointly solved under a unified cross-layer framework.

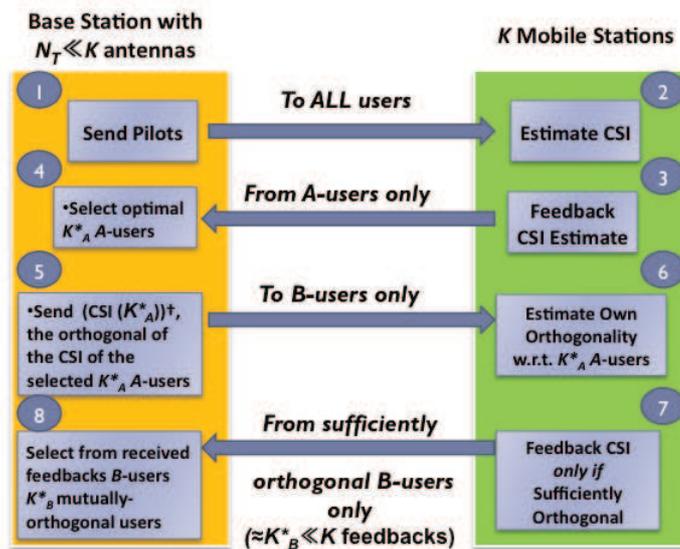
Suppose, for instance, two classes of service A and B , where class A 's traffic has higher priority and *Quality of Service* (QoS) requirements than class B 's traffic. If the BS feeds back information about A -users' CSI to B -users, then the latter can measure how orthogonal their channels are to those of A -users. Hence, a B -user feeds back its CSI only if its channel is *sufficiently* orthogonal to those of the scheduled A -users. This way, the number of feedbacks may be made as small as desired (by defining *how sufficient* should be), which in turn reduces the resources required for feedback, renders exhaustive search computationally feasible at the BS and allows for feedback only from users who are likely to be scheduled for transmission. Ultimately, the BS simply picks, from those feedbacks, the B -users who not only are orthogonal to A -users, but are also orthogonal to each other. Again, this search is made possible owing to the reduction of the number of received feedbacks. Fig. 4.1 illustrate the main idea behind the proposed approach, compared with the conventional approach.

The main contributions of this work are threefold:

- A Multi-user MIMO feedback scheme that, w.r.t. conventional schemes, is service-differentiated (i.e. supports prioritized/multiclass traffic) and more power-and-bandwidth efficient (In conventional approaches, all users need to feed back their CSI. In the proposed scheme, only ZFBF-optimal users (i.e. users that are likely to be scheduled for next transmission) feed back their CSI).
- For delay-sensitive users, a performance analysis that quantitatively assesses, in Rayleigh fading channels, the trade-off between the transmit power and guaranteed transmission rates on one hand, and the required number of streams/antennas and the incurred delays on the other.
- A delay-aware scheduling scheme to support the QoS requirements of such-constrained users.



(a) Conventional Approach



(b) Proposed Approach

Figure 4.1: Comparison of conventional and proposed channel feedback approaches in closed-loop multi-user MIMO. The proposed scheme requires twice as many time-slots as the conventional, yet tremendously reduces the number of feedbacks and the computational burden of the user selection process at the base station.

4.2 System Model

We consider a multi-user MIMO system, made up by a BS and K active Mobile Stations (MS), each equipped with multiple receive antennas. We denote by N_T the number of transmit antennas at the BS and, for simplicity, we assume that all MSs are equipped with equal number of antennas, N_R . Of particular interest is the case where $N_T < K$, the scheduling problem becoming challenging (as only a maximum of N_T MSs can only be scheduled for transmission at a given time).

4.2.1 Traffic Model

We assume 2 Classes of Service (CoS):

- A high priority CoS, which we denote by A and refer to as the class of *Constant Bit-Rate* (CBR) *users*. Owing to their delay constraints, these users are given priority over the other class's users in channel feedback, scheduling and stream (eigenmode) preferences.
- A low priority CoS, which we denote by B and refer to as the class of *Best Effort* (BE) *users*. These users have no Quality of Service (QoS) guarantees. Therefore, their traffic is scheduled depending on their SINR and on their channels' orthogonality to the CBR users', i.e. how much interference they would cause to CBR users, should they be scheduled for transmission at the next time slot.

The K active users consist of K_A A -users and $K_B = K - K_A$ B -users, among which resp. K_A^* and K_B^* users are selected at each transmission time slot to be served (hence $K_A^* + K_B^* \leq N_T$). Traffic queues Q_A and Q_B relative to each CoS are characterized by independent packet arrivals and exponentially-distributed service times, therefore modeled as $M/M/K_A^*$ and $M/M/K_B^*$ queues with arrival rates λ_A, λ_B and service times μ_A, μ_B , respectively [49].

4.2.2 Channel Model

Channels between the BS and each CBR user a_k are denoted as $\mathbf{H}_{a_k} \in \mathbb{C}^{N_R \times N_T}$, $1 \leq k \leq K_A$, and between the BS and each BE user b_j are denoted as $\mathbf{H}_{b_j} \in \mathbb{C}^{N_R \times N_T}$, $1 \leq j \leq K_B$. They are assumed to be frequency-flat and slowly time-varying, obeying the conventional block-fading law of coherence time¹ T and with entries (channel gains) following Dent's model [39]. We model

¹i.e. T is the time interval during which the channel remains constant, before changing to a new independent realization.

noises \mathbf{z}_k (resp. \mathbf{z}_j) relative to channels $\mathbf{H}_{a_k}, 1 \leq k \leq K_A$ (resp. $\mathbf{H}_{b_j}, 1 \leq j \leq K_B$) as Zero Mean White Gaussian (ZMWG) with variances $\sigma_{a_k}^2$ (resp. $\sigma_{b_j}^2$).

The coherence time T is decomposed into two phases:

- A training phase of duration T_p , during which the BS sends pilots to intended receivers who estimate channels from the received training sequence and feedback their respective CSI estimates. Owing to channel identifiability requirement, a complex-valued training sequence $\mathbf{X}_p \in \mathbb{C}^{N_t \times L}$ of pilot symbols is transmitted by BS to the receivers, where L is the training sequence length. Both the received signals \mathbf{Y}_k and noises \mathbf{Z}_k are, therefore, matrices in $\mathbb{C}^{N_R \times L}$.
- A transmission phase of duration $T_d = T - T_p$ during which the BS uses the fed-back CSI to adapt its data transmission signals to the channels of intended scheduled receivers. In this phase, received signals are symbol-wise detected. Therefore, transmitted signals by BS are vectors of symbols, $\mathbf{x} \in \mathbb{C}^{N_R \times 1}$. It follows that the received signals \mathbf{y}_k and noises \mathbf{z}_k are vectors in $\mathbb{C}^{N_R \times 1}$.

4.3 A QoS-aware Channel Estimation and Feedback Scheme

4.3.1 Proposed Scheme

The proposed scheme is as follows:

1. First, BS sends pilots to all users.
2. From received pilots, K_A A -users estimate their inward channels $\{\mathbf{H}_{a_k}, 1 \leq k \leq K_A\}$ and feedback their estimates to BS.
3. Upon reception of CSI feedback from A -users, BS selects relevant K_A^* users w.r.t. the delay guarantees (see the scheduling algorithm proposed later).

4. After that, BS sends to B -users the normalized sum $\mathbf{P} = \sum_{k=1}^{K_A^*} \frac{\mathbf{H}_{a_k}^\dagger}{\|\mathbf{H}_{a_k}^\dagger\|^2}$ of the K_A^* users' channels conjugate transposes, multiplied by the training matrix \mathbf{X}_p . Therefore, each B -user $b_j, 1 \leq j \leq K_B$ receives the following signal matrix:

$$\mathbf{Y}_{b_j} = \mathbf{H}_{b_j} \mathbf{P} \mathbf{X}_p + \mathbf{Z}_{b_j} = \underbrace{\left[\mathbf{H}_{b_j} \sum_{k=1}^{K_A^*} \frac{\mathbf{H}_{a_k}^\dagger}{\|\mathbf{H}_{a_k}^\dagger\|^2} \right]}_{\subset \text{Null}\{\mathbf{H}_{a_1}, \dots, \mathbf{H}_{a_{K_A^*}}\}} \mathbf{X}_p + \mathbf{Z}_{b_j} \quad (4.1)$$

The term $\Delta_j \triangleq \left[\mathbf{H}_{b_j} \sum_{k=1}^{K_A^*} \frac{\mathbf{H}_{a_k}^\dagger}{\|\mathbf{H}_{a_k}^\dagger\|^2} \right]$ represents the projection of user b_j 's channel on the null space of the K_A^* users' channels, i.e. *a measure of the orthogonality of user b_j vis-a-vis the K_A^* users*. Extreme cases are *perfect orthogonality* ($\Delta_j = \mathbf{H}_{b_j}$) and *colinearity* ($\Delta_j = \mathbf{0}$), with gray zones in between.

Therefore, we suggest that user b_j compares the scalar $\|\Delta_j\|_2$ to a given threshold ϵ . Should this norm be larger than the aforementioned threshold, MS b_j may infer that its channel is *sufficiently orthogonal* to K_A^* users' and is therefore allowed to feedback its CSI to the BS. Otherwise, it refrains from feeding back its CSI as it virtually stands no chance of being scheduled for next transmission (being non-ZFBBF-optimal). In conclusion, a B -MS $b_j, 1 \leq j \leq K_B$, feeds back the following CSI:

$$\text{CSI}(b_j) = \begin{cases} \mathbf{H}_{b_j}, & \text{if } \|\Delta_j\| \geq \epsilon \\ \emptyset, & \text{otherwise} \end{cases} \quad (4.2)$$

Observe that threshold ϵ may be arbitrarily chosen so as to gage how fewer feedbacks BS should expect on the average, hence providing the network operator much flexibility about system design and capacity planning. Later in III-B, we will provide theoretical insights on the minimal value of ϵ that would ensure a targeted number κ of B -users feedback their CSI to BS, with an arbitrary probability p_0 .

5. Ultimately, If more than K_B^* feedbacks are received, then BS is free to choose among them the K_B^* B -users that maximize the sum-capacity. This selection may now be effortlessly performed through exhaustive search, as the search set has been significantly reduced by constraining the feedbacks as we suggested.

Thus, owing to the proposed training-based channel estimation and feedback scheme, a B -MS can tell if its channel is (sufficiently) orthogonal to the A -MSs' without even knowing their individual channels (i.e. without cooperation among nodes).

4.3.2 On The Optimal Feedback Threshold ϵ Under Gaussian Approximation

Let ν be the number of feedbacks received from B -users. In the proposed scheme, parameter ϵ is the threshold that defines whether or not a B -receiver can feed back its CSI to the BS. If ϵ is too small, too many feedbacks will be received by the BS, i.e. $\nu \gg K_B^*$. Contrarily, if ϵ is too large, then not enough feedbacks will be received by the BS, i.e. $\nu < K_B^*$. Therefore, it is interesting to determine the *largest* value of ϵ that would ensure a targeted number κ of B -users feedback their CSI to BS, with an arbitrarily high probability p_0 . By definition, for $1 \leq j \leq K_B$, we have:

$$\Delta_j \triangleq \left[\mathbf{H}_{b_j} \sum_{k=1}^{K_A^*} \frac{\mathbf{H}_{a_k}^\dagger}{\|\mathbf{H}_{a_k}^\dagger\|^2} \right]. \quad (4.3)$$

The entries of the channel matrices are usually assumed to be i.i.d. zero-mean, unit-variance, Gaussian distributed [10, 15]. Thus, $\forall 1 \leq l, m \leq N$, the probability density function (pdf) of the (l, m) th element $\delta_{lm}^{(j)}$ of Δ is given by:

$$p(\delta_{lm}^{(j)}) = \frac{1}{\pi} K_0 \left(\left| \delta_{lm}^{(j)} \right| \right), \quad (4.4)$$

where K_0 is the zeroth modified Bessel function of the second kind:

$$K_0(y) = \int_0^\infty \frac{\cos(yt)}{\sqrt{1+t^2}} dt. \quad (4.5)$$

Unfortunately, the cumulative distribution function (cdf) of a random variable whose pdf is given by the zeroth modified Bessel function of the second kind, K_0 , is not known in closed form, thus there is little hope in determining a single-letter expression of the cdf of $\|\Delta_j\|$, in the general case. To make the problem tractable, we use a Gaussian approximation and model the eigenvalues of $\|\Delta_j\|$ as standard normal variables². Subsequently, $\|\Delta_j\|$, being the Euclidean norm of standard

²We stress that Gaussian approximation is only used to make the problem tractable and may be inaccurate in reality, in general. However, such approximation may hold true if the number of antennas (i.e. the matrix size of Δ_j) is very large, owing to the Central Limit Theorem.

normal variables, follows a Rayleigh distribution:

$$p_{\|\Delta_j\|}(\|\Delta_j\| = y) = y \exp\left(\frac{-y^2}{2}\right). \quad (4.6)$$

Now, in order for a B -user to be entitled to feed back its CSI, $\|\Delta_j\|$ must be higher than a certain value $\epsilon \geq 0$. This event has the following probability (which is also the cdf of $\|\Delta_j\|$ in ϵ):

$$p(\|\Delta_j\|^2 \geq \epsilon) = \int_{\epsilon}^{\infty} y \exp\left(\frac{-y^2}{2}\right) dy = \exp\left(\frac{-\epsilon^2}{2}\right). \quad (4.7)$$

Thus, the probability p_0 that κ feedbacks are received by BS is given by:

$$p_0 = \left(\exp\left(\frac{-\epsilon^2}{2}\right)\right)^{\kappa}. \quad (4.8)$$

This yields:

$$\epsilon = \sqrt{\frac{2}{\kappa} \log\left(\frac{1}{p_0}\right)}. \quad (4.9)$$

To summarize:

Theorem 6 *If the eigenvalues of the matrices $\Delta_j \triangleq \left(\mathbf{H}_{bj} \sum_{k=1}^{K_A^*} \frac{\mathbf{H}_{a_k}^\dagger}{\|\mathbf{H}_{a_k}^\dagger\|^2}\right)$, $1 \leq j \leq K_B$, can be approximated as standard normal variables, then the largest value of the feedback threshold ϵ that ensures that κ B -users feedback their CSI to BS with an arbitrarily high probability p_0 is given by:*

$$\boxed{\epsilon(\kappa, p_0) = \sqrt{\frac{2}{\kappa} \log\left(\frac{1}{p_0}\right)}}. \quad (4.10)$$

Fig. 4.2 illustrates the targeted number κ of B -users who are likely to feedback their CSI with a probability higher than p_0 , versus the optimal orthogonality threshold ϵ , for different values of p_0 . We can readily verify that the required threshold decreases as the number of targeted feedbacks and/or the probability of success increase (as it becomes less likely that B -users meet the feedback requirement).

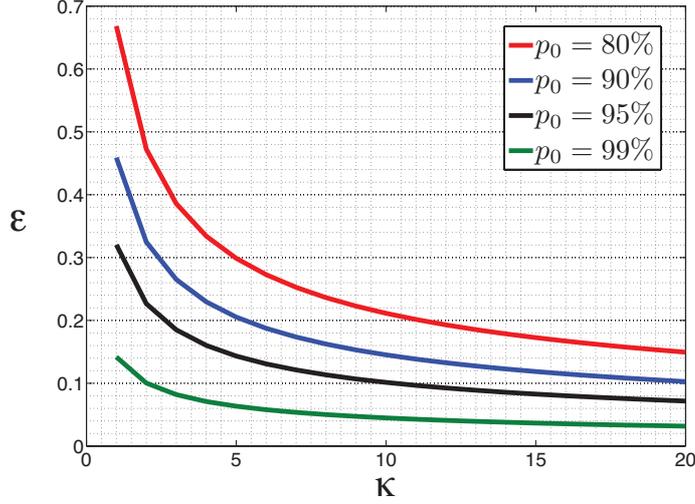


Figure 4.2: Targeted number κ of B -users who are likely to feedback their CSI with a probability higher than p_0 , versus the optimal orthogonality threshold ϵ , for different values of p_0

4.4 On The Required Number of Streams to Meet a Delay Constraint

Under the previous assumptions, we attempt in this paragraph to determine how many streams³ are required to meet a given delay bound of delay-sensitive users. We shall neglect the propagation delay. Therefore, a packet intended for a CBR user a_k may be delayed for two reasons:

1. Because of buffer congestion, if other CBR packets are ahead of it in the queue. In that case, the packet will encounter a *queuing delay* T_Q , as it has to wait its turn.
2. Because of destination unavailability, if the packet can be scheduled for transmission (no packets ahead of it in the queue), but the intended user is in a deep fade (unreachable under the current power constraint). In that case, the packet will encounter an *outage delay* T_{out} , as it has to wait for its destination to become reachable again.

Thus, the guaranteed delay D should be higher than a lower bound D_{min} that is given by:

$$D_{min} = T_Q + T_{out}. \quad (4.11)$$

The mean queuing delay T_Q of an $M/M/K_A^*$ queue is given by [49]:

$$T_Q = \frac{1}{\mu_A K_A^* - \lambda_A} \quad (4.12)$$

³If channels are rich-scattered (full-rank), then this number is also equal to the number of required transmit antennas.

On the other hand, the average outage delay for wireless channels is given by [50]:

$$T_{out} = \frac{P_{out}}{\eta} \quad (4.13)$$

where P_{out} is the probability of outage and η , called the average *Level Crossing Rate* (LCR), is proportional to the inverse of the average fade duration. Thus, we get:

$$K_A^* = \frac{1}{\mu_A \left(D_{min} - \frac{P_{out}}{\eta} \right)} + \frac{\lambda_A}{\mu_A}. \quad (4.14)$$

The closed form of this minimal number of streams depends on assumptions on interference.

4.4.1 In the Absence of Interference

Assuming perfect interference pre-cancellation (ZFBF), P_{out} [51] and η [50] in a wireless channel with Rayleigh fading are given by:

$$P_{out}(\gamma_0) = 1 - \frac{1}{\left(1 + \frac{\gamma_0}{\Omega_D} \right)} \quad (4.15)$$

$$\eta(\gamma_0, K_A^*) = \frac{\pi f_m}{1 + \frac{\gamma_0}{\Omega_D}} \sqrt{\frac{\Omega_D}{2\gamma_0}} \quad (4.16)$$

where f_m, γ_0, Ω_D denote the channel frequency, the outage SNR threshold and the average fade power of intended K_A^* users, respectively. This yields the following result:

Theorem 7 *Under Rayleigh fading and assuming no interference, in order to satisfy a delay requirement D_{min} of delay-sensitive (CBR) users, a MIMO BS should allocate no less than K_A^* transmit beams to such CBR-users, with K_A^* given by the following:*

$$\boxed{K_A^* = \frac{1}{\mu_A \left(D_{min} - \frac{1}{\pi f_m} \sqrt{\frac{\gamma_0}{2\Omega_D}} \right)} + \frac{\lambda_A}{\mu_A}} \quad (4.17)$$

where $f_m, \gamma_0, \Omega_D, \lambda_A, \mu_A$ denote the channel frequency, the outage SNR threshold, the average fade power of intended users, the average packet arrival of CBR traffic and the average service time per antenna of CBR traffic, respectively.

4.4.2 In the Presence of Interference

We assume that each stream among the K_A^* streams incurs interference from the other $K_A^* - 1$ streams. In such a case, denoting by Ω_I the average power of the interferers, P_{out} [51] and η [50] in a wireless channel with Rayleigh fading are given by:

$$P_{out}(\gamma_0) = 1 - \frac{1}{\left(1 + \frac{\gamma_0 \Omega_I}{\Omega_D}\right)^{K_A^* - 1}} \quad (4.18)$$

$$\eta(\gamma_0, K_A^*) = \frac{\sqrt{2\pi} f_m \Gamma(K_A^* - \frac{1}{2})}{\Gamma(K_A^* - 1)} \left(\frac{1}{1 + \frac{\gamma_0 \Omega_I}{\Omega_D}}\right)^{K_A^* - 1} \sqrt{\frac{\gamma_0 \Omega_I}{\Omega_D}}. \quad (4.19)$$

Owing to the multiplication theorem, the Gamma function is such that:

$$\frac{\Gamma(K_A^* - \frac{1}{2})}{\Gamma(K_A^* - 1)} = \frac{\sqrt{\pi}}{2^{2n+1}} \Gamma(2(K_A^* - 1)) \quad (4.20)$$

$$= \frac{\sqrt{\pi}}{2^{2n+1}} \binom{2(n-1)}{n-1}. \quad (4.21)$$

This yields the following:

Theorem 8 *Under Rayleigh fading and assuming co-channel interference, in order to satisfy a delay requirement D_{min} of delay-sensitive (CBR) users, the minimum number K_A^* of streams to be allocated by a MIMO BS to these users is solution to the following equation:*

$$K_A^* - \frac{1}{\mu_A \left(D_{min} - \frac{\left(1 + \frac{\gamma_0 \Omega_I}{\Omega_D}\right)^{K_A^* - 1} - 1}{\frac{\pi f_m}{2^{2K_A^* - 1}} \binom{2(K_A^* - 1)}{K_A^* - 1}} \sqrt{\frac{\gamma_0 \Omega_I}{\Omega_D}} \right)} - \frac{\lambda_A}{\mu_A} = 0 \quad (4.22)$$

where $f_m, \gamma_0, \Omega_D, \Omega_I, \lambda_A, \mu_A$ denote the channel frequency, the outage SNR threshold, the average fade power of intended users, the average fade power of co-channel interferers, the average packet arrival of CBR traffic and the average service time per antenna of CBR traffic, respectively.

A solution to the previous equation seems difficult to track in closed form, yet it may be approached by means of numerical search.

4.5 Proposal of a delay-aware scheduling scheme

We start by highlighting the need for a delay-aware scheduling scheme to meet the requirements of delay-sensitive applications.

4.5.1 Need for a Delay-Aware Scheduling Scheme

The probability that a user i has better SNR γ_i than that (γ_j) of a user j is given by:

$$p(\gamma_i \geq \gamma_j) = \int_{\gamma_j}^{\infty} f_i(\gamma_i) d\gamma_i \quad (4.23)$$

where f_i is the fading probability density function relative to user i . Given the aforementioned channel assumptions, such probability density is exponentially-distributed [?]:

$$f_i(\gamma_i) = \frac{1}{\bar{\gamma}} e^{-\gamma_i/\bar{\gamma}} \quad (4.24)$$

where $\bar{\gamma}$ denotes the average SNR. In order for user i to be among K_A^* scheduled users at a given time, its SNR γ_i needs to be better than, at least, $K_A - K_A^*$ A -users at that time. This event has the following probability:

$$p_i = \sum_{n=1}^{K_A^*-1} \left(\prod_{j=1}^{K_A-K_A^*+n} p(\gamma_i \geq \gamma_j) \right) \quad (4.25)$$

$$= \sum_{n=1}^{K_A^*-1} \left(\prod_{j=1}^{K_A-K_A^*+n} \int_{\gamma_j}^{\infty} f(\gamma_i) d\gamma_i \right). \quad (4.26)$$

Let T denote the maximum number of scheduling periods a packet of user i can be delayed without infringing its delay constraint D_i . As users are scheduled based on their channel condition only, scheduling at time t is independent of previous schedules. Thus, the number of times χ_i a user i is scheduled within T periods is binomial-distributed $B(T, p_i)$. Precisely, the probability that user i is scheduled χ_i times within a period of T time slots is given by:

$$\text{Prob}(\chi_i \leq T) = \binom{T}{\chi_i} p_i^{\chi_i} (1 - p_i)^{T-\chi_i} \quad (4.27)$$

where p_i is given by (25), (26). The expected value of such distribution is:

$$\mathbb{E}\{\chi_i\} = T p_i = T \sum_{n=1}^{K_A^*-1} \left(\prod_{j=1}^{K_A-K_A^*+n} \int_{\gamma_j}^{\infty} f(\gamma_i) d\gamma_i \right). \quad (4.28)$$

This means that, in a period of time T , user i will be scheduled $T p_i$ times, on the average. As the number of candidate users increases, $p_i \rightarrow 0$ (product of real numbers in $(0, 1)$), and so does $\mathbb{E}\{\chi_i\}$. Thus, under a sum-capacity maximizing scheduling policy, a user is unlikely to be

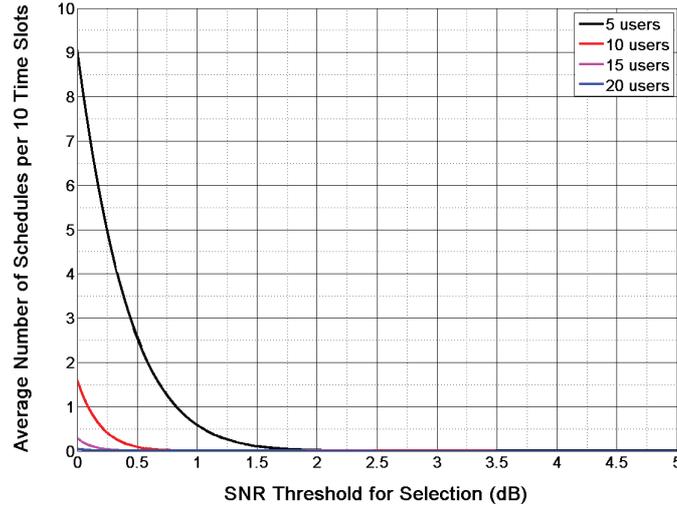


Figure 4.3: Average number of schedules (time slots) a CBR user is scheduled for transmission within $T = 10$ time slots, if a capacity-maximizing scheduling is used. Observe that as the number of candidate users increases, each user becomes less likely to be scheduled within such a period of time.

scheduled in a short period of time if the number of candidate users is large. As an illustrative example, Fig. 4.3 plots this distribution versus SNR for different numbers of candidate users. Precisely, we consider the average number of schedules (time slots) a CBR user is scheduled for transmission within an observation period of $T = 10$ time slots, if a capacity-maximizing scheduling scheme is used. Observe that as the number of candidate users increases, each user becomes less likely to be scheduled within such a time frame.

4.5.2 Proposed Delay-Aware Scheduling Scheme

Fig. 4.4 presents a flowchart of the proposed delay-aware scheduling algorithm:

- First, schedule the $a_k, 1 \leq k \leq K_A^*$ users whose packets have queuing delays about to exceed their guaranteed delays D_k .
- Subsequently select, among the remaining packets in the queue, K_B^* packets relative to the K_B^* B -users that are most orthogonal to the already-scheduled A -users.

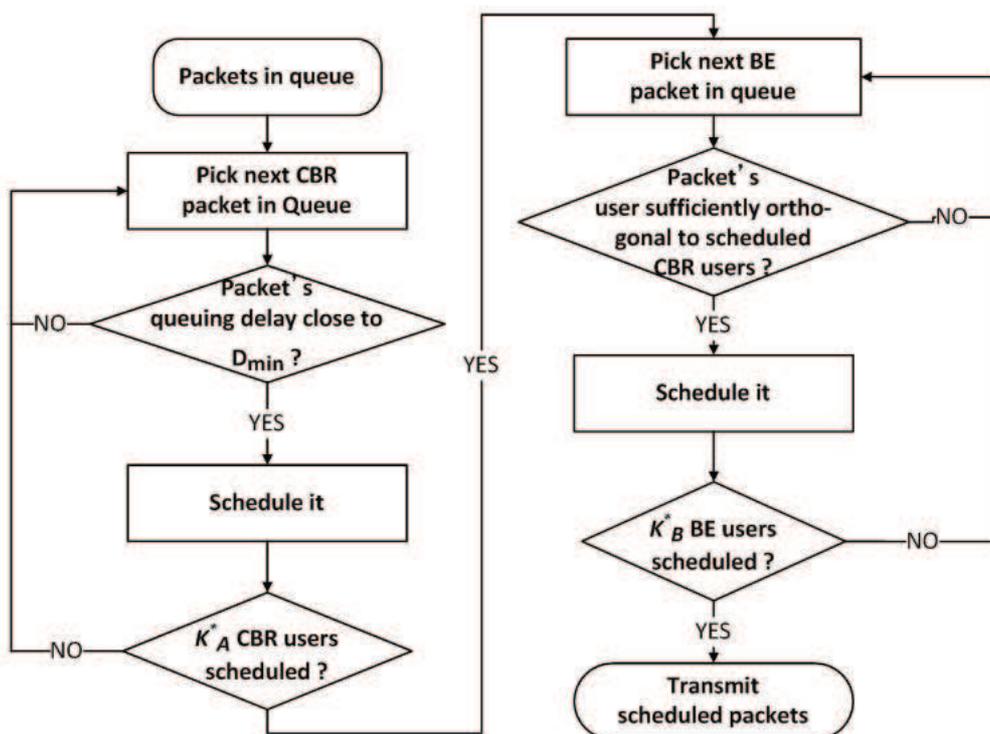


Figure 4.4: Proposed delay-aware scheduling algorithm. While conventional sum-capacity maximizing scheduling approaches schedule all users only based on their mutual orthogonality (irrespective of their queuing delays), the proposed scheme schedules delay-sensitive packets based on their queuing delay and best-effort packets based on their orthogonality w.r.t. scheduled delay-sensitive users, thus ensuring a high QoS (in terms of instantaneous delay guarantees and incurred interference) for the latter.

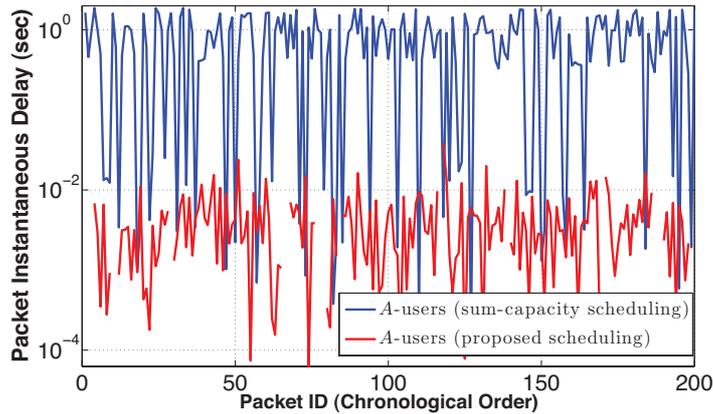


Figure 4.5: Comparison of instantaneous delays incurred by A -users' packets when using conventional and proposed scheduling policies. Owing to the by-the-order-of-magnitude delay enhancements, results are reported in logarithmic scale

4.6 Simulation Results

4.6.1 On Incurred Delays

Fig. 4.5 illustrates the packet instantaneous delay in the proposed scheduling algorithm for both A -users and B -users in logarithmic scale. The maximum tolerable delay was set to $D = 3 \times 10^{-2}$ sec. First, we observe that the proposed scheduling scheme satisfies the delay constraint of A -users, as they maintain a constant delay below the guaranteed bound. Quite the reverse, the delay in the case of sum-capacity maximizing scheduling is more significant. This is attributed to the fact that such scheduling schemes schedule packets based on *their channel conditions* rather than *their queuing delays*. As explained in the previous section, a user is less likely to be scheduled by conventional scheduling schemes as the number of candidate users becomes large. Hence, in order to meet delay requirements of delay-sensitive users, it is necessary to account for queuing delays, as in the proposed scheduling scheme (Fig. 3). A similar conclusion may be drawn from Fig. 4.6 where the average delay versus the total number of users is reported.

4.6.2 On Achievable Throughputs

Fig. 4.7 illustrates the sum-throughput versus SNR of the scheduled users, for both the conventional and the proposed scheduling schemes. An initial population of 50 users (10 A -users, 40 B -users) and a MIMO 8×8 system were considered. Therefore, at any given time slot, a maximum of 8 users can be scheduled. The sum-throughput of the proposed scheme was evaluated when, of the 8 scheduled users, 1, 2, 4, 6 A -users (resp. 7, 6, 4, 2 B -users) were scheduled. The

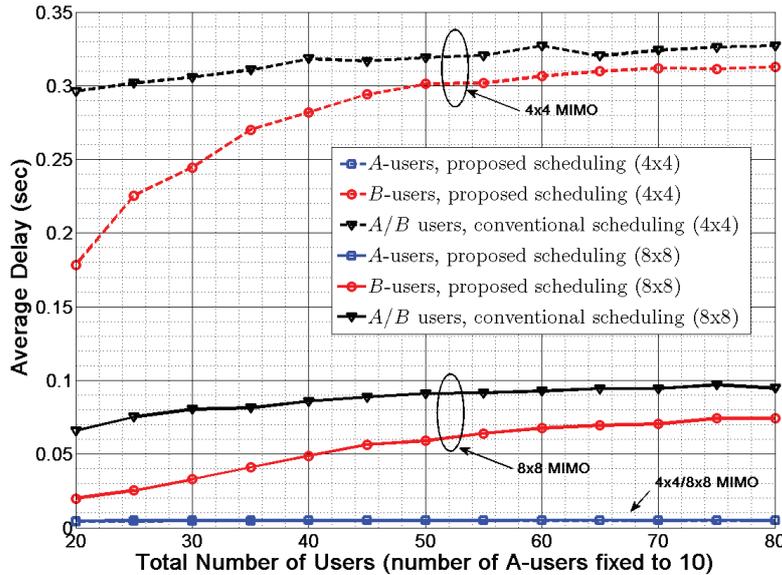


Figure 4.6: Average delay versus total number of users (10 of which are A -users), for different numbers of transmit antennas. Here, the conventional scheme is sum-capacity maximizing scheduling, which aims at scheduling users only based on channel condition, thereby disregarding the queuing delays incurred by their packets.

considered setting is a worst-case scenario for the proposed scheduling in the sense that the scheduled A -users were intentionally considered mutually interfering, while the scheduled B -users were perfectly mutually-orthogonal (i.e. non interfering with each other). Hence, we observe a degradation of the achievable sum-throughput as the ratio of A -users among the total scheduled users increases. This reduced throughput naturally trades for the support of delay-requirement of such users. However, we observe that the throughput degradation is less significant when only a few A -users are scheduled per time slot. Hence, as long as the ratio of scheduled A -users per scheduled B -users is kept to a minimum, their mutual interference would not significantly degrade the sum-throughput with the proposed scheduling scheme w.r.t. to capacity-maximizing scheduling schemes.

4.7 Conclusion

In this chapter, we considered the problem of scheduling, in closed-loop multi-user MIMO systems, a large set of users with different Quality of Service (QoS) requirements, supplied by a single BS by means of Zero-Forcing Beamforming (ZFBF). Unlike related work where all candidate users are required to feed back their CSI to BS, we provide a power-and-bandwidth efficient feedback scheme in which users may tell if they are ZFBF-optimal and feed back their CSI only in

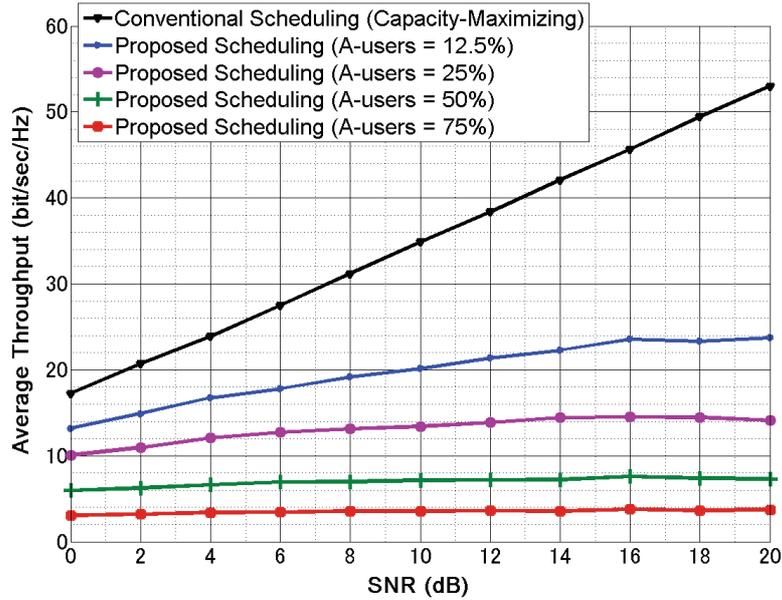


Figure 4.7: Average throughput versus SNR when using the conventional and the proposed scheduling approaches.

such a case, thereby reducing the number of required feedbacks and the computational burden of exhaustive search for best users at the BS. Subsequently, we demonstrated that conventional sum-capacity maximizing scheduling policies fall short to meet the requisites of delay-sensitive applications, and we provided appropriate delay-aware scheduling scheme for such-constrained users. Reported simulation results showed that the proposed scheduling scheme successfully meets both average and instantaneous delay constraints of delay-sensitive applications. Besides, we observe that, in order to minimize the degradation of the sum-throughput, the ratio of scheduled A -users per scheduled B -users per time slot needs to be kept to a minimum.

Chapter 5

Conclusions And Perspectives

Both the wireless industry and government frequency regulation bodies (such as the FCC) forecast , for the near futur, significant demand for bandwidth (data rates) that is triggered by an increase in the number of networked devices, a huge growth of video traffic and the emergence of more wireless applications. Further, it is expected that such demand in bandwidth is unlikely to be met by the current physical-layer channel capacity, particularly for wireless communications.

Recently, multiple-input multiple-input (MIMO) technology, the use of multiple antennas at the transmitter and the receiver, has emerged as a potential solution to meet the huge demand in wireless bandwidth. Early works on MIMO [7, 15] predict a linear growth in capacity with the number of antennas, which could allow for unprecedented wireless data rates.

Regrettably, some theoretical requirements and assumptions have challenged its implementation in practice.

In this work, we provided our (however humble) contributions towards this end by addressing some of these challenges in both single-user (SU-MIMO) and multi-user (MU-MIMO) settings.

We started by addressing MIMO challenges pertaining to single-user systems. First, we considered in chapter 2 the need for a fast channel feedback phase, as MIMO capacity gains are conditional upon the availability of channel estimates at the transmitter. From the proposed feedback scheme, we derived an application to information security by providing a secrecy scheme

through which the confidentiality of a two-way MIMO communication can be guaranteed.

Subsequently, we considered in chapter 3 the keyhole problem where the propagation environment has a single degree of freedom regardless of the number of transmit antennas. Related literature seems to consider such degeneration hopeless. Contrarily, we showed that the use of relay-assisted communications as well as a careful power allocation (that we devised) makes the keyhole effect mitigation feasible.

Finally, we addressed in chapter 4 some research challenges pertaining to multi-user systems. We considered the problem of user scheduling in MU-MIMO. We provided an efficient feedback scheme where the number of required feedbacks and the computational burden of exhaustively searching for best users at the transmitter's side are substantially reduced. Afterwards, we provided a QoS-aware scheduling scheme that allows to meet the demand of delay-constrained users.

As of the date of this dissertation, some challenges remain open and could be an interesting extension to this work.

In the area of CSI estimation and feedback, the feasibility of closed-loop MIMO at very-high receiver velocities (e.g. 100 km/hr or more) is challenging at best. At such speeds, the channel's coherence time is too small to allow for a cost-efficient (low-signaling) closed-loop MIMO.

In the area of cooperative communications, distributed MIMO systems (so-called Network MIMO) and the efficient deployment of MIMO femtocells are also interesting capacity-enhancement approaches through transmit cooperation, cross-cell interference mitigation and cell coverage extension.

Lastly, the interference channel remains an open problem. Despite over 30 years of research work, little is known in this area on the capacity limits of such channel. Open questions include the capacity region of the interference channel, the practical deployment of Han-Kobayashi interference mitigation approaches and, from an industrial perspective, the realization of a blind (transparent) maximum-likelihood detector for multi-user MIMO.

Appendix A

Proof of Proposition 1

From Theorem 1, we have:

$$C^d = \max_{p(\mathbf{x}, \mathbf{x}_1)} \min \{I(\mathbf{x}, \mathbf{x}_1; \mathbf{y}), I(\mathbf{x}; \mathbf{y}_1 | \mathbf{x}_1)\} \quad (\text{A.1})$$

To determine closed forms for the mutual information expressions, we need the probability densities $p(\mathbf{x})$, $p(\mathbf{x}_1)$ and $p(\mathbf{x} | \mathbf{x}_1)$. As for the first two distributions, it is known that circularly-symmetric Gaussian inputs are capacity-maximizers of the MIMO channel [15]. To determine the third distribution, related works such as [3, 52] decompose the relay's output \mathbf{x}_1 as follows:

$$\mathbf{x}_1 = \mathbf{x}_{10} + \mathbf{x}_{11} \quad (\text{A.2})$$

such that \mathbf{x}_{10} is *independent* from the transmitter's input, \mathbf{x} . Besides, it is also argued that the mutual information is maximized when the independent component (\mathbf{x}_{10}) is maximized, see [52], Appendix B. Therefore, we neglect, for simplicity, the contribution of \mathbf{x}_{11} to the mutual information and subsequently consider that $p(\mathbf{x} | \mathbf{x}_1) \approx p(\mathbf{x})$.

Now, we shall develop each mutual information separately. By definition, the first mutual information is given by [41]:

$$I(\mathbf{x}, \mathbf{x}_1; \mathbf{y}) \triangleq h(\mathbf{y}) - h(\mathbf{y} | \mathbf{x}, \mathbf{x}_1) \quad (\text{A.3})$$

$$= h(\mathbf{y}) - h(\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{x}_1 + \mathbf{z} | \mathbf{x}, \mathbf{x}_1) \quad (\text{A.4})$$

$$= h(\mathbf{y}) - h(\mathbf{z}) \quad (\text{A.5})$$

As the noise and channel inputs are Gaussian, then so is the channel output \mathbf{y} :

$$h(\mathbf{y}) = \log \left((2\pi e)^N \left| \mathbb{E} \left\{ \mathbf{y} \mathbf{y}^\dagger \right\} \right| \right) \quad (\text{A.6})$$

The fact that \mathbf{x} , \mathbf{x}_1 are zero-mean and independent of \mathbf{z} yields:

$$\mathbb{E} \left\{ (\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{x}_1) \mathbf{z}^\dagger \right\} = \mathbf{0} \quad (\text{A.7})$$

$$\mathbb{E} \left\{ \mathbf{z} (\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{x}_1)^\dagger \right\} = \mathbf{0} \quad (\text{A.8})$$

Meanwhile:

$$\mathbb{E} \left\{ (\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{x}_1) (\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{x}_1)^\dagger \right\} = \mathbf{H}\mathbf{Q}_{\mathbf{x}\mathbf{x}}\mathbf{H}^\dagger + \mathbf{G}\mathbf{Q}_{\mathbf{x}_1\mathbf{x}_1}\mathbf{G}^\dagger + \mathbf{H}\mathbf{Q}_{\mathbf{x}\mathbf{x}_1}\mathbf{G}^\dagger + \mathbf{G}\mathbf{Q}_{\mathbf{x}_1\mathbf{x}}\mathbf{H}^\dagger \quad (\text{A.9})$$

Therefore:

$$h(\mathbf{y}) = \log \left((2\pi e)^N \left| \frac{1}{\sigma_z^2} \mathbf{H}\mathbf{Q}_{\mathbf{x}\mathbf{x}}\mathbf{H}^\dagger + \mathbf{G}\mathbf{Q}_{\mathbf{x}_1\mathbf{x}_1}\mathbf{G}^\dagger \right| \right) \quad (\text{A.10})$$

where

$$\mathbf{Q}_{\mathbf{x}\mathbf{x}_1} \triangleq \mathbb{E} \left\{ \mathbf{x} \mathbf{x}_1^\dagger \right\} = \mathbf{0} \quad (\text{A.11})$$

$$\mathbf{Q}_{\mathbf{x}_1\mathbf{x}} \triangleq \mathbb{E} \left\{ \mathbf{x}_1 \mathbf{x}^\dagger \right\} = \mathbf{0} \quad (\text{A.12})$$

$$\mathbb{E} \left\{ \mathbf{z} \mathbf{z}^\dagger \right\} = \frac{1}{\sigma_z^2} \mathbf{I}_N \quad (\text{A.13})$$

On the other hand, \mathbf{z} is zero-mean Gaussian with variance σ_z^2 , therefore:

$$h(\mathbf{z}) = \log \left((2\pi e)^N \left| \sigma_z^2 \mathbf{I}_N \right| \right) \quad (\text{A.14})$$

Finally, we get:

$$\boxed{I(\mathbf{x}, \mathbf{x}_1; \mathbf{y}) = \log \left| \mathbf{I}_N + \frac{1}{\sigma_z^2} \mathbf{G}\mathbf{Q}_{\mathbf{x}_1\mathbf{x}_1}\mathbf{G}^\dagger + \frac{1}{\sigma_z^2} \mathbf{H}\mathbf{Q}_{\mathbf{x}\mathbf{x}}\mathbf{H}^\dagger \right|} \quad (\text{A.15})$$

As for the second mutual information, we have [3]:

$$I(\mathbf{x}; \mathbf{y}_1 | \mathbf{x}_1) \triangleq h(\mathbf{y}_1 | \mathbf{x}_1) - h(\mathbf{y}_1 | \mathbf{x}, \mathbf{x}_1) \quad (\text{A.16})$$

$$\begin{aligned} &= h(\mathbf{F}\mathbf{x} + \mathbf{z}_1 | \mathbf{x}_1) - h(\mathbf{F}\mathbf{x} + \mathbf{z}_1 | \mathbf{x}, \mathbf{x}_1) \\ &= h(\mathbf{y}_1) - h(\mathbf{z}_1) \end{aligned} \quad (\text{A.17})$$

since \mathbf{x}_1 is independent of \mathbf{x} and \mathbf{z}_1 . Besides,

$$h(\mathbf{y}_1) = \log \left| (2\pi e)^N \mathbb{E} \left\{ \mathbf{y}_1 \mathbf{y}_1^\dagger \right\} \right| \quad (\text{A.18})$$

$$= \log \left((2\pi e)^N \left| \sigma_{\mathbf{z}_1}^2 \mathbf{I}_N + \mathbf{F} \mathbf{Q}_{\mathbf{x}\mathbf{x}} \mathbf{F}^\dagger \right| \right) \quad (\text{A.19})$$

and:

$$h(\mathbf{z}_1) = \log \left((2\pi e)^N \left| \sigma_{\mathbf{z}_1}^2 \mathbf{I}_N \right| \right) \quad (\text{A.20})$$

Therefore:

$$\boxed{I(\mathbf{x}; \mathbf{y}_1 | \mathbf{x}_1) = \log \left| \mathbf{I}_N + \frac{1}{\sigma_{\mathbf{z}_1}^2} \mathbf{F} \mathbf{Q}_{\mathbf{x}\mathbf{x}} \mathbf{F}^\dagger \right|} \quad (\text{A.21})$$

Q.E.D. ■

References

- [1] “Mobile broadband: The benefits of additional spectrum,” Federal Communications Commission TECHNICAL PAPER, Nov. 2010.
- [2] C.E. Shannon, “A mathematical theory of communication,” Bell System Technical Journal, vol.27, pp.379 – 423, July 1948.
- [3] T.M. Cover and J.A. Thomas, Elements of Information Theory, John Wiley & sons, 1991.
- [4] S. Santini, “We are sorry to inform you ...,” Computer, vol.38, no.12, pp.128 – 127, Dec. 2005.
- [5] J. Salz, “Digital transmission over cross-coupled linear channels,” AT&T Technical Journal, vol.64, no.6, pp.1147 – 1159, July/Aug. 1985.
- [6] D. Gerlach and A. Paulraj, “Base station transmitter antenna arrays with mobile to base feedback,” 1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, (Asilomar’1993), pp.1432 – 1436 vol.2, Nov. 1993.
- [7] G.J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” Bell Labs Technical Journal, vol.1, pp.41 – 59, 1996.
- [8] J. Li and P. Stoica, MIMO radar signal processing, J. Wiley & Sons, 2008.
- [9] P.D. Arapoglou, K. Liolis, M. Bertinelli, A. Panagopoulos, P. Cottis, and R. De Gaudenzi, “Mimo over satellite: A review,” IEEE Communications Surveys Tutorials, vol.13, no.1, pp.27 – 51, quarter 2011.
- [10] A. Goldsmith, Wireless Communications, Cambridge University Press, New York, USA, 2005.

- [11] B. Hassibi and B. Hochwald, "How much training is needed in multiple - antenna wireless links?," *IEEE Transactions on Information Theory*, vol.49, no.4, pp.951 – 963, Apr. 2003.
- [12] S. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," *2001 IEEE International Symposium on Information Theory, (ISIT'2001)*, p.321, 2001.
- [13] D. Love, J. Heath, R.W., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol.49, no.10, pp.2735 – 2747, Oct. 2003.
- [14] "3gpp - ts 36.104: Base station (bs) radio transmission and reception (release 9)," *3GPP TSG RAN-1 Working Group*, Dec. 2009.
- [15] E. Telatar, "Capacity of multi - antenna gaussian channels," *European Transactions on Telecommunications*, vol.10, pp.585 – 595, 1999.
- [16] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Mimo wireless channels: capacity and performance prediction," *IEEE Global Telecommunications Conference, 2000 (GLOBE-COM'00)*, pp.1083 – 1088, 2000.
- [17] D. Chizhik, G. Foschini, M. Gans, and R. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Transactions on Wireless Communications*, vol.1, no.2, pp.361 – 368, apr 2002.
- [18] P. Almers, F. Tufvesson, and A. Molisch, "Keyhole effect in mimo wireless channels: Measurements and theory," *IEEE Transactions on Wireless Communications*, vol.5, no.12, pp.3596 – 3604, Dec. 2006.
- [19] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol.29, no.3, pp.439 – 441, may 1983.
- [20] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol.24, no.3, pp.528 – 541, Mar. 2006.
- [21] L. Withers, R. Taylor, and D. Warne, "Echo-mimo: A two-way channel training method for matched cooperative beamforming," *IEEE Transactions on Signal Processing*, vol.56, no.9, pp.4419 –4432, Sept. 2008.

- [22] O. Souihli and T. Ohtsuki, "Transparent inband feedback for training-based mimo systems," *Wireless Personal Communications*, pp.1 – 13, 2010. 10.1007/s11277-010-0152-z.
- [23] G. Levin and S. Loyka, "On the outage capacity distribution of correlated keyhole mimo channels," *IEEE Transactions on Information Theory*, vol.54, no.7, pp.3232 – 3245, July 2008.
- [24] A. Maaref and S. Aissa, "Impact of spatial fading correlation and keyhole on the capacity of mimo systems with transmitter and receiver csi," *IEEE Transactions on Wireless Communications*, vol.7, no.8, pp.3218 – 3229, Aug. 2008.
- [25] G. Levin and S. Loyka, "Multi-keyhole mimo channels: Asymptotic analysis of outage capacity," *2006 IEEE International Symposium on Information Theory*, pp.1305 – 1309, July 2006.
- [26] C. Zhong, S. Jin, K. Wong, and M. McKay, "Ergodic mutual information analysis for multi-keyhole mimo channels," *IEEE Transactions on Wireless Communications*, vol.PP, no.99, pp.1 – 10, 2011.
- [27] H. Shin and J.H. Lee, "Performance analysis of space-time block codes over keyhole nakagami-m fading channels," *IEEE Transactions on Vehicular Technology*, vol.53, no.2, pp.351 – 362, Mar. 2004.
- [28] S. Sanayei, A. Hedayat, and A. Nosratinia, "Space time codes in keyhole channels: Analysis and design," *IEEE Transactions on Wireless Communications*, vol.6, no.6, pp.2006 –2011, June 2007.
- [29] H. Zhao, Y. Gong, Y.L. Guan, and Y. Tang, "Performance analysis of m-psk/m-qam modulated orthogonal space-time block codes in keyhole channels," *IEEE Transactions on Vehicular Technology*, vol.58, no.2, pp.1036 – 1043, Feb. 2009.
- [30] N. Zlatanov, Z. Hadzi-Velkov, and G. Karagiannidis, "Level crossing rate and average fade duration of the double nakagami-m random process and application in mimo keyhole fading channels," *IEEE Communications Letters*, vol.12, no.11, pp.822 –824, Nov. 2008.
- [31] Y. Gong and K. Letaief, "On the error probability of orthogonal space-time block codes over keyhole mimo channels," *IEEE Transactions on Wireless Communications*, vol.6, no.9, pp.3402 – 3409, Sep. 2007.

- [32] O. Souihli and T. Ohtsuki, "Cooperative diversity can mitigate keyhole effects in wireless mimo systems," 2009 IEEE Global Telecommunications Conference (GLOBECOM'2009), pp.1 – 6, Dec. 2009.
- [33] O. Souihli and T. Ohtsuki, "The mimo relay channel in the presence of keyhole effects," 2010 IEEE International Conference on Communications (ICC), pp.1 – 5, May 2010.
- [34] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," IEEE Transactions on Information Theory, vol.48, no.6, pp.1277 – 1294, Jun 2002.
- [35] O. Souihli and T. Ohtsuki, "Joint feedback and scheduling scheme for service-differentiated multiuser mimo systems," IEEE Transactions on Wireless Communications, vol.9, no.2, pp.528 – 533, Feb. 2010.
- [36] O. Souihli and T. Ohtsuki, "Joint feedback and scheduling scheme for service-differentiated multiuser mimo systems," 2009 IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), pp.1 –5, Sept. 2009.
- [37] G. Scutari, D. Palomar, and S. Barbarossa, "The mimo iterative waterfilling algorithm," IEEE Transactions on Signal Processing, vol.57, no.5, pp.1917 – 1935, may 2009.
- [38] S. Ohno and G. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless ofdm," IEEE Transactions on Communications, vol.50, no.12, pp.2113 – 2123, Dec 2002.
- [39] P. Dent, G. Bottomley, and T. Croft, "Jakes fading model revisited," Electronics Letters, vol.29, no.13, pp.1162 – 1163, June 1993.
- [40] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading mimo channels with channel estimation error," IEEE Transactions on Information Theory, vol.52, no.5, pp.2203 – 2214, may 2006.
- [41] T. Cover and A. Gamal, "Capacity theorems for the relay channel," IEEE Transactions on Information Theory, vol.25, no.5, pp.572 – 584, Sep. 1979.
- [42] J.L.W.V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," Acta Mathematica, vol.30, pp.175 – 193, 1906.

- [43] S. Jafar and A. Goldsmith, "Multiple - antenna capacity in correlated rayleigh fading with channel covariance information," *IEEE Transactions on Wireless Communications*, vol.4, no.3, pp.990 – 997, May 2005.
- [44] K. Kobayashi, T. Ohtsuki, and T. Kaneko, "Mimo systems in the presence of feedback delay," 2006 *IEEE International Conference on Communications (ICC '06)*, pp.4102 – 4106, June 2006.
- [45] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *IEEE Journal on Selected Areas in Communications*, vol.21, no.5, pp.684 – 702, June 2003.
- [46] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol.49, no.7, pp.1691 – 1706, July 2003.
- [47] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol.26, no.8, pp.1341 – 1365, Oct. 2008.
- [48] Y. Al-Harathi, A. Tewfik, and M.S. Alouini, "Opportunistic scheduling with quantized feedback in wireless networks," *International Conference on Information Technology: Coding and Computing, 2005 (ITCC'2005)*, pp.716 – 722 Vol. 2, Apr. 2005.
- [49] L. Breuer and D. Baum, *An introduction to queueing theory and matrix-analytic methods*, Bibliografia. Índex, Springer, 2005.
- [50] M. Simon and M. Alouini, "On the difference of two chi-square variates with application to outage probability computation," *IEEE Transactions on Communications*, vol.49, no.11, pp.1946 – 1954, Nov. 2001.
- [51] L. Yang and M. Alouini, "On the average outage rate and average outage duration of wireless communication systems with multiple cochannel interferers," *IEEE Transactions on Wireless Communications*, vol.3, no.4, pp.1142 – 1153, July 2004.
- [52] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of mimo relay channels," *IEEE Transactions on Information Theory*, vol.51, no.1, pp.29 – 43, Jan. 2005.