UNIVERSITÀ DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Wireless Systems with Quantized Information: Multiuser MIMO and Networked Control Systems

Ph.D. THESIS

Author: Matteo Trivellato
Director of Ph.D. School: Prof. Matteo Bertocco
Supervisor: Prof. Nevio Benvenuto

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

SCUOLA DI DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE
XXI CICLO
Indirizzo: Ingegneria elettronica e delle telecomunicazioni

Abstract

This thesis deals with two multiuser wireless communication systems: i) the multiple-input multiple-output (MIMO) broadcast channel (BC) where a multi-antenna transmitter serves multiple users along spatially multiplexed channels and ii) a networked control system (NCS) where spatially distributed sensors, controllers and actuators exchange information via a digital wireless network in order to estimate or control a dynamical system.

In MIMO BC, channel state information at transmitter (CSIT) is essential to achieve spatial multiplexing across users. Special interest is on frequency division duplexing systems where CSIT is provided through limited uplink feedback (FB) from the receivers. Either in case of single antenna or multi-antenna receivers the main contributions are: i) the design of novel linear transceiver strategies that account for limited CSIT, ii) the proposal of channel quantization techniques and FB strategies that exploit spatial and time correlation of the MIMO channel and iii) the derivation of efficient and robust user selection schemes for the maximization of the achievable throughput. In MIMO downlink systems the potential gains of multiuser over more conventional single user transmission strategies are also evaluated in a multi-cell cellular network where coordination among spatially distributed base stations and higher order sectorization are investigated as possible methods to mitigate inter-cell interference. In case of multiuser MIMO orthogonal frequency division multiplexing (OFDM) downlink systems we provide non trivial generalizations of channel quantization strategies proposed for single carrier flat fading systems. Interestingly, concentrating FB bits to characterize only a portion of the available bandwidth at receivers and the possibility of exploiting multiuser diversity can increase significantly the achievable throughput.

In NCSs where system measurements come from multiple spatially distributed sensors, the main contribution is the generalization of estimation and control techniques to account for wireless link inefficiencies: i) packet loss, ii) delays and iii) signal quantization. In particular sensors, controller and actuator share a common frequency resource motivating a cross layer optimization of i) signal/measurements quantization processes and ii) network resource allocation. Even with small transmission bandwidth, single-hop communication protocols with binary phase shift keying provide close to optimum performance in applications dealing with state estimation or state control of a stable system. This support the widespread use of low-cost sensors for these applications.

Sommario

Questa tesi si occupa di due sistemi di comunicazione wireless: i) il canale broadcast (BC) multiple-input multiple-output (MIMO) dove un trasmettitore equipaggiato con piu antenne serve più utenti lungo canali multiplati spazialmente e ii) e un sistema di controllo connesso attraverso una rete wireless (NCS) dove sensori distribuiti nello spazio, controllori e attuatori scambiano informazioni attraverso una rete wireless digitale con lo scopo di stimare o controllare un sistema dinamico.

Nel sistema MIMO BC è essenziale avere informazioni sullo stato del canale al trasmettitore (CSIT) in modo da ottenere multiplazione spaziale degli utenti. In modo particolare vengono considerati sistemi duplex a divisione di frequenza dove CSIT viene fornita dai ricevitori attraverso canali di feedback (FB) a banda limitata. Considerando ricevitori con una o più antenne i contributi principali sono i seguenti: i) il progetto di nuove strategie di trasmissione e ricezione lineari che tengano in considerazione la conoscenza limitata del canale in trasmissione, ii) la proposta di techniche di quantizzazione del canale e strategie di FB che sfruttino la correlazione spaziale e temporale del canale MIMO e iii) la derivazione di schemi di selezione degli utenti robusti ed efficienti con lo scopo di massimizzare il throughput del sistema. I potenziali vantaggi di techniche di trasmissione MIMO multiutente rispetto a più convenzionali strategie MIMO singolo-utente sono stati verificati anche in reti cellulare multi cella, dove la possibiltà di coordinare la trasmissione tra stazioni radio base distribuite nello spazio e una più fine settorizzazione all'interno di ciascuna cella sono state studiate come possibili metodi per mitigare l'interferenza tra le varie celle. In sistemi MIMO multiutente con multiplazione ortogonale a divisione di frequenza (OFDM) diamo generalizzazioni non banali di strategie di quantizzazione di canale proposte per sistemi singola portante con canali non dispersivi. In questo caso un risultato interessante è che la concentrazione di tutti i bit di FB nella caratterizzazione di solo una parte dello banda disponibile e la possibilità di sfruttare la diversità multiutente possono portare significativi guadagni al throughput del sistema.

In NCS dove misure dello stato del sistema arrivano da sensori distribuiti nel dominio dello spazio, il contributo principale è la generalizzazione di techniche di stima e controllo che tengano in considerazione le problematiche del canale wireless: i) perdita di pacchetti, ii) ritardi e iii) quantizzazione dei segnali. Inoltre, sensori, controllore e attuatore condividono una banda di trasmissione comune, e questo suggerisce un ottimizzazione di tipo cross-layer delle seguenti problematiche: i) processi di quantizzazione di segnali e misure e ii) allocazione di risorse nella rete. Utilizzando protocolli di comunicazione single-hop e una modulazione binaria di fase è possibile ottenere prestazioni quasi ottime in problemi di stima e controllo dello stato per sistema stabili. E questo addirittura con una piccola banda di trasmissione.

Tutto ciò supporta l’utilizzo di sensori a basso prezzo in applicazioni di questo tipo.

”Everything should be made as simple as possible, but not simpler.“

Albert Einstein

Acknowledgments

This Ph.D. thesis concludes an exciting period of my life in which I had the opportunity to study, work and exchange ideas with many wonderful people.

First I want to thank my advisor, Prof. Nevio Benvenuto, who helped me in acquiring a scientific but also practical approach in solving problems. He guided me during my research work but also gave me enough freedom to develop and investigate my own ideas. Another person I really want to thank is Federico Boccardi that suggested me novel and exciting research directions at the beginning of my "journey". It was wonderful to work with him during the last three years because we could always learn from each other, sharing a common enthusiasm for the research. I am also deeply indebt to Howard Huang that supervised me during my internship at Bell Laboratories. From him I learnt how approaching problems in a schematic way usually helps to find practical and effective solutions. Moreover he involved me in exciting projects where theoretical studies were always accompanied by practical considerations. Other two people I really want to thank are Stefano Tommasin and Ermanna Conte. It was really nice to work, discuss and exchange ideas with them and I have to say that together with Prof. Benvenuto we were really a good team.

Thanks also to Prof. Silvano Pupolin that was my advisor at the beginning of my Ph.D. and to many colleagues and friends at the university of Padova: Antonio Assalini, Mior Alessandra, Anahita Goljahani, Simone Merlin, Filippo Tosato, Daniele Veronesi and Emiliano Dall'Anese with whom I exchanged opinions and ideas. A special thank also to the people in the Wireless research group at Bell Labs in Crawford Hill that never said "no" when I knocked at their doors for questions or suggestions.

Finally, thanks to the members of my family and my girlfriend Veronica for their support, patience and useful advice during this "journey".

Contents

List of symbols

$\mathbb{N}$: natural numbers

$\mathbb{R}$: real numbers

$\mathbb{C}$: complex numbers

$\mathcal{A}$: generic set of elements

a: real or complex scalar value

$\mathbf{a}$: real or complex vector

$\mathbf{A}$: real or complex matrix

$\mathbf{I}$: identity matrix

$\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$: Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$

$\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$: complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$

$|\mathcal{A}|$: cardinality of set $\mathcal{A}$

$|a|$: absolute value of scalar number a

$\mathcal{N}(\mathbf{A})$: null space of matrix $\mathbf{A}$

$\mathbf{A}^T$: transposition of matrix $\mathbf{A}$

$\mathbf{A}^H$: Hermitian of matrix $\mathbf{A}$

$\mathbf{A}^{-1}$: inverse of matrix $\mathbf{A}$

$[\mathbf{A}]_{(\ell,m)}$: (ℓ, m) element of matrix $\mathbf{A}$.

$||\mathbf{a}||$: euclidian norm of vector $\mathbf{a}$

$\mathrm{E}\,[a]$: expectation of the scalar a

$\mathrm{diag}(\mathbf{a})$: diagonal matrix whose diagonal is the vector $\mathbf{a}$

$\rho(\mathbf{A})$: rank of matrix $\mathbf{A}$

$\lambda_i(\mathbf{A})$: ith eigenvalue of matrix $\mathbf{A}$. The eigenvalues are in decreasing order

$\mathrm{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_N)$: block diagonal matrix whose blocks on the main diagonal are given by the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_N$

$\mathbf{A}^\dagger$: Moore-Penrose inverse of matrix $\mathbf{A}$

tr($\mathbf{A}$): trace of matrix $\mathbf{A}$

$Co(\cdot)$: convex hull operator

$\Re$: real part

$\Im$: imaginary part

List of acronyms

ACK: Acknowledgment

ad-RA: adaptive rate allocation

AWGN: Amplitude white Gaussian noise

BC: Broadcast channel

BD: Block diagonalization

BeFB: Best RB feedback

BF: Beamformer/Beamforming

BFB: Basic feedback strategy

BLAST: Bell Labs Layered Space-Time

BPSK: Binary PSK

BS: Base station

CDF: Cumulative density function

CDI: Channel direction information

CLB: Closed-loop BLAST

CQI: Channel quality information

CSI: Channel state information

CSIT: Channel state information at transmitter

DFB: *Distributed FB* among RBs

DFT: Discrete Fourier Transform

DPC: Dirty paper coding

DSN: Digital sensor network

DSP: Digital signal processing

EV-DO: Evolution data optimized

FB: Feedback

FDD: Frequency division duplexing

FSC: *Frequency selective* Rayleigh fading MIMO *channel*

GUS: Greedy user selection

HFB Hierarchical feedback

IC: Interference channel

Id-C: *Ideal* i.i.d. MIMO *channel*. The channel frequency response in an OFDM system is modelled as constant within a RB and independent across different RBs

KKT: Karush-Kuhn-Tucker

LAN: Local Area Network

LBG: Linde, Buzo and Gray

LOS: Line of sight

LQG: Linear quadratic Gaussian

LTE: Long term evolution

MAC: Multiple access channel or Medium access control depending on the context

MESC: Maximum estimated SINR combiner

MET: Multi-user eigenmode transmission

MH-R: *multi-hop* communication with packet *retransmissions*

MIMO: Multiple input-multiple output

ML: maximum likelihood

MMSE: Minimum mean square error

MRC: Maximum ratio combining/combiner

MSE: Mean square error

MT: Mobile terminal

MU: Multi user

NACK: Not acknowledgment

NCS: Networked control system

OFDM: Orthogonal frequency division multiplexing

PCSIT: Perfect channel state information at transmitter

PDF: Probability density function

PFB: Predictive feedback

PFB-SE: PFB with state evolution

PSK: Phase-shift keying

QBC: Quantization based combining

QEV: Predictive feedback with *quantization* of the *error vector*

QoS: Quality of service

QPSK: Quadrature PSK

QUB: Quantization upper bound

RAS: Receive antenna selection

RB: Resource block

RBCM-Q: RB channel matrix quantization

RBCV-Q: RB channel vector quantization

RF: Radio frequency

RM: Predictive feedback with unitary *rotation matrix*

RVQ: Random vector quantization

SDMA: Spatial division multiple access

SH-nR: *Single-hop* communication protocol with *no* packet *retransmission*

SH-R: *Single-hop* communication protocol with packet *retransmission*

SINR: Signal-to-interference-plus-noise ratio

SM: Spatial multiplexing

SNR: Signal-to-noise ratio

SPC: Sum power constraint

STBC: Space time block coding

SU: Single user

SUS: Semi orthogonal user selection

SVD: Singular value decomposition

TCP: Transmission control protocol

TD: Time domain

TDD: Time division duplexing

TDMA: Time division multiple access

TD-UQ: Time-domain uniform quantization

U: Unitary

ZF: Zero forcing

ZF-1: *ZF* BF with at most *one* stream per active user. ZF BF where each active user can be served only along its dominant eigenmode

ZF-M: *ZF* BF with possible *multiple* streams per selected user. General MET technique, based on ZF BF, which allows the selection of multiple streams per user

Chapter 1

Introduction

A multi-user wireless communication system is a competitive environment where multiple users compete for a common resource, the transmission bandwidth [1]. From an information theoretic point of view we can distinguish between three different multiuser communication systems [2]. The first configuration is the interference channel (IC) where neither transmitters not receivers can cooperate [3, 4, 5]. A typical IC is a sensor network where spatially distributed sensors exchange messages using a common frequency resource. The second one is the multiple access channel (MAC) where no-cooperating transmitters communicate with a single receiver [6, 7, 8]. This configuration models the uplink of a cellular network where the base station can jointly decode the independent messages coming from the different users. The last one is the broadcast channel (BC) where a single transmitter serves multiple no-cooperating receivers [9, 10, 11, 12, 13]. This models the downlink of a cellular system where a base station jointly encodes and transmits independent messages to different users which cannot cooperate.

In this thesis we study two different multiuser wireless systems. The first part considers a BC where both transmitter and receivers might be equipped with multiple antennas. The potential gains of multiple-input multiple-output (MIMO) transmission techniques over more conventional single-antenna transmission strategies were highlighted since the pioneering works by Foschini and Gans [14] and Telatar [15]. Even if most of the research activities in the last decade focused on single-user (SU) MIMO where multiple spatial channels are allocated to a single user, recently there has been a shift to multiuser configurations [16]. In particular in MIMO BC the additional degrees of freedom provided by multiple antennas are used to serve multiple users along spatially multiplexed channels. Unfortunately, differently from SU MIMO where channel state information at transmitter (CSIT) is optional, in a MIMO BC, CSIT is essential to achieve spatial multiplexing across users. In a time division duplexing (TDD) system channel knowledge at transmitter can be obtained through channel estimation in the uplink, exploiting channel reciprocity. Differently in a frequency division duplexing (FDD) system the transmitter must rely on uplink feedback (FB) from the users to obtain CSIT. The main contribution of the first part of the thesis is the proposal of transceiver architectures and channel feedback strategies for MIMO BC with limited uplink feedback from receivers. Even with low rate connections between transmitter and receivers we can significantly improve the network throughput with respect to SU MIMO techniques, supporting the development of MU MIMO

deployments in next generation cellular systems. This first part of the thesis covers Chapters 2-6.

The second part of the thesis considers a special interference channel, a networked control system (NCS) [17]. In a NCS the aim is to estimate or control one or more dynamical systems, using multiple sensors, actuators and controllers that are not physically co-located and need to exchange information via a wireless digital communication network. In NCSs measurements and control packets are subject to random delay and loss, [18]. Moreover, as to each component is effectively allocated only a small portion of the available bandwidth, measurements and control information need to be quantized and this affects the stability of the system, [19]. The main contribution of this part of the thesis is the generalization of classic estimation and control techniques to account for wireless link inefficiencies: i) packet loss, ii) delays and iii) signals quantization. Moreover we show how a cross-layer design of communication and estimation/control systems might provide significant performance improvements over a separated approach. A deeper understanding of the inter-connections between communication and control systems, together with the widespread proliferation of wireless sensor networks, promises a tremendous improvement of our capabilities of monitoring and controlling the environment. This second part of the thesis covers Chapters 7-8.

In more details, the contributions of this thesis can be summarized as follows.

- In Chapter 2 we consider a MIMO BC with perfect CSIT where both transmitter and receivers might be equipped with multiple antennas. After reviewing the capacity achieving dirty paper coding (DPC) we describe a suboptimum linear transmission strategy denoted as multiuser eigenmode transmission (MET). MET is based on zero forcing (ZF) beamforming (BF), therefore each active user sees no interference from other users' messages. In case the total number of users' receive antennas is larger than the number of transmit antennas, MET selects a set of active users and a set of eigenmodes per each user in order to maximize the weighted sum rate. A simple greedy eigenmode selection algorithm is described which provides close to optimum performance. A simplified version of MET which serves each selected user only along its dominant eigenmode performs close to MET as the number of users increases. Numerical comparisons with other linear precoding schemes confirm the effectiveness of MET.

- In Chapter 3 we consider a MIMO BC with limited FB from single antenna receivers and investigate three different problems: i) beamformer design, ii) feedback signalling optimization and iii) user selection. After reviewing ZF beamforming we propose a new minimum mean square error (MMSE) beamformer under incomplete CSI that takes into account the channel quantization error. Recalling a well known result under perfect CSIT, MMSE BF shows significant performance improvements in case of randomly selected users but gives reduced gains with respect to ZF BF in case of opportunistic user selection. As a second contribution we propose channel quantization techniques and various feedback strategies based on the Lloyd-Max algorithm [20] that exploit both spatial and time correlation of the MIMO channel. In particular we derive a hierarchical FB (HFB) approach where FB bits are accumulated over multiple signalling intervals in order to index a much larger codebook. Moreover we propose predictive FB where both

transmitter and users predict the evolution of the channel vector and the users adjust the prediction by feeding back a quantized version of the prediction error to the transmitter. Finally we design two user selection algorithms based on [21] that rely on users FB and show improved performance with respect to state-of-the-art algorithms. Numerical results provide a comparison between the proposed schemes.

- In Chapter 4 we still consider a MIMO BC with limited feedback but, differently from Chapter 3, users are equipped with multiple antennas. We propose solutions for i) transceiver design and ii) channel quantization that exploit the additional degrees of freedom provided by multiple antenna receivers. Under the assumption of at most one stream per selected user we propose a first technique based on ZF beamforming and maximum estimated signal-to-interference plus noise ratio (SINR) combiner (MESC). We provide an analytic characterization of the achievable throughput of the proposed combiner in case of many users and show how additional receive antennas or higher multiuser diversity can reduce the required feedback rate to achieve a target throughput. Moreover we extend channel quantization techniques and feedback strategies introduced in Chapter 3 exploiting multiple receive antennas. As a second technique we propose unitary (U) BF with MMSE combiner extending [22]. The codebook is designed to comprise an high number of unitary matrices to be used as tentative precoders. U-BF simplifies the control signalling and provides very competitive performance for low FB rates. Numerical results validate the effectiveness of the proposals with respect to state-of-the-art techniques.

- In Chapter 5 we evaluate the performance of MU MIMO in a real multi-cell packet-based cellular network with full frequency reuse, using SU MIMO and DPC as terms of comparison. Fairness among user is guaranteed by the multiuser proportional fair scheduling algorithm [23]. In a TDD system, under the assumption of perfect CSIT, we investigate two approaches to mitigate inter-cell interference: i) network MIMO where transmission is coordinated among spatially distributed base stations and ii) higher order sectorization where parallel spatial channels are created physically rather than through beamforming. Network MIMO requires user messages and channel state information to be shared among the coordinated bases, resulting in the need for enhanced backhaul capabilities. Nevertheless coordination among co-located sectors of a cell already provides a significant throughput gain over a SU MIMO baseline. The potential gains of MU MIMO over SU MIMO are observed even in a FDD system with low rate uplink FB channels, supporting the development of MU MIMO in next generation cellular deployments.

- In Chapter 6 the single carrier system model used in Chapter 3 is extended to a MU MIMO orthogonal frequency division multiplexing (OFDM) downlink system with limited FB from single antenna receivers. To reflect the operation of 4th generation wireless communication systems the available bandwidth is divided into resource blocks (RBs) whose number of subcarriers reflects the coherence bandwidth of the channel. The chapter contains two main contributions. Firstly we provide joint conditions on the channel coherence bandwidth and the FB rate per RB that allow for a simpler quantization of the

RB channel matrix (space-frequency) by a space vector, causing negligible performance loss in terms of system achievable throughput. As a second contribution we investigate the trade-off between accurate channel knowledge and frequency/multiuser diversity. It is seen that even for a moderate number of users in the network, concentrating all the available FB bits in characterizing only one RB provides a significant gain in system throughput over a more classical distributed approach and this result is validated both analytically and by simulations.

- In Chapter 7 we start dealing with the second topic of the thesis. The chapter considers state estimation in NCSs where observations from multiple sensors are subject to random delays and packet losses. We derive the minimum error covariance estimator and the optimum estimator with constant gains as a low-complexity solution. Generalizations to account for the effects of measurements quantization and limited transmission bandwidth are investigated for a stable system. Assuming a simple scalar system, we show how the proposed framework can be exploited for the design of NCSs. In particular we investigate i) cross-layer optimization of quantization processes and network resource allocation and ii) comparison between single-hop and multi-hop communication protocols. We show that simple BPSK and single-hop communication protocols provide close to optimum performance in applications dealing with state estimation of a stable system.

- In Chapter 8 we assume a NCS with single-hop communications and a TCP-like protocol between controller and actuator. We solve the problem of optimum control around a target state for a stable system in case of both packet drops and signal quantization. Generalization for unstable systems is also given for large bandwidth transmissions. Moreover we derive the limiting behavior of the system in the infinite horizon and propose a general framework for cross-layer optimization of signal quantization and network resource allocation. As an example of application, we consider a simple scalar, stable system and compare network resource allocation in the presence of i) low-cost sensors using a fix modulation and ii) long-term future sensors capable of rate adaptation. Interestingly, almost optimal control is achievable with small bandwidth transmissions and simple BPSK, supporting the use of low-cost sensors in applications dealing with state control in stable systems.

- In Chapter 9 we conclude the thesis summarizing the main findings of the different chapters.

Chapter 2

Precoding schemes for multiuser MIMO downlink

The demand for higher speed communications in future wireless cellular networks motivated an intensive study of multi-antenna transmission techniques which can provide significant performance gains over conventional single-antenna transmission strategies [14, 15]. Much of the MIMO research in the last decade has focused on single-user (SU) MIMO techniques where the multiple spatial channels are allocated to a single user during a given transmission interval. But lately, there is an increasing attention to multiuser (MU) configurations, where a multi-antenna transmitter serves multiple users over spatially multiplexed channels [16]. Differently from SU MIMO transmissions where channel state information at transmitter (CSIT) is optional, in MU MIMO CSIT is essential to achieve spatial multiplexing across users.

In this chapter we consider a time division duplexing (TDD) system where the same band is used for the uplink and the downlink. In this case CSIT for the downlink can be obtained through channel estimation on the uplink and the assumption of ideal CSIT becomes reasonable. Ideal CSIT is not an appropriate assumption in frequency division duplexing (FDD) systems where the base station must rely on uplink feedback (FB) from the users to obtain CSI. Transmission strategies based on limited FB are investigated in Chapters 3-6.

It has recently been shown [9] (see also [10, 11, 12, 13]) that the capacity region of the MIMO Broadcast Channel (BC) can be achieved by means of a nonlinear transmission technique known as "dirty paper coding" (DPC) [24]. Because of the considerable complexity of DPC, practical implementations using suboptimum non-linear techniques are an active area of research [25], [26], [27], [28], [29].

Linear beamforming techniques with lower complexity have also been proposed in which the transmitted signal is a spatially multiplexed, linear combination of the users' data signals. Beamforming is also known as linear precoding or spatial division multiple access (SDMA). One class of beamforming techniques for the case of a single-antenna receivers is based on zero forcing (ZF) [28], [30], [21], [31], where each user receives only its desired signal with no interference. ZF beamforming provides good performance in high signal-to-noise ratio (SNR) scenarios [31], while for a lower SNR, a better solution is given by minimum mean square error (MMSE) beamforming that balances the effects of noise and multiuser interference [26].

However, the marginal gain of MMSE beamforming with respect to ZF vanishes when an opportunistic scheduling is considered for downlink transmission and ZF achieves close to optimum performance [28].

Extensions of the zero forcing technique to the case of multiple antenna receivers appear in [32], [33], [34], where multiple spatial streams (or eigenmodes) are transmitted to each user with no interuser interference, resulting in a block diagonal transmit covariance matrix. In this class of techniques, denoted as block diagonalization (BD), the eigenmodes of a given user are active regardless of their spatial relationship with other users' modes. Due to the interuser orthogonality requirement, this assumption may lead to poor performance when modes are highly correlated. To address this drawback, alternative solutions have been proposed which perform receive antenna selection [35] or iterative optimization of transmitter and receivers [36].

Using linear transmission techniques like ZF or BD, the number of independent spatial streams is at most equal to the number of transmit antennas. In practical downlink cellular networks the number of users is much larger than the number of transmit antennas, therefore the transmitter must select a set of "active users" for receiving data and a set of modes per each user. This user selection could be performed optimally using a brute-force search over all possible combinations of users and modes, but due to the high complexity when the number of users is large, suboptimum techniques based on a greedy selection algorithm have been shown to provide near-optimum performance for the case of single antenna receivers [21].

In this chapter after reviewing the capacity achieving DPC we describe a linear transceiver technique known as multi-user eigenmode transmission (MET), where the eigenmodes satisfy a ZF criterion and the transmitter determines the set of active eigenmodes with no restrictions on how they are distributed among the users. By providing greater flexibility in determining the active eigenmodes, MET is a generalization of [37] and the basic BD techniques, and it provides performance gains in cases where eigenmodes are highly correlated. Each active user under MET uses a linear combiner across all receive antennas, independent of the number of eigenmodes it receives, thereby achieving a combining gain advantage over [35]. In order to simplify the eigenmode selection, we propose an extension of the greedy user selection technique in [21]. As a simplified version of MET we also propose a ZF beamforming strategy where each active user can be served only along its dominant eigenmode. This technique is less complex to implement, requires less control signalling and provides close to optimum performance in practical environments.

The chapter is organized as follows. In Section 2.1 we introduce the system model for the MIMO BC, then in Section 2.2 we review the capacity achieving dirty paper coding and describe the suboptimum multiuser eigenmode transmission. Section 2.3 provides numerical comparisons between various state-of-the-art MU MIMO transmission strategies and finally Section 2.4 concludes the chapter summarizing the main findings.

The material in this chapter was partially presented in [38] and published under a more general form in [39].

2.1 System model

As shown in Figure 2.1, the downlink of a cellular network can be modelled as a broadcast channel (BC) where a transmitter with M antennas transmits distinct data signals to K users, each with N antennas. Under a narrowband channel assumption, the baseband received signal by the kth user ($k = 1, \ldots, K$) during time slot n is

$$\mathbf{y}_k(n) = \mathbf{H}_k(n)\mathbf{x}(n) + \mathbf{n}_k(n) \tag{2.1}$$

where $\mathbf{H}_k(n)$ is the kth user's $N \times M$ channel matrix, $\mathbf{x}(n)$ is the M-dimensional transmitted signal vector, and $\mathbf{n}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is the additive white Gaussian noise vector. Note that we assume a block fading model for the channel so that it is static over the time slot. The transmitted signal is a summation of the signals for each user, and in general, these signals could be non-linearly processed. If we denote with $\mathcal{S}(n)$ the set of users selected at the transmitter in time slot n, using linear precoding, the transmitted signal is given by

$$\mathbf{x}(n) = \sum_{k=1}^{|\mathcal{S}(n)|} \mathbf{G}_k(\mathcal{S}(n))\mathbf{d}_k(n), \tag{2.2}$$

where $\mathbf{G}_k(\mathcal{S}(n))$ is the $M \times L_k(n)$ linear precoder matrix, $\mathbf{d}_k(n)$ is the $L_k(n)$-dimensional symbol vector for the kth user and $L_k(n)$ represents the number of streams transmitted to the kth user. The transmit covariance is given by

$$\mathbf{Q}(n) = \sum_k \mathbf{G}_k(\mathcal{S}(n))\mathrm{E}(\mathbf{d}_k(n)\mathbf{d}_k^H(n))\mathbf{G}_k^H(\mathcal{S}(n)) \tag{2.3}$$

and we impose a sum power constraint P among the M base station antennas, i.e.

$$\mathrm{E}[\mathbf{x}^H(n)\mathbf{x}(n)] = \mathrm{tr}(\mathbf{Q}(n)) \leq P\,. \tag{2.4}$$

A generalization of the proposed signal model for a MIMO-OFDM system is given in Chapter 6.

2.2 Multi-user MIMO

MU-MIMO refers to a general class of transmission techniques where multiple users are simultaneously served over common spectral resources during a given transmission interval. In this section we first review the capacity achieving DPC and then describe multiuser eigenmode transmission.

2.2.1 Capacity achieving DPC

MU-MIMO performance is measured using a multidimensional capacity region. Different points within this region can be obtained by changing the transmission strategy, e.g. by changing the partitioning of power among users and adjusting the transmit covariances. For a given

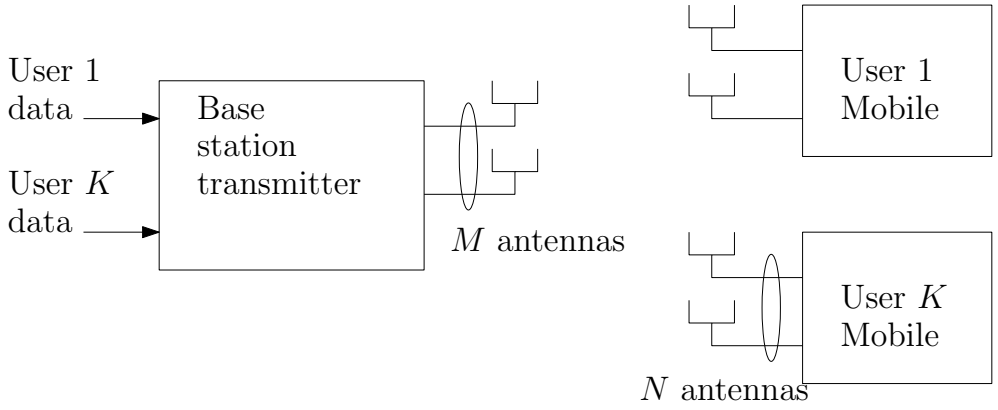

Figure 2.1: System block diagram where K users, each with N receive antennas contend for service from a base station with M transmit antennas.

set of user channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$, which we denote collectively as $\bar{\mathbf{H}}$, and a given power constraint P, we denote by Co the convex hull operation and by π a user ordering. The capacity region of the MIMO BC can be written as

$$\mathcal{C}_{BC}(\bar{\mathbf{H}}, P) = Co\left(\bigcup_{\pi} \mathcal{R}_{\pi}\right) \tag{2.5}$$

where

$$\begin{aligned} \mathcal{R}_{\pi} \quad = \quad & \left\{ \underline{R} = [R_1, \ldots, R_K] : R_{\pi(k)} = \frac{\left|\mathbf{I}_N + \mathbf{H}_{\pi(k)} \sum_{j=k}^{K} \mathbf{Q}_{\pi(j)} \mathbf{H}_{\pi(k)}^H\right|}{\left|\mathbf{I}_N + \mathbf{H}_{\pi(k)} \sum_{j=k+1}^{K} \mathbf{Q}_{\pi(j)} \mathbf{H}_{\pi(k)}^H\right|}, \right. \\ & \left. \sum_{j=1}^{K} \mathrm{tr}\left(\mathbf{Q}_j\right) \leq P \right\}. \end{aligned} \tag{2.6}$$

and $\underline{R}$ denotes a rate vector for the K users.

It was shown recently in [9] that the capacity region (2.5) can be achieved using a coding technique known as DPC [24] which assumes perfect CSI at transmitter and receiver. In DPC users are coded in an ordered fashion such that a given user sees no interference from users encoded before it. In this sense, DPC accomplishes for the downlink at the transmitter what successive interference cancellation does for the uplink at the receiver. During each slot the transmitter determines the point within the capacity region (2.5) which maximizes the weighted sum rate metric

$$\underline{R}_{BC}(\boldsymbol{\alpha}(n), \bar{\mathbf{H}}(n), P) = \underset{\mathbf{r}(n) \in \mathcal{C}_{BC}(\bar{\mathbf{H}}(n), P)}{\arg\max} \sum_{k=1}^{K} \alpha_k(n) R_k(n). \tag{2.7}$$

where $\alpha_k(n)$, $k = 1, \ldots, K$ are the quality of service (QoS) weights. For given $\{\alpha_k\}$ and MIMO channel realizations, the set of rates $R_{BC,1}, \ldots, R_{BC,K}$ that maximizes the metric can be computed numerically [32]. The computation also provides the set of transmit covariance

matrices $\mathbf{Q}_k$ for each of the users. The actual throughput during this interval is then given by the element sum of the rate vector $\underline{R}_{BC}$:

$$R_{BC}(\boldsymbol{\alpha}(n), \bar{\mathbf{H}}(n), P) = \sum_{k=1}^{K} R_{BC,k}(\boldsymbol{\alpha}(n), \bar{\mathbf{H}}(n), P). \tag{2.8}$$

We underline that MU-MIMO has several advantages over time-multiplexed SU-MIMO. Firstly, because the former is a generalization of the latter, the MU-MIMO achievable weighted sum rate will be at least as large as that of SU-MIMO for a given set of channel realizations and QoS weights. Secondly, multiple antennas are required only at the base station, for either receiving or transmitting, because multiplexing gains can be achieved even if users are equipped with only a single antenna. Thirdly, the multiplexing gains in MU-MIMO can be achieved even with highly correlated base station antennas. For example, in line-of-sight channels, the rank deficiency of the MIMO channel matrix for SU-MIMO does not allow for spatial multiplexing (SM). However with multiple users, sufficient spatial separation among the users will ensure low correlation among their spatial channels, resulting in SM gain. A practical consequence of this advantage is that the base station array antennas can be very closely spaced, resulting in a more compact array compared to SU-MIMO.

As discussed at the beginning of the chapter a relative disadvantage of MU-MIMO compared to SU-MIMO is that CSI is required at the transmitter in order to achieve SM across multiple users [16].

2.2.2 Multiuser eigenmode transmission

The high complexity of DPC motivated intensive research activities in the design of suboptimum transmission strategies. In this Section we consider a ZF precoding technique based on BD and denoted as MET, where the number of streams transmitted to each user is chosen to maximize the weighted sum rate metric and where the signal is received by each active user with no interference, [38, 39]. This technique has advantages over other generalized ZF techniques because it provides flexibility in the number of streams transmitted per user. In the following, for ease of notation, we drop the time index n.

Let us fix the set of served users $\mathcal{S}$, and for the kth user select the set of transmitted eigenmodes S_k and assume they are indexed from 1 to $L_k = |S_k|$ The channel of the kth user can be decomposed using the singular value decomposition (SVD) as $\mathbf{H}_k = \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{V}_k^H$, where the eigenvalues in $\boldsymbol{\Lambda}_k$ are arranged so that the ones associated with the allocated streams $\mathcal{S}_k$ of user k appear in the leftmost L_k columns. We denote these eigenvalues as $\Lambda_{k,1}, \ldots, \Lambda_{k,L_k}$ The kth user's receiver is a linear combiner given by the Hermitian transposition of the leftmost L_k columns of $\mathbf{U}_k$ which we denote as $\mathbf{u}_{k,1} \ldots \mathbf{u}_{k,L_k}$. Likewise, we denote the leftmost L_k columns of the right eigenvector matrix $\mathbf{V}_k$ as $\mathbf{v}_{k,1} \ldots \mathbf{v}_{k,L_k}$. From (2.1) and (2.2) the received signal for user k after the linear combiner can be written as

$$\mathbf{r}_k = \left[\mathbf{u}_{k,1} \ldots \mathbf{u}_{k,L_k}\right]^H \mathbf{y}_k \tag{2.9}$$

$$= \boldsymbol{\Gamma}_k \mathbf{G}_k \mathbf{d}_k + \boldsymbol{\Gamma}_k \sum_{j \in \mathcal{S}, j \neq k} \mathbf{G}_j \mathbf{d}_j + \mathbf{n}'_k \tag{2.10}$$

where $\mathbf{\Gamma}_k = [\Lambda_{k,1}\mathbf{v}_{k,1} \ldots \Lambda_{k,L_k}\mathbf{v}_{k,L_k}]^H$ is a $L_k \times M$ matrix and $\mathbf{n}'_k$ is the processed noise. By defining

$$\tilde{\mathbf{H}}_k = \left[\mathbf{\Gamma}_1^H \ldots \mathbf{\Gamma}_{k-1}^H \, \mathbf{\Gamma}_{k+1}^H \ldots \mathbf{\Gamma}_{|\mathcal{S}|}^H\right]^H, \tag{2.11}$$

the zero-forcing constraint requires that the columns of $\mathbf{G}_k$ lie in the null space of $\tilde{\mathbf{H}}_k$. Hence if we consider the SVD of $\tilde{\mathbf{H}}_k$:

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k \tilde{\mathbf{\Lambda}}_k \left[\tilde{\mathbf{V}}_k^{(1)} \; \tilde{\mathbf{V}}_k^{(0)}\right]^H, \tag{2.12}$$

where $\tilde{\mathbf{V}}_k^{(0)}$ corresponds to the right eigenvectors associated with the null modes, the precoding matrix of user k is given by $\mathbf{G}_k = \tilde{\mathbf{V}}_k^{(0)}\mathbf{C}_k$, where $\mathbf{C}_k \in \mathbb{C}^{(M-\sum_{j\in\mathcal{S},j\neq k} L_j)\times L_k}$ is determined later[1].

Note that since $\tilde{\mathbf{H}}_k\tilde{\mathbf{V}}_k^{(0)} = \mathbf{0}$ for all $k \in \mathcal{S}$, it follows that $\mathbf{\Gamma}_k\mathbf{G}_j = \mathbf{\Gamma}_k\tilde{\mathbf{V}}_j^{(0)}\mathbf{C}_j = \mathbf{0}$ for $j \neq k$ and any choice of $\mathbf{C}_j$. Therefore from (2.10), the received signal for the kth user after combining contains no interference:

$$\mathbf{r}_k = \mathbf{\Gamma}_k\mathbf{G}_k\mathbf{d}_k + \mathbf{n}'_k. \tag{2.15}$$

We perform the SVD

$$\mathbf{\Gamma}_k\tilde{\mathbf{V}}_k^{(0)} = \overline{\mathbf{U}}_k \left[\overline{\mathbf{\Lambda}}_k \; \mathbf{0}\right] \left[\overline{\mathbf{V}}_k^{(1)}\overline{\mathbf{V}}_k^{(0)}\right]^H, \tag{2.16}$$

where $\overline{\mathbf{\Lambda}}_k$ is the $L_k \times L_k$ diagonal matrix of eigenvalues, and assign $\mathbf{C}_k = \overline{\mathbf{V}}_k^{(1)}$. From (2.15), the resulting weighted rate for the kth user is

$$\alpha_k \sum_{j\in S_k} \log\left(1 + \overline{\sigma}_j^{(k)^2} w_j^{(k)}\right), \tag{2.17}$$

where $\overline{\sigma}_j^{(k)^2}$ is the jth diagonal element of $\overline{\mathbf{\Lambda}}_k^2$ ($j \in S_k$), α_k is the weighting coefficient for user k, $\mathbf{W}_k$ is the $L_k \times L_k$ diagonal matrix of powers allocated to the eigenmodes, and $w_j^{(k)}$ is the jth diagonal element. Therefore the total transmitted power for user k is $\text{tr}\left(\mathbf{G}_k\mathbf{W}_k\mathbf{G}_k^H\right) = \text{tr}\left(\mathbf{W}_k\right)$. For a given selection of users and eigenmodes $\mathcal{T}$, as determined by $\mathcal{S}$ and S_k, $k \in \mathcal{S}$,

[1] From the relation between the dimension of the null space and rank of $\tilde{\mathbf{V}}_k^{(1)}$, the following constraint has to be satisfied in order to build the set of precoding matrices for the selected users in $\mathcal{S}$:

$$\sum_{j\in\mathcal{S},j\neq k} L_j < M \quad \forall k \in \mathcal{S}. \tag{2.13}$$

From (2.13) the number of modes allocated to the kth user satisfies $L_k \leq M - \sum_{j\in\mathcal{S},j\neq k} L_j$. It follows that the number of allocated modes is upperbounded by the number of transmit antennas: $\sum_{k\in\mathcal{S}} L_k \leq M$. We note that it is possible to allocate all the M modes if the channels are statistically independent. We recall that in the BD scheme [32, 33, 34] the constraint to be satisfied in the construction of the precoding matrices is $\sum_{j\in\mathcal{S},j\neq k} N = N(|S| - 1) < M$, whereas in the BD scheme with receive antenna selection [35] the constraints become less restrictive

$$\sum_{j\in\mathcal{S},j\neq k} N'_j < M \quad \forall k \in \mathcal{S} \tag{2.14}$$

where $N'_k \leq N$ is the number of receive antennas selected for the kth user. We note that (2.13) is similar to (2.14) except that instead of using a subset of receive antennas we use a subset of eigenmodes.

INIT: $\mathcal{T}_0 \leftarrow \emptyset$, $\tilde{R}(\mathcal{T}_0) \leftarrow 0$
for $n = 1$ to $\max(KN, M)$ **do**

$$t_n \leftarrow \underset{t \in \mathcal{T}_A \setminus \mathcal{T}_{n-1}}{\arg\max} \tilde{R}(\mathcal{T}_{n-1} \cup \{t\}) \tag{2.19}$$

  if $\tilde{R}(\mathcal{T}_{n-1} \cup \{t_n\}) < \tilde{R}(\mathcal{T}_{n-1})$ **then**
    $\mathcal{T} \leftarrow \mathcal{T}_{n-1}$
    break
  else
    $\mathcal{T}, \mathcal{T}_n \leftarrow \mathcal{T}_{n-1} \cup \{t_n\}$
  end if
end for

Table 2.1: Pseudo-code of greedy eigenmode selection algorithm .

the power allocation problem under sum power constraint can be written as

$$\tilde{R}(\mathcal{T}) = \max_{w_j^{(k)}, k \in \mathcal{S}, j \in S_k} \sum_{k \in \mathcal{S}} \alpha_k \sum_{j \in S_k} \log\left(1 + \overline{\sigma}_j^{(k)^2} w_j^{(k)}\right) \tag{2.18}$$

$$\text{subject to} \quad \begin{cases} w_j^{(k)} \geq 0, \quad k \in \mathcal{S}, j \in S_k \\ \sum_{k \in \mathcal{S}} \sum_{j \in \mathcal{S}_k} w_j^{(k)} \leq P \end{cases}$$

and the resulting optimization can be solved using waterfilling.

We emphasize that optimization (2.18) is performed for a given user and eigenmode allocation. The allocation itself could be performed in a brute-force manner by considering all possible sets of up to M eigenmodes. Due to the high computational complexity of the brute-force case (see [40]) we propose a generalization of the greedy allocation algorithm proposed in [21]. We define $\mathcal{T}_A$ to be the set of all K users' eigenmodes. Assuming $N < M$, each user has at most N eigenmodes, and there are a total of KN eigenmodes in set $\mathcal{T}_A$. On the nth iteration, let t_n be the candidate eigenmode chosen among any of the available eigenmodes from any user. The eigenmode t_n is added to the set of active eigenmodes only if the weighted throughput increases. The proposed greedy algorithm is summarized in Tab. 2.1.

Let's consider as a special case the maximization of the sum rate, i.e. all users have the same QoS weights $\alpha_k = 1$, $k = 1, \ldots, K$. On the first iteration, the selected eigenmode t_1 will be the globally dominant eigenmode. In other words, its eigenvalue is the largest among all users' modes. Note however that the chosen set $\mathcal{T}$ will not necessarily contain the dominant eigenmodes of each user. Note also that not all eigenmodes will necessarily be active. Numerical examples in Section 6.5 show the distribution of allocated eigenmodes. Even if this greedy algorithm is suboptimum, we show in Section 2.3 that it achieves a good balance between performance and complexity. Moreover it is also totally flexible, as it can handle any combination of M, K, and N.

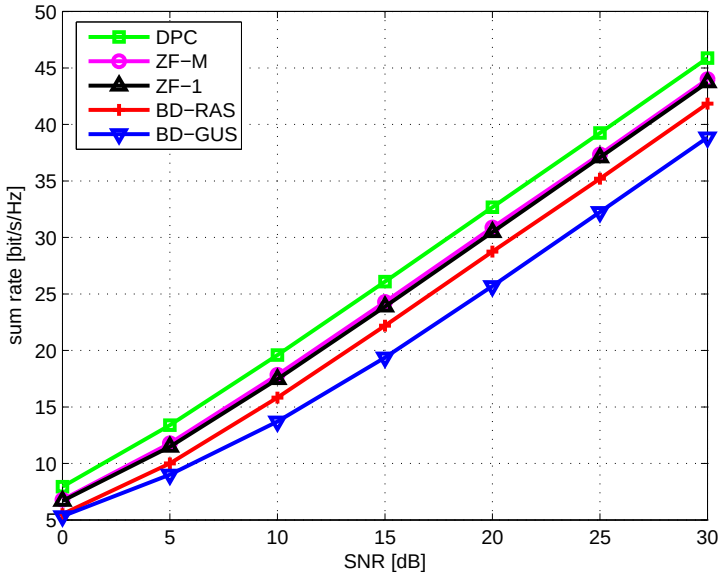

Figure 2.2: Sum rate versus average SNR of each user, for $N = M = 4$ antennas and $K = 20$ users.

2.3 Simulation results

In this section, we assume an independent and identically distributed complex Gaussian channel ($[\mathbf{H}_k]_{(i,j)} \sim \mathcal{CN}(0,1)$) where the channel matrix $\mathbf{H}_k$ is assumed to be perfectly known both at the transmitter and at the kth receiver. In Figure 2.2 we set $M = 4$, $N = 4$ and $K = 20$ and compare the optimum DPC, BD and MET in terms of average sum rate versus average user SNR. We consider two types of MET: i) the general MET which allows the selection of multiple streams per user (denoted as ZF-M) and ii) a special version of MET where each active user can be served only along its dominant eigenmode (denoted as ZF-1). We consider two types of BD as well, each using greedy user selection (GUS) [37]. Under BD-GUS, each selected user employs all N antennas. Under BD-RAS we use a modified version of GUS where each candidate user selects the best subset of N receive antennas [35]. Interestingly ZF-M gives the best performance among the linear beamforming options, but for a moderate number of users, e.g. $K = 20$, ZF-1 is already very close to the more general ZF-M. For SNR=10 dB ZF-M achieves about 90% of the DPC sum rate.

In Figure 2.3, we investigate the probability of mode selection for an active user, for $M = 4, 12$, $N = 4$, $K = 5, 20$, and different values of SNR. We use the greedy user and eigenmode selection algorithm described in Tab. 2.1 and consider the following three categories: i) only the dominant eigenmode is active as in ZF-1 (eigenmode 1), ii) only one non-dominant eigenmode (eigenmode i with $i \geq 2$) is active and iii) multiple eigenmodes are active. We note that when the ratio M/K is small, the probability of transmitting on only the dominant eigenmode for an active user is very high. On the other hand, we note that in general the probability of selecting a non-dominant mode or multiple modes for the same user, is not small. We emphasize that for a given transmission interval, MET chooses the best option between:

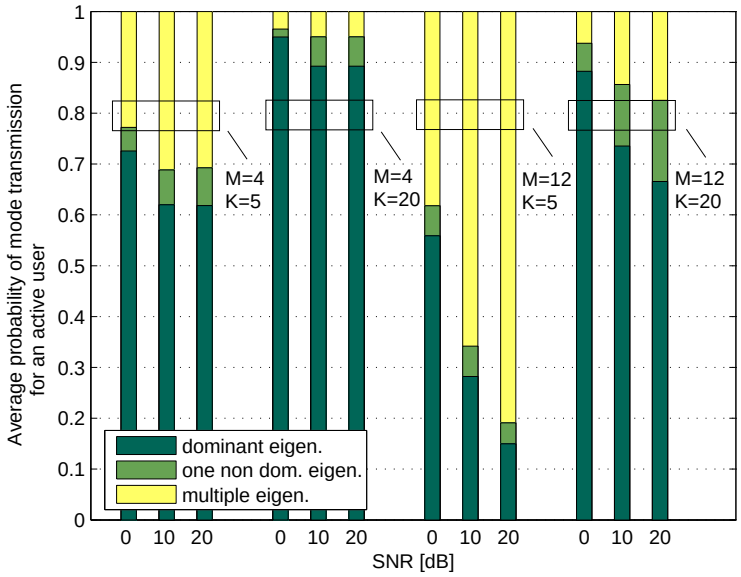

Figure 2.3: Probability of mode transmission for an active user, with $M = 4, 12$, $N = 4$, $K = 5, 20$, and different values of average SNR.

i) spatially multiplexing different independent streams to only one user (SU MIMO with spatial multiplexing), ii) spatially multiplexing independent streams to different users (multiuser MIMO with rank 1 transmission to each active user) or iii) a hybrid solution where some users are served with rank 1 transmission whereas other users receive multiple independent streams. In other words, any "mode switching" between SU and MU MIMO is performed automatically by the greedy eigenmode selection algorithm. We also observe that in a multiuser scenario allocating the dominant eigenmodes as done in [33] and [36], or selecting the users without considering the problem of the eigenmode allocation [37] are clearly suboptimum policies.

In Figure 2.4 we compare the performance of general ZF-M and ZF-1. For ZF-1 power allocation among eigenmodes is performed using waterfilling as in ZF-M, however, it can be shown that a simplified algorithm which allocates equal power across modes results in nearly identical performance, especially at high SNR. We note that when the ratio K/M is small, the gap between the two considered schemes is not negligible, whereas when K/M is large and multiuser diversity is exploitable, the gap decreases.

Finally in Fig. 2.5 we set $M = N = 4$, SNR $= 15$ dB and show the achievable sum rate of ZF-M, ZF-1, DPC, BD-GUS and BD-RAS as a function of the number of users K. It is interesting to observe how ZF-M outperforms all other linear precoding strategies and achieve a sum rate very close to the upper bound given by DPC even for moderate K. Again ZF-1 provides almost the same performance of MET for K/M large enough. We notice that even if the sum rate of BD-RAS has been shown to scale optimally when the number of users goes to infinity [39], for a practical number of users it has a significant gap with both ZF-M and DPC.

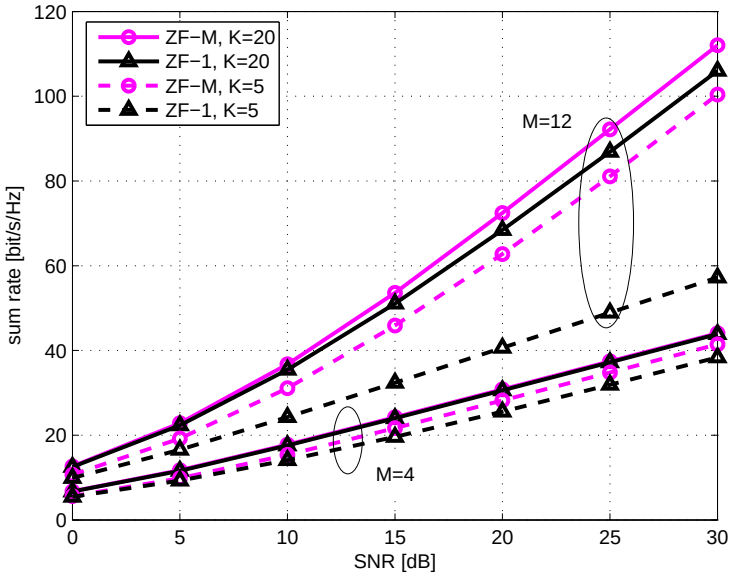

Figure 2.4: Comparison between ZF-M and ZF-1 with water-filling power allocation for $M = 4, 12$*,* $N = 4$ *and* $K = 5, 20$*.*

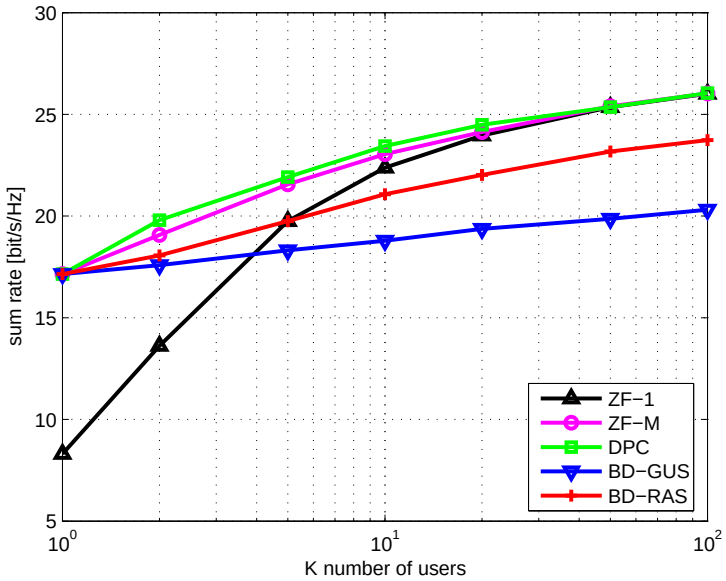

Figure 2.5: Sum rate versus number of users K *for* $M = N = 4$ *and* $SNR = 15$ *dB.*

2.4 Conclusions

Multiuser eigenmode transmission (MET) is a multiuser MIMO technique that uses linear zero-forcing beamforming to flexibly allocate multiple data streams to multiple users. Mode switching between single-user and multiuser transmission is based on quality of service weights and

channel measurements and is performed on a frame-by-frame basis to maximize the weighted sum rate. MET has the flexibility of transmitting spatially multiplexed streams to a single user, however in moderately loaded systems, its strategy for maximizing the sum rate typically results on transmitting a single stream to multiple users. Numerical results reveal that MET outperforms other state-of-the-art linear beamforming techniques and perform very close to the optimum DPC in most practical network conditions.

Chapter 3

Multiuser MIMO downlink with limited feedback and single antenna receivers

As explained in Chapter 2, multiuser MIMO downlink systems require CSIT to achieve spatial multiplexing across users. Differently from TDD systems where channel knowledge can be acquired from channel estimation in the uplink, in FDD systems CSIT can be obtained only by setting up an explicit feedback (FB) channel from each user. As the number of bits required to describe the channels grows as the product of the number of transmit and receive antennas, the channel delay spread and the number of users [41], only by a proper optimization of the FB signalling its impact on the network throughput can be limited.

Recently, this problem has received a lot of attention and various aspects have been investigated including transmitter and receiver design [42], [43], [44] and feedback optimization in both SU and MU systems [45], [46], [47] [48]. Typically, the FB bits are used to index a set of vectors (or codewords) in a codebook $\mathcal{C}$ which is known to the transmitter and all receivers. For example, B bits per feedback interval can be used to index a codebook with 2^B vectors. For a transmitter, each codeword in $\mathcal{C}$ is a multi-dimensional vector that characterizes the MIMO channel for that user or more generally provides information on the reconstruction of the user's channel. A well-designed codebook will contain codewords that effectively span the set of MIMO channels experienced by the users [46], [47]. In SU systems it has been shown that only a few FB bits (roughly on the order of transmit antennas) are needed to achieve near perfect-CSIT performance. Differently in downlink channels accurate channel knowledge is essential to avoid multiuser interference and a severe degradation of the achievable throughput [42], [49]. As with perfect CSIT, an opportunistic user selection approach can increase the performance of this systems leading to asymptotically optimum performance when the number of users goes to infinity [45],[42],[44] .

In this chapter we focus on single-antenna mobile terminals and investigate three different problems: i) beamformer design, ii) channel quantization and feedback signalling optimization and iii) user selection. Firstly, we revise ZF beamforming and propose a new MMSE beamformer under incomplete CSIT that takes into account the quantization error of the channel

vector. Recalling a result known under perfect CSIT, MMSE BF shows significant performance improvements in case of randomly selected users but gives reduced gains with respect to ZF BF in case of opportunistic user selection. Secondly, we propose various channel quantization techniques and feedback strategies based on the Lloyd-Max algorithm [20] that exploit both spatial and time correlation of the MIMO channel. In particular we derive a hierarchical FB (HFB) approach where FB bits are accumulated over multiple signalling intervals in order to index a much larger codebook. Moreover we propose new predictive FB strategies where both transmitter and users predict the evolution of the channel vector and users adjusts the prediction by feeding back a quantized version of the prediction error to the transmitter. Finally we propose two greedy user selection algorithms based on [21] that rely on users FB and show improved performance with respect to state-of-the-art algorithms. Numerical results provide an useful comparison between the proposed schemes.

The chapter is organized as follows. Section 3.1 introduces the system model for a MIMO BC with limited uplink feedback and Section 3.2 derives ZF and MMSE beamformers, providing for both techniques tight approximations of the users' estimated achievable signal-to-interference plus noise ratio (SINR). Then, Section 3.3 addresses the problems of channel quantization and optimization of the FB signalling, while Section 3.4 propose greedy user selection algorithms for the maximization of the achievable throughput. Numerical comparisons of the proposed techniques are given in Section 3.5 and Section 3.6 concludes the chapter summarizing the main findings.

Some of the material in this chapter has been published in [50], [51], [52] and [53]. Moreover part of the results was presented at the 3GPP long term evolution (LTE) meetings [54], [55].

3.1 System model

We consider the downlink of a cellular system where a transmitter has M antennas and K users have one antenna each, i.e. $N = 1$. Transmission is performed in time slots of size T and in each time slot users feed back a partial CSI, which is used by the transmitter to schedule downlink transmissions and design the beamformer. The transmitted signal $\mathbf{x}$ and the signal received by user k, $\mathbf{y}_k$, are modelled as in (2.2) and (2.1), respectively, with $L_k = 1$ for each selected user. Since $N = 1$ we denote the channel and beamformer of user k as the $1 \times M$ vector $\mathbf{h}_k(n)$ and the $M \times 1$ vector $\mathbf{g}_k(n)$, respectively. From (2.2) and (2.1) the achievable SINR for user k is given by

$$\text{SINR}_k(n) = \frac{|\mathbf{h}_k(n)\mathbf{g}_k(n)|^2}{1 + \sum\limits_{i \in \mathcal{S}(n) \setminus \{k\}} |\mathbf{h}_k(n)\mathbf{g}_i(n)|^2} \ . \tag{3.1}$$

Under Gaussian codes and minimum-distance decoding, the achievable rate for user k is given by

$$R_k(n) = \log(1 + \text{SINR}_k(n)) \ . \tag{3.2}$$

The transmission strategy we propose develops in three phases: i) users channel estimation

and FB of channel information to the transmitter, ii) user selection and beamformer design at the transmitter and iii) users data demodulation.

In the first phase each user perfectly measures its channel vector $\mathbf{h}_k(n)$ once a slot, based on pilot signals transmitted on each of the M antennas. In particular, as in [42], we assume that each user estimates two quantities: $i)$ the channel direction information (CDI), namely

$$\tilde{\mathbf{h}}_k(n) = \frac{\mathbf{h}_k(n)}{||\mathbf{h}_k(n)||} \tag{3.3}$$

and $ii)$ a channel quality information (CQI) related to the user's achievable rate or equivalently its SINR. Each user feeds back to the transmitter a quantized version of the CDI and the unquantized CQI assuming a zero-delay and error free uplink control channel. The CDI FB consists of B bits per slot, used at the transmitter to reconstruct user's channel vector. The channel reconstruction algorithm depends on the FB strategy adopted at receivers. For instance, for the basic FB (BFB) strategy (see also Section 3.3.1) the channel direction is quantized according to minimal chordal distance [42] using a codebook with 2^B unit-norm codewords. In this case the index of the best codeword is fed back to the transmitter as quantized CDI and the reconstructed channel is simply the best codeword. More details about the proposed FB strategies are given in Section 3.3. We note that the unit-norm reconstructed channel vectors of all users are stored at the transmitter into the matrix

$$\bar{\mathbf{H}} = [\bar{\mathbf{h}}_1(n)^T, \ldots, \bar{\mathbf{h}}_K(n)^T]^T \, . \tag{3.4}$$

In phase 2, the transmitter uses the CQI and CDI information from all receivers to determine the set of active users $\mathcal{S}$ and the precoding vectors $\mathbf{g}_k, k \in \mathcal{S}$ for each data stream d_k. In phase 3, each user in $\mathcal{S}$ estimates its equivalent channel using dedicated pilots [49] and demodulates the data.

We notice that the actual achievable rate of a served user in phase 3 is a function of its MISO channel, the transmit beamforming weights and the residual multiuser interference. During phase 1, a user does not have *a priori* knowledge of the CDI vectors for the other users in $\mathcal{S}$, and hence it cannot know what the beamforming weights or interference will be. We therefore propose to use the *expected* SINR as the CQI by making a judicious prediction on the interference statistics. We notice that in cellular standards, the quantization of CQI feedback uses a large number of bits in order to closely match the channel quality with a fine granularity of modulation and coding options. We use this observation to justify our assumption that the CQI is an unquantized analog value. This is a typical assumption found in literature [42, 49].

We notice that the transmitter does not have an exact knowledge of the achievable rate (3.2) which could be obtained only after a second uplink FB from users after phase 3. On the other hand, practical systems such as 1xEV-DO [56] make use of fast incremental redundancy coding in order to cope with residual channel uncertainty, i.e. the effective coding rate is adapted such that, eventually, it is slightly less than (3.2) even though the latter is not known. As in [42, 43, 49] we optimistically assume that thanks to fast incremental redundancy or some other higher level medium access protocol (3.2) is achievable.

3.2 Beamformer design

In this section we briefly review ZF BF and derive a new MMSE BF under incomplete CSI assumptions. For ease of notation we drop the slot index n.

3.2.1 Zero-forcing beamforming

Let us denote with $\bar{\mathbf{H}}(\mathcal{S})$ the matrix containing as rows the reconstructed channel vectors of the selected users. By denoting with $\mathbf{W}(\mathcal{S}) = \bar{\mathbf{H}}(\mathcal{S})^{\dagger}$ the right pseudo-inverse of $\bar{\mathbf{H}}(\mathcal{S})$ the ZF transmit matrix is given by

$$\begin{aligned}\mathbf{G}(\mathcal{S}) &= \mathbf{W}(\mathcal{S})\text{diag}(\mathbf{p})^{1/2} \\ &= \bar{\mathbf{H}}(\mathcal{S})^{H}\left(\bar{\mathbf{H}}(\mathcal{S})\bar{\mathbf{H}}(\mathcal{S})^{H}\right)^{-1}\text{diag}(\mathbf{p})^{1/2}\ ,\end{aligned} \tag{3.5}$$

where $\mathbf{p}$ is the vector of power normalization coefficients imposing the power constraint P on the transmitted signal. Under the assumption of equal power distribution across users, $\mathbf{p}$ has elements[1]

$$p_k = \frac{P}{|\mathcal{S}| \cdot ||\mathbf{w}_k||^2}\ , \tag{3.6}$$

where $\mathbf{w}_k$ is the k-th column of $\mathbf{W}(\mathcal{S})$. We recall that, by construction $\mathbf{w}_k$ is orthogonal to $\bar{\mathbf{h}}_i^H$ for $i \in \mathcal{S} \setminus \{k\}$ and $|\bar{\mathbf{h}}_k\mathbf{w}_k| = 1$ for every k. We also note that the computation of $\mathbf{G}(\mathcal{S})$ requires only the CDI feedback from the terminals.

CQI feedback

As explained in Section 3.4 the proposed user selection algorithms require that the transmitter can estimate the achievable user rates (3.2), i.e. the SINRs (3.1). In the following we derive various approximations of the expected SINR. We emphasize that the analysis we develop is based on the following assumptions: i) Rayleigh fading channels with i.i.d. elements $\sim \mathcal{CN}(0,1)$ ii) basic FB strategy iii) quantization error vector isotropically distributed in the hyperspace orthogonal to the selected codeword iv) independent codebooks for different users. As a consequence it strictly holds for random vector quantization (RVQ) [42] (see also Section 3.3.1) but it also provides a good approximation for BFB in case of well-designed codebooks and large B. Numerical simulations revealed that the proposed analysis gives a tight approximation even for other FB strategies explained in Section 3.3.

Let us define the angle $\theta_k \in [0, \pi/2]$ between complex vectors $\tilde{\mathbf{h}}_k$ and $\bar{\mathbf{h}}_k$, such that

$$\cos\theta_k = |\tilde{\mathbf{h}}_k\bar{\mathbf{h}}_k^H|\ . \tag{3.7}$$

One naïve approach is for the transmitter to assume that there is no quantization error in the CDI report, such that $|\mathbf{h}_k\mathbf{w}_i| \approx 0$ for $i \in \mathcal{S} \setminus \{k\}$, and $|\mathbf{h}_k\mathbf{w}_k| \approx ||\mathbf{h}_k||\cos\theta_k$. In other

[1] This last restriction is clearly sub-optimum but allows to derive a good approximation of the expected SINR of each user that otherwise would be difficult to predict. Moreover, as verified in Chapter 2 under perfect CSIT, there is marginal loss in achievable throughput when considering ZF with equal power allocation instead of optimum water-filling power allocation.

words, the channels $\mathbf{h}_k$ are approximated with the projection along their quantized direction. With this assumption (3.1) can be roughly approximated as

$$\text{SINR}_k \approx p_k ||\mathbf{h}_k||^2 \cos^2 \theta_k \triangleq \gamma_k^{(ZF,1)} . \tag{3.8}$$

As p_k are known to the transmitter, whereas $\mathbf{h}_k$ is perfectly known only at the receiver, the required CQI feedback from the terminals would be $g(\mathbf{h}_k) = ||\mathbf{h}_k||^2 \cos^2 \theta_k$.

One better approximation can be derived by taking into account the quantization error, defined as $\mathbf{e}_k = \tilde{\mathbf{h}}_k - (\tilde{\mathbf{h}}_k \bar{\mathbf{h}}_k^H)\bar{\mathbf{h}}_k$, with square norm $||\mathbf{e}_k||^2 = \sin^2 \theta_k$. Let us introduce the unit-norm vectors $\tilde{\mathbf{e}}_k = \mathbf{e}_k / ||\mathbf{e}_k||$ and $\tilde{\mathbf{w}}_k = \mathbf{w}_k / ||\mathbf{w}_k||$. Then, by decomposing $\tilde{\mathbf{h}}_k$ as

$$\tilde{\mathbf{h}}_k = (\tilde{\mathbf{h}}_k \bar{\mathbf{h}}_k^H)\bar{\mathbf{h}}_k + \mathbf{e}_k \tag{3.9}$$

(3.1) can be rewritten as

$$\text{SINR}_k = \frac{\frac{P}{|\mathcal{S}|}||\mathbf{h}_k||^2 \left|(\tilde{\mathbf{h}}_k \bar{\mathbf{h}}_k^H)(\bar{\mathbf{h}}_k \tilde{\mathbf{w}}_k) + \mathbf{e}_k \tilde{\mathbf{w}}_k\right|^2}{1 + \frac{P}{|\mathcal{S}|}||\mathbf{h}_k||^2 \sin^2 \theta_k \sum\limits_{i \in \mathcal{S} \setminus \{k\}} |\tilde{\mathbf{e}}_k \tilde{\mathbf{w}}_i|^2} \tag{3.10}$$

Let us focus on the term at the numerator. Say $\phi_k \in [0, \pi/2]$ the angle between vectors $\bar{\mathbf{h}}_k$ and $\tilde{\mathbf{w}}_k$, which is in general non-zero[2], then by construction of the ZF beamformer

$$\cos \phi_k = |\bar{\mathbf{h}}_k \tilde{\mathbf{w}}_k| = \frac{1}{||\mathbf{w}_k||} . \tag{3.11}$$

Moreover, if ϕ_k is small enough, i.e. the users selected for transmissions have nearly orthogonal reported channels, the k-th error vector is orthogonal to the k-th beamforming vector such that we can approximate

$$\mathbf{e}_k \tilde{\mathbf{w}}_k \approx 0 . \tag{3.12}$$

We now focus on the sum at the denominator of (3.10). The unit vectors $\tilde{\mathbf{e}}_k$ and $\tilde{\mathbf{w}}_i$ are both isotropically distributed on the $(M-1)$-dimensional hyperplane orthogonal to $\bar{\mathbf{h}}_k$. Moreover, as the directional distribution of $\tilde{\mathbf{w}}_i$ on this hyperplane depends only on $\bar{\mathbf{h}}_j$ for $j \in \mathcal{S} \setminus \{k, i\}$, it follows that $\tilde{\mathbf{w}}_i$ is independent of the quantization error $\tilde{\mathbf{e}}_k$, for any $i \neq k$. Hence, the inner product $|\tilde{\mathbf{e}}_k \tilde{\mathbf{w}}_i|$ is Beta-distributed with parameters $(1, M-1)$, and mean value $1/(M-1)$. By taking the expectation of (3.10) w.r.t. the interference term and noticing that the SINR function is monotonic with this term, Jensen's inequality yields the following lower-bound

$$\begin{aligned} \text{E}\left[\text{SINR}_k\right] &\geq \frac{\frac{P}{|\mathcal{S}|}||\mathbf{h}_k||^2 \left|(\tilde{\mathbf{h}}_k \bar{\mathbf{h}}_k^H)(\bar{\mathbf{h}}_k \tilde{\mathbf{w}}_k) + \mathbf{e}_k \tilde{\mathbf{w}}_k\right|^2}{1 + \frac{P}{|\mathcal{S}|}||\mathbf{h}_k||^2 \sin^2 \theta_k E\left[\sum\limits_{i \in \mathcal{S} \setminus \{k\}} |\tilde{\mathbf{e}}_k \tilde{\mathbf{w}}_i|^2\right]} \\ &= \frac{\frac{P}{|\mathcal{S}|}||\mathbf{h}_k||^2 \left|(\tilde{\mathbf{h}}_k \bar{\mathbf{h}}_k^H)(\bar{\mathbf{h}}_k \tilde{\mathbf{w}}_k) + \mathbf{e}_k \tilde{\mathbf{w}}_k\right|^2}{1 + \frac{P}{|\mathcal{S}|}\frac{|\mathcal{S}|-1}{M-1}||\mathbf{h}_k||^2 \sin^2 \theta_k} \end{aligned} \tag{3.13}$$

[2] $\phi_k = 0$ only if $\bar{\mathbf{h}}_k$ is orthogonal to $\bar{\mathbf{h}}_i$ for all $i \in \mathcal{S} \setminus \{k\}$.

Feedback methods for ZF precoder		
Method 1		
$g(\mathbf{h}_k)$		SINR$_k$ approximation
$\|\|\mathbf{h}_k\|\|^2 \cos^2\theta_k$		$\gamma_k^{(ZF,1)} = p_k g(\mathbf{h}_k)$
Method 2		
$g(\mathbf{h}_k)$		E [SINR$_k$] lower bound
$\frac{\frac{P}{M}\|\|\mathbf{h}_k\|\|^2 \cos^2\theta_k}{1+\frac{P}{M}\|\|\mathbf{h}_k\|\|^2 \sin^2\theta_k}$		$\gamma_k^{(ZF,2)} = \frac{M}{\|\mathcal{S}\|\|\mathbf{w}_k\|^2} g(\mathbf{h}_k)$
Method 3		
$g_1(\mathbf{h}_k)$	$g_2(\mathbf{h}_k)$	E [SINR$_k$] lower bound
$\|\|\mathbf{h}_k\|\|^2$	$\cos^2\theta_k$	$\gamma_k^{(ZF,3)} = \frac{p_k\|\|\mathbf{h}_k\|\|^2 \cos^2\theta_k}{1+\frac{P}{\|\mathcal{S}\|}\frac{\|\mathcal{S}\|-1}{M-1}\|\|\mathbf{h}_k\|\|^2 \sin^2\theta_k}$

Table 3.1: Feedback methods and SINR estimation methods.

From (3.13), using (3.11) and the approximation (3.12) we get the following

$$\mathrm{E}\left[\mathrm{SINR}_k\right] \gtrapprox \frac{p_k||\mathbf{h}_k||^2 \cos^2\theta_k}{1+\frac{P}{|\mathcal{S}|}\frac{|\mathcal{S}|-1}{M-1}||\mathbf{h}_k||^2 \sin^2\theta_k} \triangleq \gamma_k^{(ZF,3)}. \tag{3.14}$$

As the cardinality of $\mathcal{S}$ is unknown to the terminals, the only way for the transmitter to compute (3.14) is by having the terminals report the square amplitude of the channel, $||\mathbf{h}_k||^2$, and the square amplitude of the quantization error, $||\mathbf{e}_k||^2$ (or, equivalently, $\cos^2\theta$) *separately*. This implies that each terminal has to send *two* CQI values, thus if the CQI are to be quantized with a finite number of bits, the precision of the reported CQI's is necessarily reduced compared to the single CQI case.

One way of reducing the CQI report to one value is by further lower-bounding (3.14), by noticing that $(i-1)/i \leq (M-1)/M$ for all $i \leq M$. Therefore, from (3.14) we get

$$\mathrm{E}\left[\mathrm{SINR}_k\right] \gtrapprox \frac{p_k||\mathbf{h}_k||^2 \cos^2\theta_k}{1+\frac{P}{M}||\mathbf{h}_k||^2 \sin^2\theta_k} \triangleq \gamma_k^{(ZF,2)}. \tag{3.15}$$

It is straightforward to show that $\gamma_k^{(ZF,2)}$ represents the exact SINR of the kth receiver when the CDIs of the selected users form a set of M orthogonal vectors. In fact in this case the precoding vectors are simply the CDIs of the selected users reconstructed by the transmitter, i.e. $\mathbf{w}_k = \bar{\mathbf{h}}_k$, the error vector $\mathbf{e}_k$ becomes strictly orthogonal to the kth precoding vector, i.e. $\mathbf{e}_k^H \mathbf{w}_k = 0$, and $\sum_{i\in\mathcal{S}\setminus\{k\}} |\tilde{\mathbf{e}}_k^H \tilde{\mathbf{w}}_i|^2 = ||\tilde{\mathbf{e}}_k^H||^2 = 1$. Under this simplification, (3.10) reduces to (3.15).

In Table 3.1, we summarize the CQI expressions and the SINR estimates of these three different approaches. The estimated throughput achieved by using the ith method is given by

$$R^{(ZF,i)}(\mathcal{S}) = \sum_{k\in\mathcal{S}} \log_2\left(1+\gamma_k^{(ZF,i)}\right) \tag{3.16}$$

where $\gamma_k^{(ZF,i)}$, $i = 1, 3, 2$ are defined in (3.8), (3.14) and (3.15), respectively.

3.2.2 MMSE beamforming

Differently from ZF BF, MMSE BF aims at minimizing the sum mean square error (MSE) of the received signals. To this end, we first decompose the channel vector relative to user k into two orthogonal vectors $\mathbf{f}_k$ and $\boldsymbol{\epsilon}_k$, parallel and orthogonal to $\bar{\mathbf{h}}_k$, respectively, with

$$\mathbf{h}_k = ||\mathbf{h}_k|| \left(\mathbf{f}_k + \boldsymbol{\epsilon}_k\right) , \tag{3.17}$$

where $\mathbf{f}_k = \cos\theta_k \bar{\mathbf{h}}_k$ and we recall $\cos\theta_k = |\tilde{\mathbf{h}}_k \bar{\mathbf{h}}_k^H|$. Let also define $\mathbf{F} = [\mathbf{f}_1^T, \ldots, \mathbf{f}_{|\mathcal{S}|}^T]^T$ and $\mathbf{E} = [\boldsymbol{\epsilon}_1^T, \ldots, \boldsymbol{\epsilon}_{|\mathcal{S}|}^T]^T$. We assume that user k divides the received signal by $\beta||\mathbf{h}_k||$, where β is a power normalization coefficient. In this case, by defining $\mathbf{N} = \text{diag}\left(||\mathbf{h}_1||, \ldots, ||\mathbf{h}_{|\mathcal{S}|}||\right)$, the normalized received signal can be written as

$$\mathbf{y}' = \beta^{-1} \left(\mathbf{F} + \mathbf{E}\right) \mathbf{G}(\mathcal{S})\mathbf{d} + \beta^{-1}\mathbf{N}^{-1}\mathbf{n} \, . \tag{3.18}$$

The problem to be solved for linear MMSE-BF design is the joint optimization of $\mathbf{G}(\mathcal{S})$ and β in order to minimize the MSE $\mathrm{E}\left[\left|\left|\mathbf{y}' - \mathbf{d}\right|\right|^2\right]$ under the sum power constraint, i.e.

$$\mathbf{G}^{(MMSE)}(\mathcal{S}) = \arg\min_{\mathbf{G},\beta} \mathrm{E}\left[\left|\left|\mathbf{y}' - \mathbf{d}\right|\right|^2\right] \tag{3.19a}$$

$$\mathrm{E}[||\mathbf{G}\mathbf{d}||^2] \leq P \tag{3.19b}$$

We notice that, differently from ZF-BF, in (3.19) we are not imposing equal power allocation among the selected users. Moreover the expectation in (3.19a) is taken with respect to data, noise and the direction of the error vectors $\boldsymbol{\epsilon}_k$, while from (3.17) we observe that $||\boldsymbol{\epsilon}_k||^2 = \sin^2(\theta_k)$. The solution of (3.19) is provided in Theorem 1 whose proof is given in Appendix A.1

Theorem 1 *Let us define the normalized matrix*

$$\bar{\mathbf{G}} = \left[\mathbf{F}^H\mathbf{F} + \mathbf{R} + \frac{\sigma_N^2}{P}\mathbf{I}\right]^{-1} \mathbf{F}^H \tag{3.20}$$

where $\mathbf{R} = \mathrm{E}[\mathbf{E}^H\mathbf{E}]$ *and* $\sigma_N^2 = \sum_{i\in\mathcal{S}} \frac{1}{||\mathbf{h}_i||^2}$. *The minimizing* β *in (3.19) is given by*

$$\beta = \sqrt{\frac{P}{tr\left(\bar{\mathbf{G}}^H\bar{\mathbf{G}}\right)}} \tag{3.21}$$

which leads to the MMSE-BF

$$\mathbf{G}^{(MMSE)} = \beta\bar{\mathbf{G}} \, . \tag{3.22}$$

□

The error correlation matrix $\mathbf{R}$ can be computed numerically as a function of the channel quantization codebook. In Lemma 1 we characterize $\mathbf{R}$ under realistic assumptions for the

channel quantization error. The proof of Lemma 1 is given in Appendix A.2.

Lemma 1 *Let us assume that $\boldsymbol{\epsilon}_k$ are statistically uncorrelated and that the unit-norm vector $\tilde{\boldsymbol{\epsilon}}_k = \boldsymbol{\epsilon}_k/||\boldsymbol{\epsilon}_k||$ assumes all directions orthogonal to $\bar{\mathbf{h}}_k$ with equal probability. We have*

$$\mathbf{R} = \mathrm{E}[\mathbf{E}^H\mathbf{E}] = \sum_{k=1}^{|\mathcal{S}|} \sin^2(\theta_k)\mathbf{A}_k^H\boldsymbol{\Xi}\mathbf{A}_k\,, \tag{3.23}$$

where $\boldsymbol{\Xi}$ is a diagonal matrix with entries

$$[\boldsymbol{\Xi}]_{p,p} = \frac{1}{2^p}\,, \quad p < M-1\,, \quad [\boldsymbol{\Xi}]_{M-1,M-1} = \frac{1}{2^{M-2}}\,, \tag{3.24}$$

$\mathbf{A}_k$ is an $(M-1)\times M$ matrix having as rows a base of the space orthogonal to $\bar{\mathbf{h}}_k$ and

$$\mathrm{E}[\tilde{\boldsymbol{\epsilon}}_k^H\tilde{\boldsymbol{\epsilon}}_k] = \mathbf{A}_k^H\boldsymbol{\Xi}\mathbf{A}_k\,. \tag{3.25}$$

□

For MMSE-BF the SINR relative to user k can be written as

$$\mathrm{SINR}_k = \frac{||\mathbf{h}_k||^2|\left(\bar{\mathbf{h}}_k\cos\theta_k + \tilde{\boldsymbol{\epsilon}}_k\sin\theta_k\right)\mathbf{g}_k|^2}{1+||\mathbf{h}_k||^2\sum_{i\neq k}|\left(\bar{\mathbf{h}}_k\cos\theta_k + \tilde{\boldsymbol{\epsilon}}_k\sin\theta_k\right)\mathbf{g}_i|^2}\,. \tag{3.26}$$

Neglecting the second term in the numerator of (3.26), i.e., $\boldsymbol{\epsilon}_k\mathbf{g}_k \simeq 0$, and taking the expectation with respect to the interference term in the denominator of (3.26), we obtain

$$\gamma_k^{(MMSE)} = \frac{||\mathbf{h}_k||^2\cos^2\theta_k|\bar{\mathbf{h}}_k\mathbf{g}_k|^2}{1+||\mathbf{h}_k||^2\cos^2\theta_k\sum_{i\neq k}|\bar{\mathbf{h}}_k\mathbf{g}_i|^2 + ||\mathbf{h}_k||^2\sin^2\theta_k\sum_{i\neq k}\mathbf{g}_i^H\mathrm{E}\left[\tilde{\boldsymbol{\epsilon}}_k^H\tilde{\boldsymbol{\epsilon}}_k\right]\mathbf{g}_i} \tag{3.27}$$

Note that for the MMSE-BF design, the transmitter must know two CQIs beyond CDI: i) the channel norm $g_1(\mathbf{h}_k) = ||\mathbf{h}_k||$ and ii) the correlation $g_2(\mathbf{h}_k) = \cos\theta_k$. Therefore each user should use Method 3 of Tab. 3.1 for CSI FB.

The estimated throughput with MMSE beamforming is given by

$$R^{(MMSE)}(\mathcal{S}) = \sum_{k\in\mathcal{S}}\log_2\left(1+\gamma_k^{(MMSE)}\right) \tag{3.28}$$

3.3 CDI feedback strategies

In this section we propose 4 different CDI FB strategies: i) Basic Feedback (BFB), ii) Hierarchical Feedback (HFB), iii) Predictive Feedback with quantization of the error vector (QEV), and iv) Predictive Feedback with Unitary Rotation Matrix (RM). A numerical comparison of the proposed strategies is given later in Section 3.5.2.

3.3.1 Basic feedback signaling

In the basic feedback Signalling (BFB), user k quantizes the "direction" of its channel vector, $\tilde{\mathbf{h}}_k = \frac{\mathbf{h}_k}{||\mathbf{h}_k||}$, to a unit norm vector $\hat{\mathbf{h}}_k$ selected from a codebook. We can have either a differ-

ent quantization codebook $\mathcal{C}_k = \{\mathbf{c}_{k,1}, \ldots, \mathbf{c}_{k,2^B}\}$ for each user, where B is the number of quantization bits and $\mathbf{c}_{k,i}$ are $M \times 1$ unit-norm vectors, or the same codebook for all users, i.e. $\mathcal{C}_k = \mathcal{C}$, $k = 1, \ldots, K$. The quantization criterion is minimum *chordal distance* (see e.g. [42] for a general definition),

$$\hat{\mathbf{h}}_k = \underset{\{\mathbf{c}_{k,j}\} \in \mathcal{C}_k}{\arg\max} |\tilde{\mathbf{h}}_k \mathbf{c}_{k,j}^H| \, . \tag{3.29}$$

which maximizes the CQI of user k in case of ZF BF (see (3.15)). Each user shares knowledge of its codebook with the transmitter, and feeds back the channel quantization index, which requires B bits per mobile. In BFB the reconstructed channel of user k is simply $\bar{\mathbf{h}}_k = \hat{\mathbf{h}}_k$.

We consider two different codebooks: i) RVQ and ii) Lloyd-Max based codebook. In RVQ the 2^B quantization codewords are independently chosen from an isotropic distribution on the M-dimensional unit sphere. As pointed out in [42], since any reasonably well-designed codebook should perform at least as well as RVQ, RVQ gives a lower bound in terms of performance.

Alternatively, codebook design can be performed following the Lloyd-Max algorithm [57] that for a given performance metric $\mu(\tilde{\mathbf{h}}_k, \mathbf{c}_i)$ derives the optimum codebook that maximizes the expectation of $\mu(\tilde{\mathbf{h}}_k, \mathbf{c}_i)$, i.e.

$$\max_{\mathcal{C}} \mathrm{E}[\mu(\tilde{\mathbf{h}}_k, \mathbf{c}_i)] \, . \tag{3.30}$$

In particular we use the Linde, Buzo and Gray (LBG) approach [58], that substitutes the expectation in (3.30) by a sample average. In details, the LBG algorithm considers a large set of N_{TS} channel realizations $\{\tilde{\mathbf{h}}_k\}$, referred as training set (TS)[3] and derives with an iterative approach the optimum codebook which maximizes the average average performance metric[4]

$$\max_{\mathcal{C}} \frac{1}{N_{TS}} \sum_{i=1}^{2^B} \sum_{\tilde{\mathbf{h}}_k \in \mathcal{R}_i} \mu(\tilde{\mathbf{h}}_k, \mathbf{c}_i,) \, , \tag{3.31}$$

where $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{2^B}\}$ is the generic codebook and $\mathcal{R}_i$ is the partition region of the training set associated to codeword $\mathbf{c}_i$.

Since system performance is measured in terms of the achievable sum rate (3.2), a criterion for codebook design is the maximization of the average CQI, e.g (3.15) for ZF BF. This could be done only by numerical methods but there is no guarantee of convergence. We notice that for a given channel realization, (3.15) is maximized by choosing the codeword $\mathbf{c}_i$ that maximizes the correlation

$$\mu(\tilde{\mathbf{h}}_k, \mathbf{c}_i) = |\tilde{\mathbf{h}}_k \mathbf{c}_i^H|^2 \, . \tag{3.32}$$

Therefore we use (3.32) as suboptimum performance metric for codebook design in (3.31).

[3] The size of TS has to scale at least linearly with the number of desired codewords to achieve good performance [20], hence the complexity of codebook design scales at least exponentially with the number of feedback bits. Nevertheless codebook generation can be performed off-line, and codebooks can be uploaded from the base station. Therefore the complexity of the algorithm is not an issue.

[4] We recall that the LBG algorithm converges to a maximum that is not guaranteed to be global, nevertheless it provides a practical way for codebook design even when the PDF of the source signal is not known or difficult to characterize.

3.3.2 Hierarchical feedback

In BFB, the time correlation of the MIMO channel is not exploited. If the channel is changing sufficiently slowly, the mobile CDI feedback could be aggregated over multiple feedback intervals so that the aggregated bits index a larger codebook. In general, a larger codebook implies more accurate knowledge of the MIMO channel at the transmitter, resulting in improved throughput. By aggregating the feedback bits over multiple intervals, the codewords can be arranged in a hierarchical tree structure so that the feedback on a given interval is an index of codewords that are the "children nodes" of a codeword indexed by previous feedback. Based on these considerations we propose an alternative hierarchical FB strategy (HFB) that adopts an incremental feedback approach for the update of the reconstructed channel at the transmitter.

As in BFB, in HFB search, user k quantizes $\tilde{\mathbf{h}}_k$ to a unit norm vector $\hat{\mathbf{h}}_k$ selected from a codebook $\mathcal{C}$ of $2^{B_{\max}}$ unit norm codevectors and using (3.29) as quantization rule. Note that here $B_{\max}$ can be larger than the FB codeword length B.

Codebook Design

We consider a variant of the LBG algorithm that proceeds iteratively by levels in the codebook design, according to the following steps:

1. From the TS, compute the optimum codebook with two codevectors by the LBG algorithm;

2. Split the TS into two subsets, where each subset collects all the channel vectors in TS at minimum chordal distance from the corresponding codevectors;

3. Recursively iterate steps 1) and 2) to each of the subsets of TS.

This binary construction procedure can be represented by a binary tree of B levels, having at level i the codewords of the optimal codebook with 2^i elements.

With the designed codebook, quantization can be performed with a binary search on the tree, thus requiring a lower computation complexity than conventional quantization, at the expense of a larger memory and a little decrease in quantization performance compared to a full tree brute force search.

As shown in Figure 3.1 for the case of $B_{max} = 6$, a binary representation (codeword) of each codevector is obtained by associating a bit to each of the two branches exiting a node and identifying a node at level i with the i bits on the branches leading from the root to the node itself. As a consequence, all nodes of the subtree departing from a node at level i have the same i most significant bits. The codeword of $i + 1$ bits associated to a channel vector can be obtained by adding one bit to the channel vector representation with i bits.

Moreover, slight changes of the channel in subsequent time slots most probably lead to codewords with the same most significant bits. This feature is the key aspect in the HFB signaling.

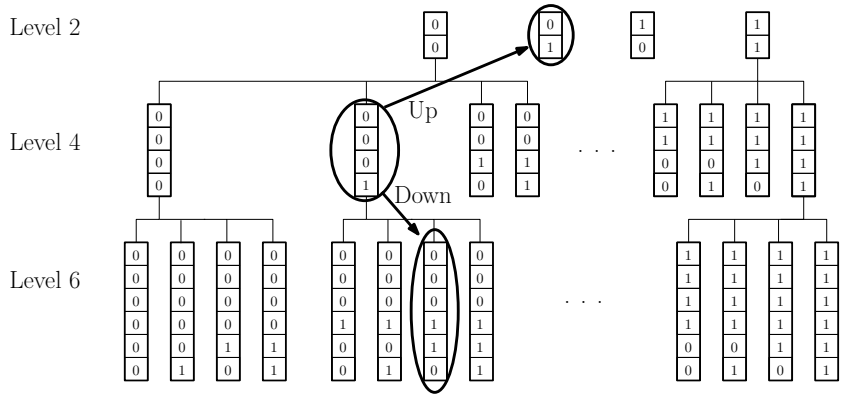

Figure 3.1: Example of tree structure for the LBG-based codebook with $B_{max} = 6$ levels and $B = 3$ bits per feedback interval. Each tree node in levels 2 and 4 have $2^{B-1} = 4$ descendants. Starting from the codeword labeled $[0001]$ in level 4, all codewords from level 2 and descendants of $[0001]$ in level 6 are considered as candidate codewords for the next time slot. If the codeword labeled $[01]$ is the best, then the feedback is $[101]$ where the first bit 1 represents an "up" transition and the remaining bits 01 give the selected codeword at level 2. If the codeword labeled $[000110]$ is the best, then the feedback is $[010]$ where the first bit 0 represents a "down" transition and the remaining bits 10 give the selected codeword which is the descendant of $[0001]$.

HFB feedback signaling

We assume that at slot $n-1$, both transmitter and user k share the reconstructed channel vector $\bar{\mathbf{h}}_k(n-1)$, represented by a binary word of variable length $L_s(n-1)$.

At slot n, user k quantizes $\tilde{\mathbf{h}}_k(n)$ into $\hat{\mathbf{h}}_k(n)$ and compare the first $L_s(n-1)$ bits of the binary representations of $\hat{\mathbf{h}}_k(n)$ and $\bar{\mathbf{h}}_k(n-1)$. The comparison leads to two cases, corresponding to a match (*Down case*) or no match (*Up Case*) between the two sequences. Let $\mathbf{i}_k(n)$ be the binary word of B bits fed back by user k at time slot n. The first bit $i_{k,1}(n)$ denotes the Up or Down case. The following bits are determined as follows:

- *Down Case.* The CSI is refined by feeding back further $B-1$ bits of the B_{max}-bits codeword. These additional bits are obtained by going down by $B-1$ levels into the quantization tree. This is performed by feeding back bits at position $L_s(n-1)+1, \ldots, L_s(n-1)+B-1$ of the codeword associated to $\hat{\mathbf{h}}_k(n)$. Moreover, $L_s(n) = L_s(n-1)+B-1$.

- *Up Case.* The CSI must be updated and the $B-1$ bits $L_s(n-1)-2(B-1)+1, \ldots, L_s(n-1)-B+1$ of the B_{max} bit codeword associated to $\hat{\mathbf{h}}_k(n)$ are fed back to transmitter. Now, $L_s(n) = L_s(n-1)-B+1$.

The proposed algorithm can be easily generalized to account for boundary conditions imposing that $B-1 \leq L_s(n) < B_{max}$. Thanks to this strategy we are able to track channel variations at the cost of an overhead of one flag bit.

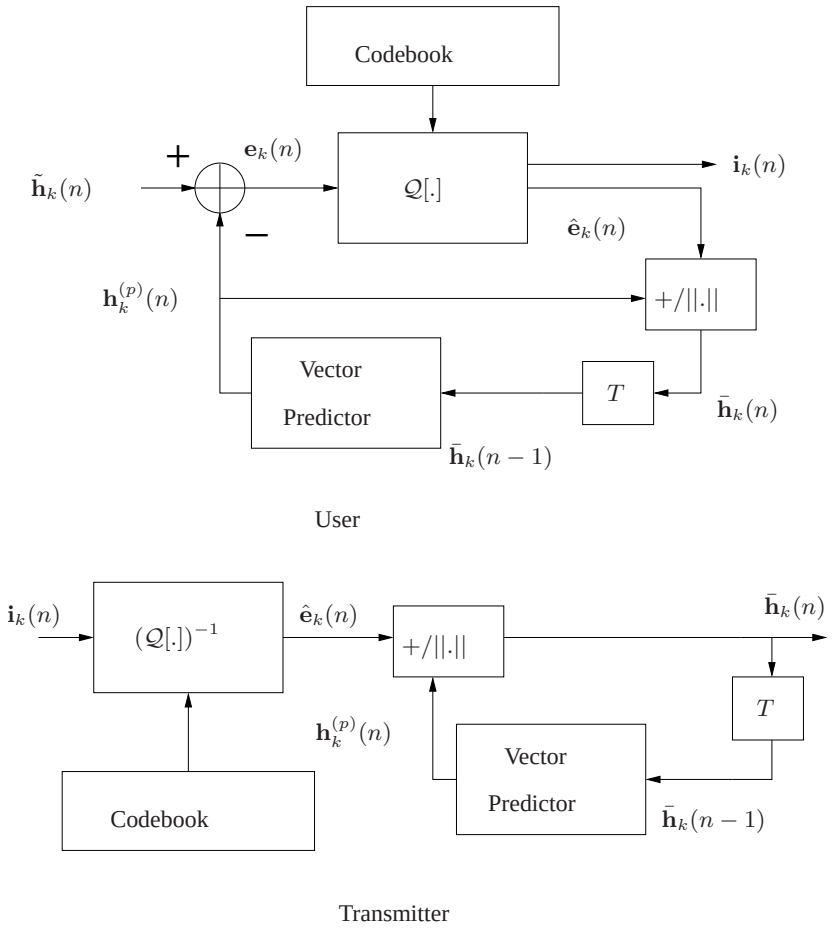

Figure 3.2: Predictive FB (PFB).

3.3.3 Predictive feedback with quantization of the error vector (QEV)

In this section we propose an alternative predictive FB (PFB) strategy where channel quantization and CSI FB is based on predictive vector quantization. As in HFB we aim at exploiting the time-correlation of the channel across different slots.

As depicted in Fig. 3.2, at slot n, both transmitter and user obtain a prediction $\mathbf{h}_k^{(p)}(n)$ of the channel direction $\tilde{\mathbf{h}}_k(n)$, based on past reproduced values $\{\bar{\mathbf{h}}_k(m)\,,\, m < n\}$. For example, a simple first order linear predictor yields

$$\mathbf{h}_k^{(p)}(n) = \bar{\mathbf{h}}_k(n-1) \tag{3.33}$$

where only the previous CSI value is used for prediction. Next, each user quantizes the prediction error $\mathbf{e}_k(n) = \tilde{\mathbf{h}}_k(n) - \mathbf{h}_k^{(p)}(n)$ and feeds back to the transmitter $\mathbf{i}_k(n)$, a binary representation of the quantized vector error $\hat{\mathbf{e}}_k(n)$ using B bits. Both transmitter and user update the reproduced channel vector $\bar{\mathbf{h}}_k(n)$ by combining the prediction with the quantized prediction error, i.e.,

$$\bar{\mathbf{h}}_k(n) = \frac{\mathbf{h}_k^{(p)}(n) + \hat{\mathbf{e}}_k(n)}{||\mathbf{h}_k^{(p)}(n) + \hat{\mathbf{e}}_k(n)||}\,, \tag{3.34}$$

denoted as $+/||.||$ in Fig. 3.2.

The codebook of the prediction error quantizer is designed with the aim of minimizing the mean square error $\mathrm{E}[||\mathbf{e}_k - \mathbf{c}_i||^2]$. As in HFB we consider the generalized LBG algorithm [58]. We follow the open loop approach, hence from a training set $\{\tilde{\mathbf{h}}_k(n)\}$ we first obtain the set of channel predictions and channel prediction errors $\{\mathbf{e}_k(n)\}$, which are then used to design the codebook using the LBG algorithm.

3.3.4 Predictive feedback with unitary rotation matrix (RM)

In this section we describe an alternative FB technique still based on quantization of the prediction error but using quantization and prediction rules different from the scheme in Section 3.3.3[5]. At slot n, both transmitter and user k obtain a prediction $\mathbf{h}_k^{(p)}(n)$ of the CDI $\tilde{\mathbf{h}}_k(n)$, based on past reconstructed vectors $\{\bar{\mathbf{h}}_k(m)\,,\, m < n\}$. As in Section 3.3.3 we consider a simple first order linear predictor (3.33). We note that, since CDIs are unit norm vectors, this predictor is the optimal first order predictor for the minimization of the chordal distance.

Since $\mathbf{h}_k^{(p)}(n)$ and $\tilde{\mathbf{h}}_k(n)$ are unit-norm vectors, we model the prediction error as a rotation vector from the predicted vector $\mathbf{h}_k^{(p)}(n)$ to the true normalized channel vector $\tilde{\mathbf{h}}_k(n)$.

In details, at slot n both user k and transmitter derive in the complex hyperspace $\mathbb{C}^{M\times 1}$ of the MIMO channel a unitary basis whose first element is given by the predicted vector $\mathbf{h}_k^{(p)}(n)$. This is done by computing the unitary $M \times M$ matrix $\mathbf{Z}_k(n)$ obtained by the Gram-Schmidt orthogonalization procedure [20] applied to the columns of $[\mathbf{h}_k^{(p)}(n)\ \ \mathbf{I}_M]$, where $\mathbf{I}_M$ is the $M \times M$ identity matrix. With this definition the components of $\mathbf{h}_k^{(p)}(n)$ in the new basis are contained in the constant vector $\mathbf{u} = \mathbf{Z}_k(n)^H \mathbf{h}_k^{(p)}(n) = [1\,0\,\ldots\,0]^T$, while the prediction error vector is defined as

$$\mathbf{e}_k(n) = \mathbf{Z}_k^H(n)\tilde{\mathbf{h}}_k(n)\,. \tag{3.35}$$

Let $\hat{\mathbf{e}}_k(n)$ be the quantized version of $\mathbf{e}_k(n)$ fed back to the transmitter. The reconstructed vector is defined as

$$\bar{\mathbf{h}}_k(n) = \mathbf{Z}_k(n)\hat{\mathbf{e}}_k(n)\,. \tag{3.36}$$

We note that $\mathbf{e}_k(n)$ is expected to lie with high probability in an hyper-cone centered around the constant vector $[1, 0, \ldots, 0]^T$ and whose surface area, although depending on channel time correlation, is usually much smaller than the complete surface area of the unitary hyper-sphere described by $\tilde{\mathbf{h}}_k(n)$. This suggests that for a target quantization distortion we need fewer codewords to quantize the prediction error $\mathbf{e}_k(n)$ than what we would need to quantize $\tilde{\mathbf{h}}_k(n)$ as in RVQ [42] or Grassmannian line packing [46].

For codebook design we use the LBG algorithm. In this case, from (3.32) and (3.36) the metric to be maximized is given by

$$\begin{aligned} \mu(\tilde{\mathbf{h}}(n), \mathbf{c}) &= \left|\tilde{\mathbf{h}}^H(n)\mathbf{Z}(n)\mathbf{c}\right|^2 \\ &= \mathbf{c}^H\mathbf{Z}^H(n)\tilde{\mathbf{h}}(n)\tilde{\mathbf{h}}^H(n)\mathbf{Z}(n)\mathbf{c}\,. \end{aligned} \tag{3.37}$$

[5]The feedback scheme described in this section is similar to the technique proposed in [59] but has been derived in a completely independent way

INIT: $\mathcal{A}_0 \leftarrow \{1, \ldots, K\}$,
$s_1 \leftarrow \arg\max_{k \in \mathcal{A}_0} \gamma_k^{(ZF,2)}$
$\mathcal{S}_1 \leftarrow \{s_1\}$
for $n = 2$ to M **do**

$$\mathcal{A}_n \leftarrow \{k \in \mathcal{A}_{n-1} : |\bar{\mathbf{h}}_k \bar{\mathbf{h}}_{s_j}^H| \leq \epsilon, j = 1, \ldots n\} \tag{3.39}$$

  if $|\mathcal{A}_n| = 0$ **then**
    $\mathcal{S} \leftarrow \mathcal{S}_{n-1}$
    break
  else
    $s_n \leftarrow \arg\max_{k \in \mathcal{A}_n} \gamma_k^{(ZF,2)}$
    $\mathcal{S}, \mathcal{S}_n \leftarrow \mathcal{S}_{n-1} \cup \{s_n\}$
  end if
end for

Table 3.2: Pseudo-code of SUS.

We follow the open loop approach, hence from a sequence of channel vectors $\{\tilde{\mathbf{h}}(n)\}$ we derive the set of channel predictions $\{\boldsymbol{h}^{(p)}(n)\}$, which are used to compute $\{\mathbf{Z}(n)\}$ in (3.37).

We notice that if we define the $M \times M$ complex matrix relative to the partition region $\mathcal{R}_i$ of the training set

$$\mathbf{A}_i = \sum_{\tilde{\mathbf{h}}(n) \in \mathcal{R}_i} \mathbf{Z}^H(n) \tilde{\mathbf{h}}(n) \tilde{\mathbf{h}}^H(n) \mathbf{Z}(n), \tag{3.38}$$

it's easy to show from (3.31) and (3.37) that the optimum codeword for the partition region $\mathcal{R}_i$ is the dominant eigenvector of matrix $\mathbf{A}_i$ normalized to unit norm.

3.4 User selection schemes

In this section, after a quick review of the user-selection scheme adopted in [42], we introduce our greedy user-selection algorithms. We notice that all algorithms aim at maximizing the sum rate of the system, nevertheless they can be easily generalized to a weighted sum rate criterion as in Chapter 2.

3.4.1 Semi-orthogonal user selection (SUS): review

Using both CQIs and CDIs from all the K users, the transmitter performs a semi-orthogonal user selection (SUS) algorithm [42] to support up to M users in each time slot. In more details, SUS sets a parameter ϵ that establishes the maximum spatial correlation allowed between reconstructed user channels. At step n, firstly it selects among the remaining users the set $\mathcal{A}_n$ of users having small correlation with already selected users. Then it chooses the user with largest CQI in the set $\mathcal{A}_n$. By using this heuristic algorithm the transmitter selects only semi-orthogonal users with large CQI. The SUS algorithm is outlined in Tab. 3.2.

1: INIT: $\mathcal{S}_0 \leftarrow \emptyset$, $U \leftarrow \{1, \ldots, K\}$, $R^{(i)}(\mathcal{S}_0) \leftarrow 0$
2: **for** $n = 1$ to M **do**
3:

$$s_n \leftarrow \arg \max_{u \in U \setminus \mathcal{S}_{n-1}} R^{(i)}(\mathcal{S}_{n-1} \cup \{u\}) \tag{3.40}$$

4: **if** $R^{(i)}(\mathcal{S}_{n-1} \cup \{s_n\}) \leq R^{(i)}(\mathcal{S}_{n-1})$ **then**
5: $\mathcal{S} \leftarrow \mathcal{S}_{n-1}$
6: **break**
7: **else**
8: $\mathcal{S}, \mathcal{S}_n \leftarrow \mathcal{S}_{n-1} \cup \{s_n\}$
9: **end if**
10: **end for**

Table 3.3: Pseudo-code of Algorithm 1 *for user selection.*

3.4.2 Improved user selection schemes

The SUS algorithm depends on the correlation parameter ϵ, which must bet set beforehand. Its value is difficult to optimize, indeed, if it is chosen too small, there are chances that very few users are scheduled for transmission. On the other hand, if it is exceedingly large, the transmitter may select unwanted users that cause too much interference.

Here we propose two greedy user selection algorithms using quantized CSIT, which, unlike the SUS algorithm, do not depend on design parameters. Both algorithms can be used in case of ZF and MMSE beamforming and in combination with any of the signalling techniques of Table 3.1. The proposals generalize the algorithm proposed in [21] under the assumption of perfect CSIT.

In Algorithm 1, users are added successively one at a time, up to a maximum of M, if the estimated achievable throughput is increased. The pseudo-code for Algorithm 1 is given in Tab. 3.3.

As explained in the numerical results, when Algorithm 1 is used with the more accurate SINR approximation given by Method 3, the system does not become interference limited for increasing SNR as it happens with Methods 1 and 2. Indeed, for high SNR both ZF BF and MMSE BF achieve the same sum rate of a simpler transmission scheme that serves only the best user in each time slot, which we denote as time division multiple access (TDMA). In fact, if the number of quantization bits is small and kept constant with SNR, eventually for high SNR transmitting to the best user using a TDMA approach turns out to be optimum in terms of sum rate. Unfortunately, the user selection Algorithm 1 applied jointly with Method 3 reduces the number of active users at high SNR, but performs worse than with Method 2 in the intermediate SNR region.

In order to overcome this drawback we propose a novel user selection mechanism (Algorithm 2) which leads to an increment of the average number of selected users. Algorithm 2 can be explained by representing the possible sets of selected user as paths in a tree of depth $\min\{M, K\}$ (see Figure 3.3). The number of nodes at the n-th level of the tree is $\frac{K!}{(K-n)!}$. The greedy selection Algorithm 1 performs a very limited search, just over $\sum_{n=1}^{\min\{M,K\}}(K-n+1)$ nodes. Our goal is to improve Algorithm 1 without adding much complexity.

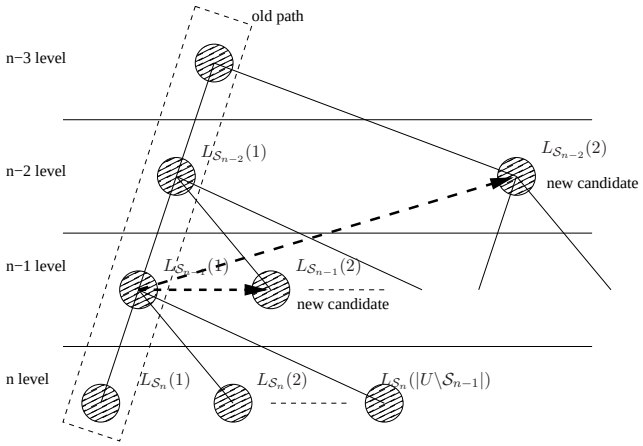

Figure 3.3: Representation as a tree search of Algorithm 2 for user selection.

Let us define $U = \{1, \ldots, K\}$ and consider a given path $\mathcal{S}_{n-1}$ of length $(n-1)$. We introduce an ordering of the nodes at level n belonging to the same subtree of $\mathcal{S}_{n-1}$, and indicate these nodes as $\{L_{\mathcal{S}_n}(j)\}_{j=1,\ldots,|U\backslash\mathcal{S}_{n-1}|}$, such that $R^{(i)}(\mathcal{S}_{n-1} \cup \{L_{\mathcal{S}_n}(1)\}) \geq \ldots \geq R^{(i)}(\mathcal{S}_{n-1} \cup \{L_{\mathcal{S}_n}(k)\}) \geq \ldots \geq R^{(i)}(\mathcal{S}_{n-1} \cup \{L_{\mathcal{S}_n}(|U\backslash\mathcal{S}_{n-1}|)\})$. When exploring a subtree of $L_{\mathcal{S}_{n-1}}(j)$, if the condition $R^{(i)}\left(\mathcal{S}_{n-1} \cup L_{\mathcal{S}_n}(1)\right) \leq R^{(i)}\left(\mathcal{S}_{n-1}\right)$ is met, instead of stopping the search as in Algorithm 1 and taking $\mathcal{S}_{n-1}$ as the final set of active users, we go back by one level and start exploring the "second best" subtree, by setting $\mathcal{S}_{n-1} = \mathcal{S}_{n-2} \cup L_{\mathcal{S}_{n-1}}(j+1)$. In order to keep control of the maximum number of visited nodes, we use the parameter $n_{\text{bs}_{\text{MAX}}}$, which represents the maximum number of trees (or backward steps) being successfully searched. The pseudo-code of Algorithm 2 is shown in Tab. 3.4.

In Fig. 3.4, we consider $M = 4$, $K = 20$, $B = 8$, $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and ZF beamforming and compare the average number of users selected by the two algorithms as a function of SNR. We notice that Algorithm 2 selects a larger number of users with respect to Algorithm 1. In particular with SNR $= 25$ dB Algorithm 2 increases the number of selected users with respect to Algorithm 1 by almost a factor of 2.

3.5 Simulation results

In this section we compare the proposed solutions for i) user selection, ii) CDI feedback and iii) beamformer design, by means of numerical simulations.

3.5.1 Greedy user selection vs SUS

We consider a flat Rayleigh fading channel model with i.i.d. elements, i.e $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Adopting ZF beamforming and simple RVQ, in Figs. 3.5 and 3.6 we compare the user selection schemes proposed in Section 3.4 in terms of sum rate vs. SNR for $K = 20$ users, $M = 4$ antennas and $B = 4, 8$ bits.

In Fig. 3.5 we compare Algorithm 1 for user selection (see Table 3.3) with SUS for the three CQI methods summarized in Table 3.1. As terms of comparison we also include the sum

INIT: $\mathcal{S}_0 \leftarrow \emptyset$, $U \leftarrow \{1,\dots,K\}$, $R^{(i)}(\mathcal{S}_0) \leftarrow 0$, $n_{bs} \leftarrow 1$
for $n = 1$ to M **do**
  $L_{\mathcal{S}_n} \leftarrow \underset{u \in U \setminus \mathcal{S}_{n-1}}{\arg\,\text{sort}} \left\{ R^{(i)}(\mathcal{S}_{n-1} \cup \{u\}) \right\}$
  $\text{pos}[n] \leftarrow 1$
  $\mathcal{S}_n \leftarrow \mathcal{S}_{n-1} \cup L_{\mathcal{S}_n}(1)$
  if $R^{(i)}(\mathcal{S}_n) \leq R^{(i)}(\mathcal{S}_{n-1})$ & $n_{bs} \leq n_{bs_{MAX}}$ **then**
    BACKOFF $\leftarrow 1$, $l \leftarrow n-1$, $\mathcal{S}^{(n_{bs})} \leftarrow \mathcal{S}_{n-1}$
    while BACKOFF & $l \geq 0$ **do**
      $\text{pos}[l] \leftarrow \text{pos}[l] + 1$
      if $\text{pos}[l] \leq |L_{\mathcal{S}_l}|$ & $R^{(i)}(\mathcal{S}_{l-1} \cup L_{\mathcal{S}_l}[\text{pos}[l]]) > R^{(i)}(\mathcal{S}_{l-1})$ **then**
        BACKOFF $\leftarrow 0$, $n_{bs} \leftarrow n_{bs} + 1$,
        $\mathcal{S}_l \leftarrow \mathcal{S}_{l-1} \cup L_{\mathcal{S}_l}[\text{pos}[l]]$, $n \leftarrow l$
      else
        $l \leftarrow l - 1$
      end if
    end while
  end if
end for
$\mathcal{S} \leftarrow \arg \underset{\{S^{(j)}\}_{j=1,\dots,n_{bs}}}{\max} R^{(i)}(\mathcal{S}^{(j)})$

Table 3.4: Pseudo-code of Algorithm 2 for user selection.

rate achievable with TDMA and ZF BF (see also Chapter 2), both under the assumption of perfect CSIT. For the SUS algorithm, we set $\epsilon = 0.4$, which was empirically found to be a good choice. The optimization of ϵ, however, remains an open problem.

We observe that Method 1 generally yields very poor performance. For low SNR, Method 2 outperforms all the other schemes at all feedback rates B. However, at high SNR, Method 2 tends to allocate too many users and the system becomes interference limited, therefore, at certain SNR depending on the number of FB bits, TDMA becomes preferable. This depends on the fact that the SINR lower-bound used by Method 2 is tight for a number of active users close to M but is loose for few active users (one or two), therefore the system wrongly estimates the sum-rate at high SNR, where in fact TDMA eventually becomes optimal[6]. This problem can be partially solved using Method 3, where a better approximation of the SINRs is guaranteed in each step of Algorithm 1 for user selection. However, for the reasons already mentioned in Section 3.4 we observe that at intermediate SNR Method 2 outperforms Method 3.

In order to improve performance of Method 3 at intermediate SNR we use Algorithm 2 for user selection (see Table 3.4). In Figure 3.6 we can see how simulations results confirm the considerations in Section 3.4 concerning the improvements of Algorithm 2 over Algorithm 1. We emphasize that by using Algorithm 2 and Method 3, the saturation of the sum rate is avoided and we are able to achieve very good performance over all the SNR range. Notice that the maximum number of backward steps used for the results of Fig. 3.6 is $n_{\text{bs}_{\text{MAX}}} = 5$.

[6]This is true only for a number of feedback bits fixed with the SNR.

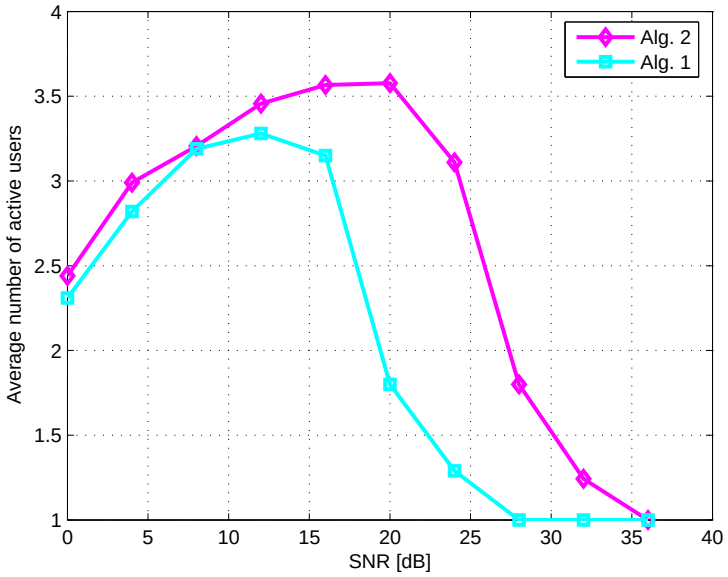

Figure 3.4: Average number of allocated users vs SNR for $K = 20$ for $M = 4$, $B = 8$ and ZF beamforming. SINRs have been calculated by using Method 3.

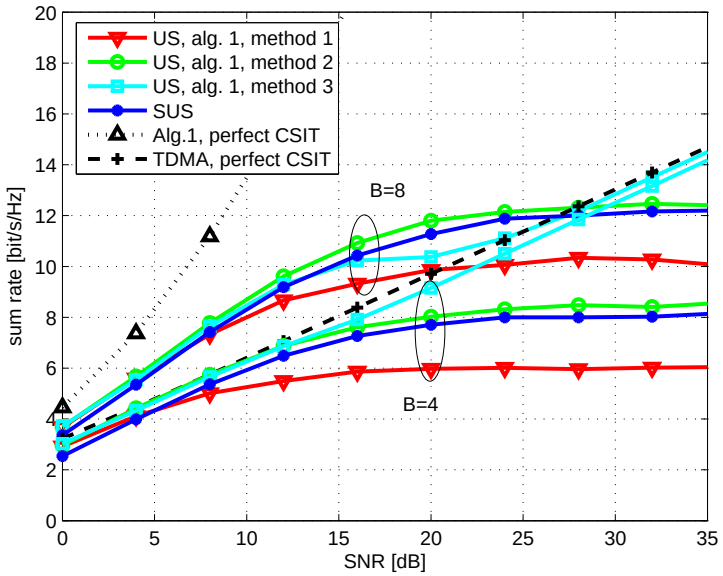

Figure 3.5: Average throughput vs SNR for $M = 4$, $K = 20$ and $B = 4, 8$. Comparison between the proposed techniques and SUS [42].

3.5.2 Comparison between CDI feedback strategies

We still consider a transmitter equipped with $M = 4$ antennas and $K = 20$ users. Differently from Section 3.5.1 the channel is modelled as *time-variant* within each time slot and flat Rayleigh fading, according to the spatial channel model (SCM) [60]. The carrier frequency is

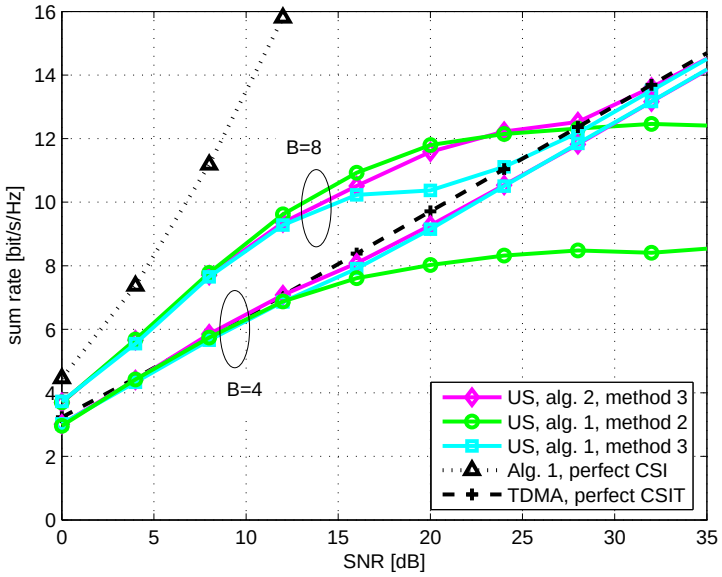

Figure 3.6: Average throughput vs SNR for $M = 4$, $K = 20$, $B = 4, 8$ and $n_{bs_{MAX}} = 5$. Comparison between Algorithm 1 and Algorithm 2.

2 GHz, the transmission bandwidth is 5 MHz and the distance between two adjacent transmit antennas is 10 wavelength. The time slot duration is $T = 0.5$ ms and each user transmits the FB once per slot. The transmitter performs user selection according to Algorithm 1, computes the CQI feedback as in Method 2, and designs the ZF beamformer at the beginning of the time slot, keeping it unchanged for the whole time slot. Under this setting the achievable sum rate is determined as

$$\mathrm{E}\left[\sum_{k=1}^{|\mathcal{S}(n)|} \bar{R}_k(t)\right], \tag{3.41}$$

with $\bar{R}_k(t)$ the achievable rate of user k at time t of slot n, i.e.,

$$\bar{R}_k(t) = \log_2\left[1 + \mathrm{SINR}_k(t)\right] \tag{3.42}$$

and extending (3.1) to time variant channels inside a time slot,

$$\mathrm{SINR}_k(t) = \frac{|\mathbf{h}_k(t)\mathbf{g}_k(n)|^2}{1 + \sum\limits_{i \in \mathcal{S}(n)\setminus\{k\}} |\mathbf{h}_k(t)\mathbf{g}_i(n)|^2}. \tag{3.43}$$

For the HFB strategy, the largest codebook size is $B_{max} = 12$, corresponding to a tree with 12 levels. The codebook for QEV and RM is designed from a TS composed of channel vectors of the SCM for users moving at 3, 50 and 130 km/h with equal probability.

We compare the following FB strategies: 1) BFB with RVQ, 2) BFB with LBG-based codebook, 3) HFB with linear search of the best codeword, 4) QEV with a simple holder as predictor, 5) RM with a simple holder as predictor, 6) ZF BF with perfect CSIT (PCSIT)

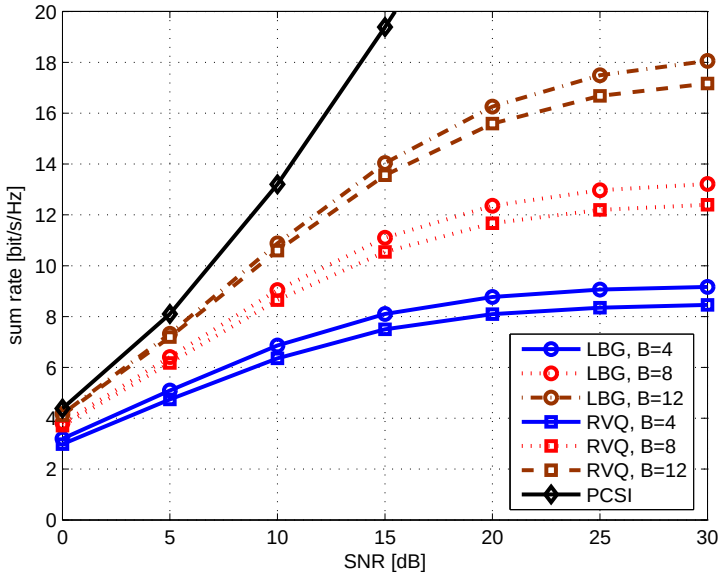

Figure 3.7: Sum rate as a function of SNR for the BFB strategy, using RVQ and LBG quantization methods. Channel with block fading, $v = 130$ km/h.

which gives an upper bound for the achievable sum rate. Moreover, as term of comparison we include 7) Predictive FB with state evolution (PFB-SE), a recursive feedback reduction method based on [61]. This scheme has been proposed to exploit the correlation of the channel in the frequency domain for a MIMO OFDM system. The same approach can be used in our scenario to exploit the channel correlation in the time domain. Quantization is performed directly on $\tilde{\mathbf{h}}_k(n)$ using a time variant quantizer. At slot $n = 1$ user quantizes $\tilde{\mathbf{h}}_k(1)$ according to the minimal chordal distance (3.29) for a codebook $\mathcal{C}$ composed of $2^{B'}$ codewords. In general, say $\mathbf{c}$ the codeword used at slot $n-1$, at slot n user uses a codebook $\mathcal{C}_{\mathbf{c}}$ containing only the 2^B codewords of $\mathcal{C}$ at minimal chordal distance from $\mathbf{c}$. In the simulations we use $B' = B + 8$.

In Fig. 3.7 we compare BFB with RVQ and BFB with LBG, in terms of sum rate as a function of SNR for a block fading (BF) channel (i.e., assuming the channel invariant during a time slot). Performance is evaluated for different FB bits B and the sum rate achievable with ZF BF and PCSIT is added as an upper bound. We note that LBG-based codebook outperforms RVQ for any SNR condition and FB rate thanks to its capability of better exploiting the spatial correlation of the channel. Moreover as in Figs. 3.5 and 3.6 the sum rate eventually saturates for high SNR, due to quantization errors that lead to inaccurate ZF beamforming and multiuser interference. From results not reported here it is also seen that when the channel is time varying within a slot, both LBG an RVQ show a performance degradation with respect to block fading, since the beamformer design is for outdated channel vectors. However, the LBG-based codebook still yields an higher sum rate than RVQ in any SNR condition.

In Figs. 3.8 and 3.9 we set the average SNR $= 15$ dB, assume a time-variant channel, and compare the proposed FB strategies in terms of sum rate as a function of FB bits B, for users moving at 3 and 130 km/h, respectively. As terms of comparison we include ZF beamforming with perfect CSIT and PFB-SE. Firstly, we notice that HFB, QEV and RM provide a significant

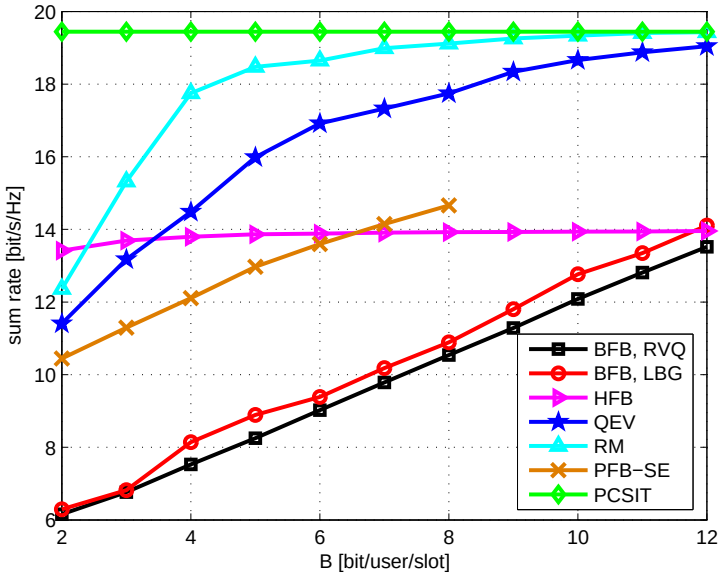

Figure 3.8: Sum rate as a function of FB bits for various FB strategies. M=4, K=20, SNR = 15 dB and $v = 3$ km/h.

gain over BFB, especially in a slowly-time variant channel, because they exploit channel time correlation. For few FB bits HFB is the best option for both $v = 3$ and 130 km/h, but as the FB rate increases predictive FB strategies, e.g. QEV and RM, become preferable and approach the sum rate achievable with perfect CSIT. For $v = 3$ km/h HFB achieves its best performance already for $B = 4$ and there is no further gain with an increment of the FB rate, as its performance is limited by the maximum size $B_{\max}$ of the codebook. For high FB rate and high speed, HFB has worse performance than BFB because the channel rapidly changes, and all users send with very high probability the most significant $B - 1$ bits of the quantized channel. In this case, the flag bit yields a rate inefficiency. Differently, predictive FB strategies are not affected by this drawback and provide higher sum rate. In particular RM always outperforms QEV because it adopts a better strategy for the quantization of the prediction error vector. Unfortunately we notice that the gap with BFB significantly reduces as the speed of the user increases.

In Figs. 3.10 and 3.11 we set $B = 6$ and compare the proposed FB strategies in terms of sum rate as a function of SNR, for $v = 3$ and 130 km/h, respectively. In the low SNR region the gap between the various strategies is small because system noise is dominant, but as SNR increases RM becomes the best option because it leads to a more accurate quantization accuracy and less multiuser diversity. In particular for $v = 3$ km/h it performs close to ZF with PCSIT already for $B = 6$. The gain with respect to BFB reduces when the user speed increases, especially at high SNR.

3.5.3 ZF beamforming vs MMSE beamforming

In Fig 3.12 we set $M = 4$, $K = 4$ and $v = 130$ km/h and compare MMSE-BF and ZF-BF in terms of sum rate as a function of SNR, for both BFB and RM. We assume dedicated channels

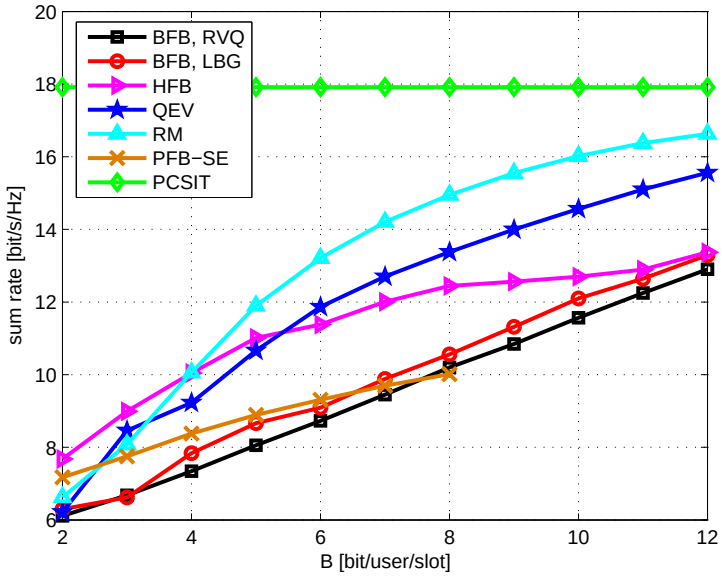

Figure 3.9: Sum rate as a function of FB bits for various FB strategies. M=4, K=20, SNR = 15 dB and $v = 130$ km/h.

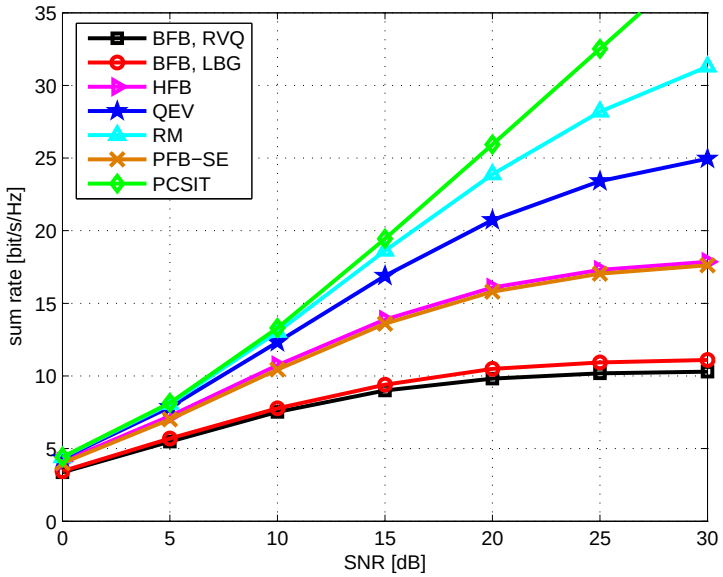

Figure 3.10: Sum rate as a function of SNR for various FB strategies and FB bits. Users moving at 3 km/h.

for the $K = 4$ users, i.e. there is no user selection (or equivalently $K = 4$ users are selected randomly). MMSE-BF is preferable because it better copes with multiuser interference caused by quantization errors, although it requires a double CQI FB. Nevertheless, we verified that even considering the mean value of $\cos\theta_k$ in BF design, MMSE-BF still highly improves ZF-BF.

Simulation results revealed that in case users are selected with an opportunistic approach

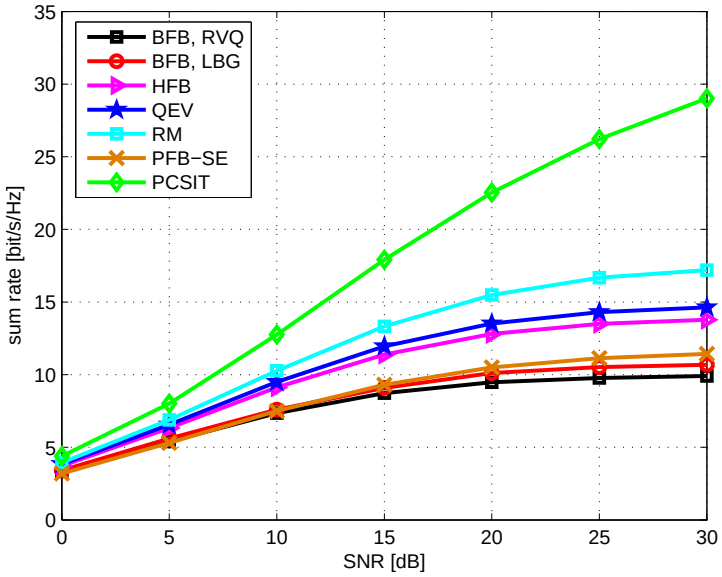

Figure 3.11: Sum rate as a function of SNR for various FB strategies and FB bits. Users moving at 130 km/h.

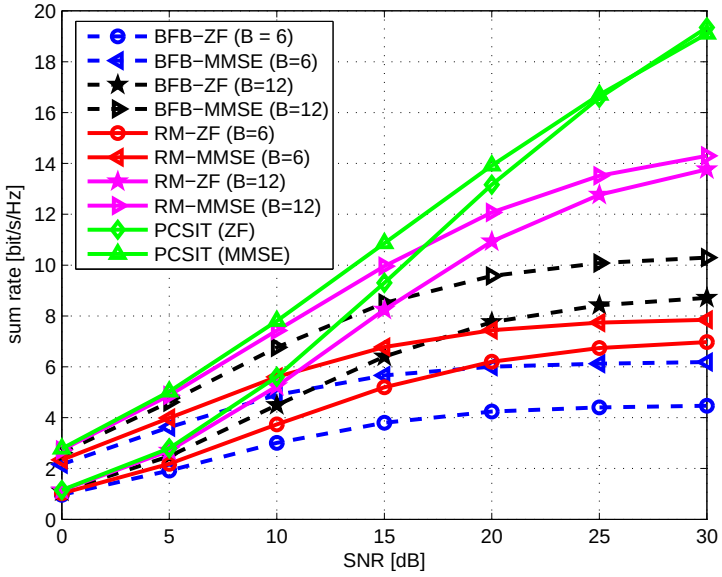

Figure 3.12: Sum rate as a function of SNR for various FB strategies and FB bits adopting both ZF-BF and MMSE-BF. Users moving at 130 km/h.

MMSE-BF does not provide gain with respect to ZF-BF. Indeed both SUS and the proposed greedy user selection algorithms try to select almost orthogonal users, thus limiting multiuser interference that MMSE-BF tries to cope with.

3.6 Conclusions

This chapter consider a MIMO-BC with limited FB and single antenna receivers and investigates three different issues: i) beamformer design ii) channel quantization and feedback optimization iii) user selection. In case of randomly selected users the proposed MMSE BF outperforms ZF BF in terms of achievable throughput, nevertheless this gain becomes negligible when users are selected by an opportunistic approach. This result recalls a well know finding under perfect CSIT.

We propose novel channel quantization algorithms and feedback strategies that exploit spatial and time correlation of the MIMO channel and lead to significant performance improvement over conventional approaches. In particular predictive FB with quantized rotation matrix achieves close to optimum performance even with a moderate number of FB bits. In slowly time variant channels and with few FB bits hierarchical FB is very competitive and is less complex to implement with respect to predictive FB strategies.

We introduce new opportunistic user selection algorithms that exploit multiuser diversity in systems where users do not have strict delay constraints. The proposals use limited uplink FB information and select the set of active user in each time slot based on a sum rate criterion. Interestingly no off-line parameter adjustments are required and they provide an improved estimation of the achievable throughput with respect to information fed back by users.

Chapter 4

Multiuser MIMO downlink with limited feedback and multiple antenna receivers

In this chapter we address the problem of transceiver design and channel quantization in a MU MIMO downlink system with limited uplink FB where users employ multiple antennas. Interestingly quantization codebook design adapts to the transceiver structure and exploits MIMO channel statistic.

A first solution considers zero-forcing (ZF) beamforming. While the codebook design and receiver combining strategies are dependent, the combining strategy can be derived for a fixed set of codebook vectors. Because CSI at transmitter (CSIT) is not ideal, each active user will experience residual interuser interference. Exploiting the additional degrees of freedom provided by multiple receive antennas, the combiner is designed to maximize the output expected signal-to-interference-plus-noise ratio (SINR), where expectation is taken with respect to the SINR as the receiver is assumed to have no apriori knowledge of the other beams used by the transmitter. We call this strategy as *maximum expected SINR combiner* (MESC). We provide an analytic characterization of the achievable throughput of the proposed combiner in the case of many users. Extending results in [62] and [42], we show how additional receive antennas or higher multiuser diversity can reduce the required feedback rate to achieve a target throughput. We show how scaling the number of feedback bits linearly with the SNR expressed in dB assures a constant gap from the achievable sum rate with perfect CSIT. Furthermore, the constant of proportionality linearly decreases with the number of receive antennas.

The quantization codebook is designed according to the Lloyd-Max algorithm [20], extending the approach used in Chapter 3 for single-antenna receivers. Following 3GPP-LTE system design guidelines we extend the FB strategies proposed in Chapter 3 giving special interest to reduced complexity and low FB rate solutions. In particular we generalize basic FB (BFB) and hierarchical FB (HFB) strategies to the case of multiple-antenna users. This is accomplished after deriving two new performance metrics for codebook design, corresponding to low and high SNR regimes. Interestingly HFB provides significant performance improvement over BFB for users with moderate mobility.

A second solution considers unitary beamforming (U-BF) [44] extending the technique in [22]. A unitary precoder allows for a perfect estimate of the user SINR at both sides of the transmission link, yielding to a design of the MMSE combiner with no approximations and a more reliable user selection. In this case adapting to the transceiver structure, the codebook design i) exploits the statistic of the channel model and ii) comprises an high number of unitary matrices. Two greedy solutions are proposed and their effectiveness is validated by simulations.

The chapter is organized as follows. In Section 4.1 we describe the system model. Section 4.2 describes ZF with MESC technique under limited feedback and Section 4.3 provides an asymptotic performance analysis of the proposed strategy in case of many users. Section 4.4 describes the LBG-based codebook design technique and generalizes basic FB and hierarchical FB to multiple antenna receivers. Then, Section 4.5 proposes a novel unitary beamforming scheme and numerical results are presented in Section 4.6. Finally Section 4.7 concludes the chapter summarizing the main findings.

Part of this chapter has been published in [63], [64], [65] and [66].

4.1 System model

We consider a narrowband multiantenna downlink channel modelled as a MIMO-BC with flat fading, where K users request service from a transmitter equipped with M antennas. Differently from Chapter 3 we assume that each user has $N > 1$ receive antennas. The discrete-time complex baseband received signal by the kth user is given by (2.1).

As shown in Chapter 2, since $N > 1$ each user could receive multiple spatial data streams [67, 39], however we transmit at most a single stream to each active user. In systems where $K \gg M$ (as in the practical cellular systems we consider), this restriction is justified through analytic [68] and empirical observations (see Chapter 2) under ideal CSIT. Moreover similar observations were made in the case of limited CSIT [62]. Furthermore, limiting the transmission to a single stream results in less uplink feedback per user, providing additional justification for the restriction.

As for single antenna users the proposed transceiver architectures comprise three phases: i) channel estimation, determination of FB information and combining vector design at each receiver, ii) user selection and beamformer design at the transmitter and iii) data demodulation at the receivers after estimating the equivalent channel by dedicated pilots[1]. We underline that the additional degrees of freedom provided by multiple antenna at the receivers implies the computation of a combining vector and a different approach for the determination of CDI and CQI.

4.2 The maximum estimated SINR combiner technique

In this section we describe a first transmission strategy based on ZF beamforming at the transmitter and maximum estimated SINR combiner (MESC) at receivers.

[1]Dedicated downlink training is not strictly necessary in case of unitary BF because each selected user already knows its beamforming vector and can estimate the equivalent channel to perform coherent detection.

4.2.1 Phase I: Determining feedback from receivers

In this section for ease of notation we refer to a BFB strategy where the reconstructed channel vector at the transmitter is simply the CDI fed back by the user. An alternative FB strategy based on HFB is described later in Section 4.4.2.

We suppose that the transmitter serves a set of users $\mathcal{S}$. The kth receiver ($k \in \mathcal{S}$) processes the received signal using a linear combiner given by a unit-norm N-dimensional vector $\mathbf{u}_k$. The signal at the combiner output is

$$\begin{aligned} r_k &= \mathbf{u}_k^H \mathbf{y}_k \\ &= \mathbf{u}_k^H \mathbf{H}_k \mathbf{g}_k d_k + \mathbf{u}_k^H \mathbf{H}_k \sum_{i \in \mathcal{S}, i \neq k} \mathbf{g}_i d_i + n'_k \end{aligned} \tag{4.1}$$

where $\mathbf{y}_k$ is the received signal given by (2.1), n'_k is the unit variance complex Gaussian processed noise, and $\mathbf{g}_i$ is the $M \times 1$ precoding vector for the ith selected user. From (4.1) the SINR for user k is

$$\text{SINR}_k = \frac{p_k |\mathbf{u}_k^H \mathbf{H}_k \mathbf{w}_k|^2}{1 + \sum\limits_{i \in \mathcal{S} \setminus \{k\}} p_i |\mathbf{u}_k^H \mathbf{H}_k \mathbf{w}_i|^2} \tag{4.2}$$

where we defined $\mathbf{g}_k = \sqrt{p_k}\mathbf{w}_k$ as the beamforming vector for user k, with $\mathbf{w}_k$ the ZF precoder to be defined in Section 4.2.2. We consider an average sum-power constraint and impose equal power-allocation for the selected users[2], i.e. $p_k = \frac{P}{|\mathcal{S}| ||\mathbf{w}_k||^2}$.

We define $\mathbf{v}_k = \mathbf{H}_k^H \mathbf{u}_k$ to be the equivalent MISO channel for the kth user and assume $\hat{\mathbf{v}}_k$ to be its unit-norm quantized version fed back as CDI. Following the same approximations used in Section 3.2.1 and assuming a full loaded system, i.e. $|\mathcal{S}| = M$, we get an approximated lower bound for the expected SINR

$$\text{E}\left[\text{SINR}_k\right] \gtrapprox \frac{\gamma_k}{||\mathbf{w}_k||^2} \triangleq \tilde{\gamma}_k \tag{4.3}$$

where we define

$$\gamma_k \triangleq \frac{\rho |\mathbf{u}_k^H \mathbf{H}_k \hat{\mathbf{v}}_k|^2}{1 + \rho ||\mathbf{u}_k^H \mathbf{H}_k - (\mathbf{u}_k^H \mathbf{H}_k \hat{\mathbf{v}}_k) \hat{\mathbf{v}}_k^H||^2} \tag{4.4}$$

as the CQI feedback with $\rho = \frac{P}{M}$. We recall that (4.4) represents the exact SINR of user k when the CDIs of the selected users form a set of M orthogonal vectors.

From (4.4), the linear detector $\mathbf{u}_k$ and the codebook vector $\hat{\mathbf{v}}_k$ are chosen according to

$$(\mathbf{u}_k, \hat{\mathbf{v}}_k) = \underset{\mathbf{u}_k \in \mathbb{C}^N, ||\mathbf{u}_k||^2 = 1, \mathbf{c}_i \in \mathcal{C}}{\arg\max} \gamma_{k,i}(\mathbf{u}_k, \mathbf{c}_i) \,, \tag{4.5}$$

where

$$\gamma_{k,i} = \frac{\mathbf{u}_k^H \mathbf{A}_k \mathbf{u}_k}{1 + \mathbf{u}_k^H \mathbf{B}_k \mathbf{u}_k} \tag{4.6}$$

[2] As in Chapter 3 this last restriction is sub-optimum but allows to derive a good approximation of the expected SINR of each user that otherwise would be difficult to predict.

and

$$\mathbf{A}_k = \rho\left(\mathbf{H}_k\mathbf{c}_i\mathbf{c}_i^H\mathbf{H}_k^H\right) \quad (4.7)$$

$$\mathbf{B}_k = \rho\left[\mathbf{H}_k(\mathbf{I}-\mathbf{c}_i\mathbf{c}_i^H)\mathbf{H}_k^H\right]. \quad (4.8)$$

We call the optimum detector $\mathbf{u}_k^H$ given by (4.5) the *maximum expected SINR combiner* (MESC). The maximizing arguments in (4.5) can be determined in a straightforward manner by considering all codewords $\mathbf{c}_i \in \mathcal{C}$. For a given codeword $\mathbf{c}_i$, the desired detector is the MMSE linear combiner which can be derived as

$$\mathbf{u}_k = (\mathbf{I}+\mathbf{B}_k)^{-1}\sqrt{\rho}\,\mathbf{H}_k\mathbf{c}_i, \quad (4.9)$$

normalized to unit norm. Then the resulting expected SINR becomes

$$\gamma_{k,i} = \rho\mathbf{c}_i^H\mathbf{H}_k^H(\mathbf{I}+\mathbf{B}_k)^{-1}\mathbf{H}_k\mathbf{c}_i. \quad (4.10)$$

We note that for the special case of low SNR, the interference becomes negligible compared to the thermal noise, and the matched filter $\mathbf{u}_k = \frac{\mathbf{H}_k\hat{\mathbf{v}}_k}{||\mathbf{H}_k\hat{\mathbf{v}}_k||}$ becomes the optimum receiver. In this case, (4.4) becomes

$$\text{Low SNR} \qquad \gamma_k = \rho\hat{\mathbf{v}}_k^H\mathbf{H}_k^H\mathbf{H}_k\hat{\mathbf{v}}_k\,. \quad (4.11)$$

From this observation, for asymptotically high B, the quantization vector that maximizes (4.11) would be, with very high probability, close to the direction of the dominant left singular vector of the channel matrix $\mathbf{H}_k$ which represents the quantization strategy proposed in [63]. Hence the proposed MESC reduces to choosing the codeword closest in direction to the dominant right singular vector of the channel matrix at low SNR and asymptotic high B.

For another special case of high SNR, thermal noise becomes negligible with respect to multiuser interference, and (4.4) reduces to

$$\text{High SNR} \qquad \gamma_k = \frac{|\tilde{\mathbf{v}}_k^H\hat{\mathbf{v}}_k|^2}{1-|\tilde{\mathbf{v}}_k^H\hat{\mathbf{v}}_k|^2} = \frac{\cos^2(\theta_k)}{\sin^2(\theta_k)} \quad (4.12)$$

where $cos(\theta_k) = |\tilde{\mathbf{v}}_k^H, \hat{\mathbf{v}}_k|$ and the expected SINR is maximized minimizing the angle θ_k between the normalized equivalent channel $\tilde{\mathbf{v}}_k$ and the quantized vector $\hat{\mathbf{v}}_k$. In other words the quantization vector is chosen as the codeword at minimum chordal distance from the space spanned by the rows of the channel matrix $\mathbf{H}_k$ completely neglecting the gain associated to the equivalent channel. This corresponds to the quantization-based combing (QBC) solution proposed in [62][3].

This analysis reveals how the proposed MESC, according to the receiver SNR and the number of feedback bits B, chooses the combining vector and the codeword that give the best

[3]We emphasize that if $M = N$ the channel vectors span $\mathbb{C}^M$ with probability one. Therefore each quantization vector has zero angle with the channel subspace and when using QBC there is no gain in increasing the feedback rate. Differently MESC accounts for the norm of the equivalent channel and even with $N \geq M$ is able to exploit additional feedback bits and provides significant gain over QBC as shown in Section 4.6.

trade-off between quantization accuracy (minimization of θ_k) and gain of the equivalent channel ($||\mathbf{v}_k||$) in order to maximize the expected SINR. For low SNR and high B, this corresponds to choosing the equivalent channel with the largest gain. Whereas at high SNR, quantization accuracy becomes the dominant factor.

4.2.2 Phase II: User selection and precoder determination at the transmitter

Using CDI vector and CQI feedback from all of the mobiles, the base station chooses a set $\mathcal{S}$ of users to serve based on a weighted sum rate given by

$$\tilde{\mathcal{R}}(\mathcal{S}) = \sum_{k \in \mathcal{S}} \alpha_k \log_2 (1 + \tilde{\gamma}_k) \tag{4.13}$$

where α_k is the QoS weight for the kth user and the CQI $\tilde{\gamma}_k$ is given by (4.3).

For a given set $\mathcal{S}$, we collect the CDI vectors of selected users in $\mathbf{\Lambda}(\mathcal{S}) = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_{|\mathcal{S}|}]^H$, and define matrix $\mathbf{W}(\mathcal{S}) = [\mathbf{w}_1, \ldots, \mathbf{w}_{|\mathcal{S}|}] = \mathbf{\Lambda}(\mathcal{S})^\dagger = \mathbf{\Lambda}(\mathcal{S})^H \left(\mathbf{\Lambda}(\mathcal{S})\mathbf{\Lambda}(\mathcal{S})^H\right)^{-1}$. The ZF beamforming vector $\mathbf{g}_k$ of the kth user is the kth column of the matrix $\mathbf{G}(\mathcal{S}) = [\mathbf{g}_1, \ldots, \mathbf{g}_{|\mathcal{S}|}] = \mathbf{W}(\mathcal{S})\text{diag}(\mathbf{p})^{1/2}$ where $\mathbf{p} = \left[p_1, \ldots, p_{|\mathcal{S}|}\right]$.

For determining set $\mathcal{S}$, we consider two different user selection schemes: 1) the greedy algorithm 1 (GUS) proposed in Section 3.4.2 and 2) the semi-orthogonal user selection (SUS) algorithm proposed in [42] (see also Section 3.4.1).

SUS is more amenable for analysis, and we use it Section 4.3. On the other hand the relative simplicity of GUS, makes it preferable to implement and we use it in Section 4.6 in numerical simulations.

4.2.3 Phase III: Data demodulation at the active receivers

Users in the active set $\mathcal{S}$, in order to perform coherent demodulation, need a coherent channel estimate of the beamformed SIMO channel which can be obtained by pilot sequences, sometimes known as *dedicated pilots*. A dedicated pilot is a training symbol transmitted to a user, using its designated beamforming vector. In turn, data demodulation can be performed using the MESC $\mathbf{u}_k$ from (4.5). Alternatively, as suggested in [63], each user could estimate the equivalent channels with respect to the precoded streams destined for other users and derive the MMSE combiner explicitly for the desired signal in the presence of this interference. Let $\mathbf{f}_k^{(j)} = \mathbf{H}_k \mathbf{g}_j$, in case of perfect channel estimation, the new MMSE combiner for user k is given by

$$\mathbf{u}_k = \left(\mathbf{I} + \sum_{j \in \mathcal{S}, j \neq k} \mathbf{f}_k^{(j)} \mathbf{f}_k^{(j)^H} \right)^{-1} \mathbf{f}_k^{(k)}, \tag{4.14}$$

normalized to have unit norm. We will investigate in Section 4.6 the conditions under which this additional MMSE processing might be useful.

4.3 Asymptotic analysis of MESC for $N < M$

In this section using RVQ we analytically characterize the performance of MESC with SUS. Because we are interested in cellular downlink systems, we study the case of many users $K \gg M$ and fewer receive antennas than transmit antennas, $N < M$. We assume each user channel matrix $\mathbf{H}_k$ has i.i.d complex zero-mean Gaussian entries with unit variance. Moreover to simplify analysis we assume that users with orthogonal CDIs are selected at transmitter and that (4.4) represents the effective SINR for user k. Under these assumptions and adopting MESC, the achievable SINR $\gamma_{k,i}$ of user k for a generic codeword $\mathbf{c}_i$ is given by (4.10) which can also be expressed as [69]

$$\gamma_{k,i} = \frac{1}{\left[\left(\mathbf{I} + \mathbf{G}^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{G}\right)^{-1}\right]_{k,k}} - 1 \tag{4.15}$$

where $\mathbf{G}$ is the unitary beamformer and $\mathbf{g}_k = \sqrt{\rho}\mathbf{c}_i$ is the precoding vector for user k. In the following we characterize the probability density function (PDF) of (4.15). To this aim, since $\mathbf{N}_k = \mathbf{H}_k^H \mathbf{H}_k$ is a central complex Wishart matrix with N degrees of freedom and covariance matrix $\mathbf{I}$, ($\mathbf{N}_k \sim \mathcal{W}(N, \mathbf{I})$), it is unitary invariant [70], i.e. $\mathbf{G}^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{G}$ has the same distribution of $\rho \mathbf{H}_k^H \mathbf{H}_k$. As a consequence the distribution of $\gamma_{k,i}$ is equivalent to the distribution of $\bar{\gamma}_{k,i} = \frac{1}{\left[\left(\mathbf{I}+\rho \mathbf{H}_k^H \mathbf{H}_k\right)^{-1}\right]_{k,k}} - 1$. The cumulative density function (CDF) of $\bar{\gamma}_{k,i}$ can be derived from [71] and under the assumptions of equal power allocation we get

$$\begin{aligned} F_{\gamma_{k,i}}(z) &= F_{\bar{\gamma}_{k,i}}(z) = P[\bar{\gamma}_{k,i} \leq z] \\ &= 1 - e^{-\frac{z}{\rho}} \sum_{n=1}^{N} \frac{E_n(z)}{(n-1)!} \left(\frac{z}{\rho}\right)^{n-1} \end{aligned} \tag{4.16}$$

where

$$E_n(z) = \begin{cases} 1 & N \geq M - 1 + n \\ \sum_{i=0}^{N-n} \binom{M-1}{i} z^i & \\ \frac{}{(1+z)^{M-1}} & N < M - 1 + n \end{cases}. \tag{4.17}$$

We underline that the analysis developed in this section explicitly refers to the more practical case $N < M$ in which (4.16) becomes

$$F_{\gamma_{k,i}}(z) = F_{\bar{\gamma}_{k,i}}(z) = 1 - \frac{e^{-\frac{z}{\rho}}}{(1+z)^{M-1}} L(z) \tag{4.18}$$

with

$$L(z) = \sum_{n=1}^{N} \frac{\sum_{i=0}^{N-n} \binom{M-1}{i} z^i}{(n-1)!} \left(\frac{z}{\rho}\right)^{n-1}. \tag{4.19}$$

The following theorem derives the CDF of γ_k for the best codeword $\hat{\mathbf{v}}_k$.

Theorem 2 *In a system with $N < M$ where channel quantization is performed according to*

(4.5), using RVQ with B feedback bits, the CDF of CQI γ_k is given by

$$F_{\gamma_k}(z) = 1 - \frac{\binom{M-1}{N-1} 2^B e^{-\frac{z}{\rho}}}{(1+z)^{M-N}}\,, \quad z = O(\rho) \gg 1\,. \tag{4.20}$$

Proof: Using Taylor series expansion we give an approximated characterization of $F_{\gamma_{k,i}}(z)$ valid for $z, \rho \gg 1$. Indeed for $L(z)$ we get

$$\begin{aligned} L(z) &= \sum_{n=1}^{N} \frac{\binom{M-1}{N-n} z^{N-n} + o\left(z^{N-n}\right)}{(n-1)!} \left(\frac{z}{\rho}\right)^{n-1} \\ &= z^{N-1} \sum_{n=1}^{N} \frac{\binom{M-1}{N-n} + o(1)}{(n-1)!\rho^{n-1}} \\ &= z^{N-1} \binom{M-1}{N-1} + o\left(z^{N-1}\right) \end{aligned} \tag{4.21}$$

and still applying Taylor approximation $z^{M-1} = (1+z)^{M-1} + o(z^{M-1})$, the CDF of $\gamma_{k,i}$ for $z, \rho \gg 1$ becomes

$$F_{\gamma_{k,i}}(z) = 1 - \frac{\binom{M-1}{N-1} e^{-\frac{z}{\rho}}}{(1+z)^{M-N}}\,. \tag{4.22}$$

Since for RVQ the codebook is composed by 2^B independent random vectors, γ_k is the maximum among 2^B i.i.d. random variables distributed according to (4.18) and its CDF results

$$F_{\gamma_k}(z) = P\left[\max_{\{\gamma_{k,i}\}_{i=1}^{2^B}} \gamma_{k,i} \leq z\right] = \left[F_{\gamma_{k,i}}(z)\right]^{2^B}\,. \tag{4.23}$$

Again using Taylor series expansion for $z = O(\rho) \gg 1$ from (4.22) and (4.23) we finally get (4.20). □

Theorem 2 gives the distribution of SINR of any given user but to find an expression of the achievable throughput, we need the distribution of γ_k for a *selected* user which depends on the user selection algorithm. We recall that according to the SUS algorithm, the $(i+1)$th selected user has the highest γ_k among $|\mathcal{A}^{(i)}|$ users with independent channels and the same average SNR. Let $\gamma_{i:U}$ be the ith largest order statistic among U i.i.d. random variables $\{\gamma_k\}$. The selection of user $i+1$ can be seen as the selection of the $(i+1)$th largest order statistic in a set with $U^{(i)} = |\mathcal{A}^{(i)}| + i$ elements all having the same statistic. An approximated expression for the achievable throughput when using MESC and SUS is given by

$$\mathrm{E}[R] \;\simeq\; \mathrm{E}\left[\sum_{i=1}^{M} \log_2\left(1 + \gamma_{i:U^{(i-1)}}\right)\right]\,, \tag{4.24}$$

where the approximation takes into account that (4.4) is the SINR after beamforming in case M users with orthogonal CDIs are selected at transmitter. In the following we derive an approximation of (4.24) in case of many users K. Applying the law of large numbers it's easy to see that $U^{(i)} \simeq K\alpha_i + O(1)$ where α_i is the probability that a user belongs to $\mathcal{A}^{(i)}$. Moreover

from [42, Theorem 1] and (4.20) one can easily derive an approximation of (4.24) in the case of many users as

$$\mathrm{E}[R] \simeq \sum_{i=1}^{M} \log_2 \left(1 + \rho \log \frac{\binom{M-1}{N-1} 2^B K \alpha_{i-1}}{\rho^{M-N}} \right) . \tag{4.25}$$

The logarithmic term $\Delta = \log \frac{\binom{M-1}{N-1} 2^B K \alpha_{i-1}}{\rho^{M-N}}$ in (4.25) can be interpreted as the SNR variation, which includes the effects of both quantization error and multiuser diversity. Interestingly for a given Δ that assures a constant gap from ZF beamforming with perfect CSIT, B and K should scale with P as

$$B + \log_2 K = (M - N) \log_2 P + c \, , \tag{4.26}$$

where $c = -(M - N) \log_2 M - \log_2 \binom{M-1}{N-1} + d$ is monotonically increasing with N and d depends on Δ. We observe that for a given K the scaling of B with $\log_2 P \simeq P_{dB}/3$ has a smaller slope with increasing N, meaning that the increment of the feedback rate necessary to assure a constant gap from ZF beamforming with perfect CSIT is smaller for increasing N. Moreover as in [42], quantities 2^B and K are interchangeable. Hence for a target sum rate, every doubling of the number of users saves one feedback bit per user. This result naturally extends the scaling rule derived in [42] for $N = 1$, showing the benefit of multiple receive antennas in systems with limited uplink feedback. Furthermore it also generalizes the results obtained in [62] for QBC in case of random or no user selection, showing how multiuser diversity can be exploited to reduce feedback overhead.

It is important to observe that (4.20), (4.25)-(4.26) are valid only in a *large user regime* as $K \to \infty$. For finite K and B, if P is too large, the system enters the *interference-limited regime*. Applying similar arguments of Theorem 2, the CDF of γ_k simplifies to

$$F_{\gamma_k}(z) = 1 - \frac{\binom{M-1}{N-1} 2^B}{(1+z)^{M-N}} , \quad z \gg 1 . \tag{4.27}$$

From [42, Theorem 2] and (4.27) an approximation of (4.24) in case of many users is

$$\begin{aligned} \mathrm{E}[R] &\simeq \frac{M}{M-N}(B + \log_2 K) + \\ &+ \frac{M \log_2 \binom{M-1}{N-1} + \sum_{i=1}^{M} \log_2 \alpha_{i-1}}{M-N} . \end{aligned} \tag{4.28}$$

Under finite B and $K \gg 1$ we see that the sum rate eventually converges to a constant value as given by (4.28) which increases with N.

4.4 Codebook Design based on the LBG algorithm

In this section we investigate the problem of codebook design assuming that the MESC algorithm described in Section 4.2 is used by each receiver. We design the codebook using the LBG algorithm described in Section 3.3.1 and repropose BFB and HFB as feedback strategies.

Here HFB is preferred to other predictive FB strategies, e.g. QEV and RM, because we are interested in low-complexity solutions and low FB rates.

Let $\mu(\mathbf{H}, \mathbf{c}_i)$ be a performance metric to be used in (3.31) with $\mathbf{H}$ substituting $\mathbf{h}$. We note that in contrast to the single-antenna case the optimization criterion for $N > 1$ has to involve the computation of the receiver combiner $\mathbf{u}_k$, which depends on the quantization codebook and the channel realization.

4.4.1 Performance metrics

A direct optimization of the quantization codebook based on (4.10) could be done by numerical methods, however the complexity is very high and there is no guarantee of convergence. Moreover (4.10) depends on the transmission power P, hence the derivation of the optimum codebook should be related to this parameter and different codebooks would be necessary for different SNR values. These motivate the derivation of a simplified performance metric as an approximation of (4.10). Indeed we consider two limit situations, low-SNR and high-SNR regions and derive suboptimal but practical performance metrics.

Codebook design for low-SNR

In the low-SNR region when interference is negligible with respect to noise, the expected SINR for the kth user can be approximated as

$$\text{Low SNR} \qquad \mathrm{E}\left[\mathrm{SINR}_k\right] = \rho\, \mathrm{E}\left[|\mathbf{u}_k^H \mathbf{H}_k \mathbf{c}|^2\right] \tag{4.29}$$

where $\mathbf{c} = \mathcal{Q}[\mathbf{H}_k, \mathbf{u}_k]$. Hence we adopt the performance metric $\mu_1(\mathbf{H}_k, \mathbf{c}) = |\mathbf{u}_k^H \mathbf{H}_k \mathbf{c}|^2$ also because the maximization problem does not depend on the constant factor ρ.

We notice that in the low SNR region the problem of codebook design simplifies to finding the best codewords for a SU-MIMO system. In this case, as stated in [46], [47], if transmit antennas are uncorrelated receive correlation and the number of receive antennas do not influence the problem of codebook design and the same codebook used for spatially uncorrelated channels may be used. The same arguments are not applicable in case of correlated transmit antennas where not only the transmit correlation matrix but also the receive correlation matrix and the number of receive antennas have to be taken into account in codebook design. The proposal applies to the general case of possible transmit and receive antennas correlation.

Since in the low-SNR region for a given codeword $\mathbf{c}$ and a channel matrix $\mathbf{H}_k$ the optimum combiner is given by the MRC with respect to the equivalent channel $\mathbf{H}_k\mathbf{c}$, we can rewrite the performance metric as

$$\mu_1(\mathbf{H}_k, \mathbf{c}) = \frac{|\mathbf{c}^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{c}|^2}{||\mathbf{H}_k \mathbf{c}||^2} = \mathbf{c}^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{c}\,. \tag{4.30}$$

And the optimum codeword for the partition region $\mathcal{R}_i$ results the dominant eigenvector of the matrix $\sum_{\mathbf{H}_k \in R_i} \mathbf{H}_k^H \mathbf{H}_k$ normalized to unit norm.

Codebook design for high-SNR

In the high-SNR region when noise is negligible with respect to multiuser interference from the expectation of (4.12) the expected SINR is given by

$$\text{High SNR} \qquad \text{E}[\text{SINR}_k] = \text{E}\left[\frac{|\tilde{\mathbf{v}}_k^H\mathbf{c}|^2}{1-|\tilde{\mathbf{v}}_k^H\mathbf{c}|^2}\right] \tag{4.31}$$

$$\geq \frac{\text{E}\left[|\tilde{\mathbf{v}}_k^H\mathbf{c}|^2\right]}{1-\text{E}\left[|\tilde{\mathbf{v}}_k^H\mathbf{c}|^2\right]}, \tag{4.32}$$

where (4.32) follows from Jensen's inequality applied to the convex function γ_k in the high SNR region. Observing that the maximization of (4.32) is achieved simply maximizing its numerator, a suboptimum performance metric in the high-SNR scenario, still independent of P, is given by

$$\mu_2(\mathbf{H}_k,\mathbf{c}) = |\tilde{\mathbf{v}}_k^H\mathbf{c}|^2 = \left|\frac{\mathbf{u}_k^H\mathbf{H}_k}{||\mathbf{u}_k^H\mathbf{H}_k||}\mathbf{c}\right|^2 = \frac{\mathbf{u}_k^H\mathbf{C}_k\mathbf{u}_k}{\mathbf{u}_k^H\mathbf{D}_k\mathbf{u}_k} \tag{4.33}$$

where $\mathbf{C}_k = \mathbf{H}_k\mathbf{c}\mathbf{c}^H\mathbf{H}_k^H$ and $\mathbf{D}_k = \mathbf{H}_k\mathbf{H}_k^H$. Given a certain codeword $\mathbf{c}$, the optimum combiner for a given channel realization is easily derived as a scaled version of $\mathbf{u}_k = \mathbf{D}_k^{-1}\mathbf{H}_k\mathbf{c}$ and the performance metric simplifies to

$$\mu_2(\mathbf{H}_k,\mathbf{c}) = \left|\frac{\mathbf{c}^H\mathbf{H}_k^H(\mathbf{D}_k)^{-1}\mathbf{H}_k}{||\mathbf{c}^H\mathbf{H}_k^H(\mathbf{D}_k)^{-1}\mathbf{H}_k||}\mathbf{c}\right|^2$$

$$= \mathbf{c}^H\mathbf{H}_k^H\left(\mathbf{D}_k\right)^{-1}\mathbf{H}_k\mathbf{c}\,. \tag{4.34}$$

As a consequence, the optimum codeword $\mathbf{c}_i$ for the partition region $\mathcal{R}_i$ results the dominant eigenvector of the matrix $\sum_{\mathbf{H}_k\in R_i}\mathbf{H}_k^H\left(\mathbf{D}_k\right)^{-1}\mathbf{H}_k$ normalized to unit norm.

4.4.2 LBG-based codebook with tree structure and hierarchical feedback

As in Section 3.3.2 we generate the quantization codebook by a tree structure. Differently from Section 3.3.2 we initialize the LBG algorithm by considering at level $L = \log_2 M$ a codebook composed of M orthogonal unit norm vectors. In particular we adopt the discrete Fourier transform (DFT) matrix which is an optimum codebook for both an i.i.d and line-of-sight (LOS) channel. It can be shown that with this approach we improve the achievable sum rate for low FB rates with respect to the technique described in Section 3.3.2, especially for $N > 1$.

We notice that for $N > 1$ the HFB signalling follows the same rules described in Section 3.3.2 with $\bar{\mathbf{h}}_k$ and $\hat{\mathbf{h}}_k$ substituted by $\bar{\mathbf{v}}_k$ and $\hat{\mathbf{v}}_k$, respectively.

4.5 Unitary beamforming with MMSE receiver

In case of U-BF if exactly M users are selected at transmitter, each one can perfectly estimate its achievable SINR even without knowing the precoding vectors of the other selected users and as anticipated in Section 4.2.1, (4.4) becomes the exact SINR for user k. Hence differently

from ZF-BF where user selection is based on estimated achievable SINRs, with U-BF user selection can rely on an exact computation of the achievable rates.

We propose a generalization of [22] that considers a set of unitary matrices as tentative beamformers and selects both users and precoder in order to maximize the weighted achievable throughput. In detail the codebook is composed of 2^B unit vectors organized in sets of M vectors forming a number of unitary matrices. Each user, assuming an MMSE receiver and M selected users as in (4.4), feeds back the index of the precoding vector that provides the highest achievable SINR[4]. Then the transmitter selects for each precoding vector the user with the highest weighted rate and chooses among the possible unitary matrices the one that provides the highest weighted throughput. This proposal generalizes the transmission scheme [22] in the design of the quantization codebook.

4.5.1 Codebook of unitary matrices

According to the transceiver scheme, codebook design for CDI quantization has to address two main issues: i) match the statistic of the channel model and ii) comprise a number of unitary matrices. In [22] there is a first attempt to address this problem with a codebook composed by $2^B/M$ unitary matrices, derived from rotations of the $M \times M$ discrete Fourier transform (DFT) matrix. This scheme shows to be efficient in case of high correlation among transmit antennas but provides worse performance with spatially uncorrelated channels. In particular for a given B the scheme strongly depends on the number of users K. Indeed as K decreases there is a corresponding increment of the probability that some vectors of each unitary matrix are not selected by any user. In other words for small K the transmitter serves with high probability fewer than M users, reducing the multiplexing gain of the system.

Focusing on spatially uncorrelated channels we propose two new codebooks that better exploit the channel statistic and comprise a larger number of unitary matrices for a given B. This last feature aims at reducing the impact of K on the transmission scheme allowing the selection the precoding matrix in a larger set. The proposed strategies are outlined in case of $M = 4$, but they can be easily generalized to any $M > 2$.

First Algorithm

Let $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_{2^B}\}$ denote the codebook of 2^B codevectors for CDI quantization which we are going to derive. In the initialization step ($i = 0$) of the algorithm the first four codevectors are chosen to form a unitary matrix where each vector is generated according to an isotropic distribution, [49]. Next, at step $i = 1, \ldots, 2^B/4 - 1$ we randomly select two couples of vectors from the set $\{\mathbf{c}_{4(i-1)+1}, \ldots, \mathbf{c}_{4i}\}$ and design two orthogonal subspaces in $\mathbb{C}^4$, e.g $\{\mathbf{c}_{4(i-1)+1}, \mathbf{c}_{4(i-1)+2}\}$ and $\{\mathbf{c}_{4(i-1)+3}, \mathbf{c}_{4i}\}$. Then a different basis is generated for each subspace that we denote as $\{\mathbf{c}_{4i+1}, \mathbf{c}_{4i+2}\}$ and $\{\mathbf{c}_{4i+3}, \mathbf{c}_{4(i+1)}\}$, respectively. The set $\{\mathbf{c}_{4i+1}, \ldots, \mathbf{c}_{4(i+1)}\}$ still represents a basis for $\mathbb{C}^4$ but at the same time we generated two other

[4]We notice that for a given B, ZF-BF and U-BF with BFB have the same computation complexity at the receiver with regard to the choice of CDI, CQI and combiner, as the quantizer performs an exhaustive search among all codevectors using (4.5), with a complexity $O(2^B)$. Differently ZF-BF-HFB with tree search has a linear complexity with B, i.e. $O(2B)$, but requires a larger memory.

basis $\{\mathbf{c}_{4(i-1)+1}, \mathbf{c}_{4(i-1)+2}, \mathbf{c}_{4i+3}, \mathbf{c}_{4(i+1)}\}$ and $\{\mathbf{c}_{4(i-1)+3}, \mathbf{c}_{4i)}, \mathbf{c}_{4i+1}, \mathbf{c}_{4i+2}\}$. With this approach at each step $i > 0$ of the algorithm we include three unitary matrices in the codebook, while only one matrix is added in the initialization step. It's easy to see that the total number of unitary matrices in the codebook is

$$N_M^{(1)} = 1 + 3\left(2^B/4 - 1\right) . \tag{4.35}$$

Notice that, differently from [22], each codeword with index in the set $\mathcal{A} = \{1, 2, 3, 4, 2^B, 2^B - 1, 2^B - 2, 2^B - 3\}$ belongs to two different tentative beamformers, e.g. $\mathbf{c}_1$ is a precoding vector for both $[\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4]$ and $[\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_7, \mathbf{c}_8]$ in the example. Moreover any other codeword in $\mathcal{C}$ with index not in $\mathcal{A}$ belongs to three different unitary matrices. We underline that thanks to the unitary constraint a user can perfectly estimate its achievable SINR when choosing a codeword, even without knowing the beamforming matrix selected at transmitter.

Second algorithm

Let $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_{2^B}\}$ denote the codebook of 2^B codevectors for CDI quantization which we are going to derive. In the initialization step ($i = 0$) of the algorithm the first four codevectors are chosen to form a unitary matrix with each vector generated according to an isotropic distribution, [49]. Next we randomly select two couples of vectors from the set $\{\mathbf{c}_1, \ldots, \mathbf{c}_4\}$ that design two orthogonal subspaces in $\mathbb{C}^4$, e.g $\{\mathbf{c}_1, \mathbf{c}_2\}$ and $\{\mathbf{c}_3, \mathbf{c}_4\}$. At each step $i > 0$ of the algorithm we add four new vectors forming a unitary matrix and where $\{\mathbf{c}_{4i+1}, \mathbf{c}_{4i+2}\}$ and $\{\mathbf{c}_{4i+3}, \mathbf{c}_{4i+4}\}$ form new basis for the subspaces designed by $\{\mathbf{c}_1, \mathbf{c}_2\}$ and $\{\mathbf{c}_3, \mathbf{c}_4\}$, respectively. With this alternative codebook design algorithm each codeword belongs to $2^B/M$ different unitary matrices. For instance with $B = 3$ codeword $\mathbf{c}_1$ belongs to the following $2^B/M = 4$ unitary matrices: $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4\}$, $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_7, \mathbf{c}_8\}$, $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_{11}, \mathbf{c}_{12}\}$ and $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_{15}, \mathbf{c}_{16}\}$. In this case the total number of unitary matrices is given by

$$N_M^{(2)} = \left(2^B/M\right)^2 \tag{4.36}$$

We note that a generalization of the proposed algorithm to a system with $M > 4$ can consider the split of the initial unitary matrix in more than two subspaces, yielding to an increment of the number of unitary matrices each codeword belongs to. We notice that even if with this second algorithm the codebook contains a larger number of tentative beamformers, it has a partial drawback. Indeed each codeword is always associated to another codeword in each tentative beamformer it belongs to. For instance $\mathbf{c}_1$ and $\mathbf{c}_2$ are always associated in the previous example. This reduces the flexibility of the user selection and can have a negative effect on the achievable throughput when the scheduling algorithm try to guarantee long term fairness among users (see also Chapter 5).

4.6 Simulation results

We consider a transmitter with $M = 4$ antennas and for the purpose of this first comparison we consider a Rayleigh fading channel model with i.i.d elements $\sim \mathcal{CN}(0, 1)$. We adopt the

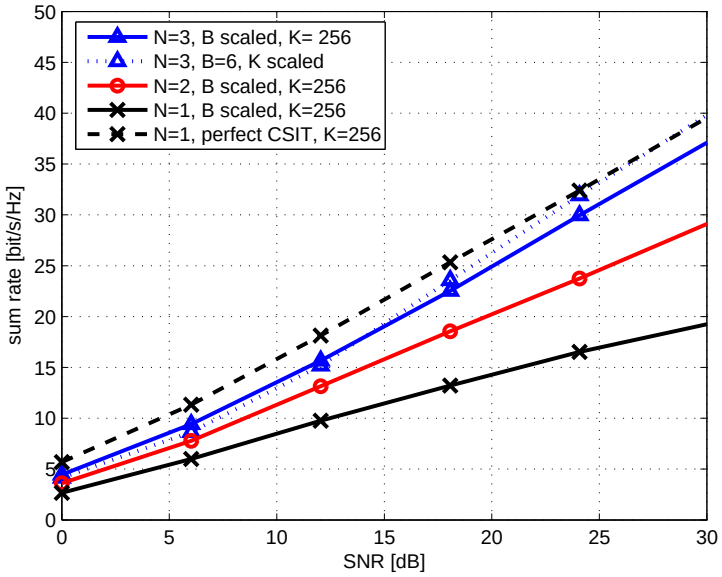

Figure 4.1: Effects of of receive antennas and multiuser diversity in the achievable throughput. MESC with SUS and RVQ. M=4, i.i.d. Rayleigh fading channel.

proposed MESC, a simple RVQ and the SUS algorithm with correlation parameter $\epsilon = 0.3$. In Fig. 4.1 we show how scaling the number of feedback bits B according to (4.26) with $c = 9$, $K = 256$ and $N = 3$, we can assure a constant gap in terms of sum rate from the curve of perfect CSIT. Differently using the same amount of feedback bits in a system with $N = 2$ or $N = 1$ is not enough to keep a constant gap from the upper bound. This could be achieved only using $N = 2$ or $N = 1$ in (4.26) and scaling B accordingly. Moreover we show how quantities 2^B and K are interchangeable. Indeed setting $B = 6$ and scaling K according to (4.26) for $N = 3$ still provides almost a constant gap from ZF with perfect CSIT.

In Figs. 4.2 and 4.3 we compare the achievable sum rate of MESC, MRC and QBC proposed in [62] as a function of SNR, for $K = 20$ users, $B = 4$ feedback bits and $N = 2, 4$ receive antennas. We still use an i.i.d. channel model and RVQ but adopt the GUS algorithm because as verified by simulations and shown in [50] it generally provides higher performance than SUS and is more practical, not requiring the optimization of the correlation parameter ϵ. As an upper bound under perfect CSIT, we show the performance of ZF-1 with optimum water-filling power allocation introduced in Chapter 2. The proposed MESC generally outperforms the other two schemes in all the SNR region with MRC approaching the proposed solution in the low SNR while QBC in the high SNR region. We notice that the gap between MESC and QBC is very small for $N = 2$ but strongly increases with more receive antennas. Moreover with $N = 2$ and fixed B, K the system becomes interference limited at high SNR as suggested by (4.28). It is also interesting to observe that designing in the second phase a new MMSE combiner for the selected users does not provide significant gain for QBC and the proposed solution. Indeed, the combiner designed in the first phase for the kth selected user would represent the MMSE combiner if the CDI of the selected users were orthogonal. As long as the

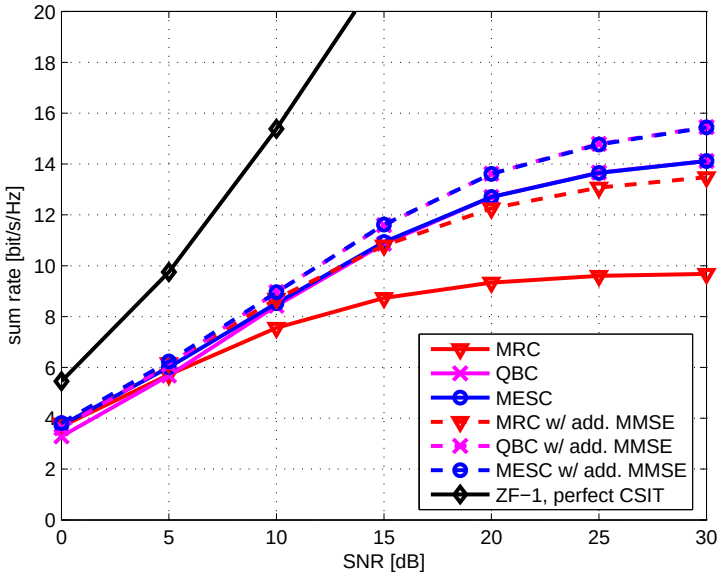

Figure 4.2: Comparison between different combining schemes: MRC, QBC, MESC. $M = 4$*,* $N = 2$*,* $K = 20$*,* $B = 4$*, RVQ.*

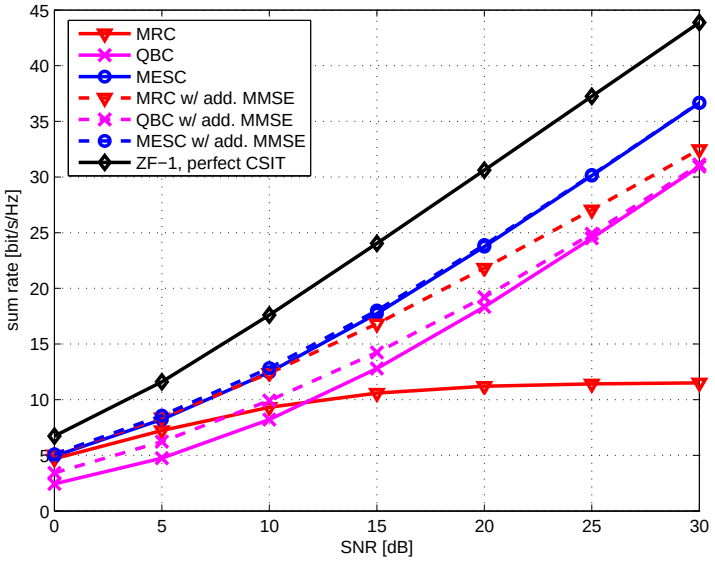

Figure 4.3: Comparison between different combining schemes: MRC, QBC, MESC. $M = 4$*,* $N = 4$*,* $K = 20$*,* $B = 4$*, RVQ.*

number of users in the system is much greater than the number of transmit antennas, the user selection algorithm used at the transmitter tries to select almost orthogonal users so that the combiner designed in the first phase becomes a close approximation of the MMSE combiner for the set of selected users.

In Fig. 4.4 using the proposed MESC we compare the performance achievable with RVQ

and the new LBG-based codebooks designed for both low (LBG-L) and high (LBG-H) SNR regimes. The transmitter is still equipped with $M = 4$ antennas, $K = 20$ users are present in the system with $N = 2$ antennas each and $B = 3$ bits are fed back by each receiver. We analyze three different channel models: the i.i.d Rayleigh fading and two spatially correlated channel models. For these last two models the transmit antennas are spaced by $\Delta_{TX} = 10$ and $2\ \lambda$, respectively, where $\lambda = c/f_c$ is the wavelength, $c = 3 \times 10^8$ m/s is the speed of light and $f_c = 2$ GHz is the carrier frequency. The channel is modelled as time-variant, flat Rayleigh fading, according to the spatial channel model (SCM) [60]. The time slot duration is $T = 0.5$ ms and the receive antennas of each mobile are spaced $\Delta_{RX} = 0.5\ \lambda$. First we observe as the LBG codebooks outperform RVQ both for independent and spatially correlated channels. Increasing the transmit antenna correlation generally reduces the spatial diversity of the system, nevertheless the proposed transmission scheme, exploiting multiuser diversity and the spatial correlation of the MIMO channel through the LBG-based codebooks, provides better performance increasing transmit antenna correlation. Interestingly the performance metric used in the design of the LBG codebook does not affect significantly system performance, anyway using LBG-L is preferable in the low SNR regime while LBG-H provides better performance for higher SNR.

The beneficial effect of transmit antenna correlation for the proposed ZF-based transmission schemes is also emphasized in Fig. 4.5. In the cases of both perfect and limited channel state information at the transmitter, increasing transmit antenna correlation provides performance improvement when multiuser diversity is available. Especially under limited feedback from receivers, transmit antenna correlation reduces the dimension of the hyperspace to quantize and for a given number of bits the resulting codebook better characterizes the space of the MIMO channel realizations. Hence even if higher transmit antenna correlation implies less spatial diversity, the higher quantization accuracy and the possibility of exploiting multiuser diversity provides significant performance improvement over uncorrelated channels.

In Figs. 4.6 and 4.7 we compare MESC-RVQ, MESC-LBG-L and U-BF for increasing number of feedback bits B. We set $M = 4$, $N = 2, 4$ and adopt an i.i.d. Rayleigh fading channel and the SCM with $\Delta_{TX} = 10\ \lambda$, $\Delta_{RX} = 0.5\ \lambda$ and $f_c = 2$ GHz, respectively in Figs. 4.6 and 4.7. We fix an average SNR $= 10$ dB and comparison is performed in term of achievable sum rate. As an upper bound we also include ZF-1 with perfect CSIT. We observe in Fig. 4.7 as MESC with the LBG-based codebook is able to exploit the spatial correlation of the MIMO channel providing higher performance than MESC with RVQ and maintaining a performance gap even when the number of feedback bits B increases and both schemes approach the upper bound provided by ZF-1 with perfect CSIT. As shown in Fig. 4.6 this gap vanishes in spatially uncorrelated channels with increasing B. Differently from MESC, the sum rate achievable with U-BF degrades with higher FB rates. Indeed for a given number of users in the network, increasing B has two opposite consequences: i) higher channel quantization accuracy ii) increment of the probability that some vectors of each unitary matrix are not selected by any user. While i) is beneficial for the achievable throughput, with ii) the transmitter serves with high probability fewer than M users reducing the multiplexing gain of the system. For a practical number of users in the network i) is dominant for small B, but eventually, when increasing the FB rate, ii) causes a performance degradation. Interestingly thanks to an higher number

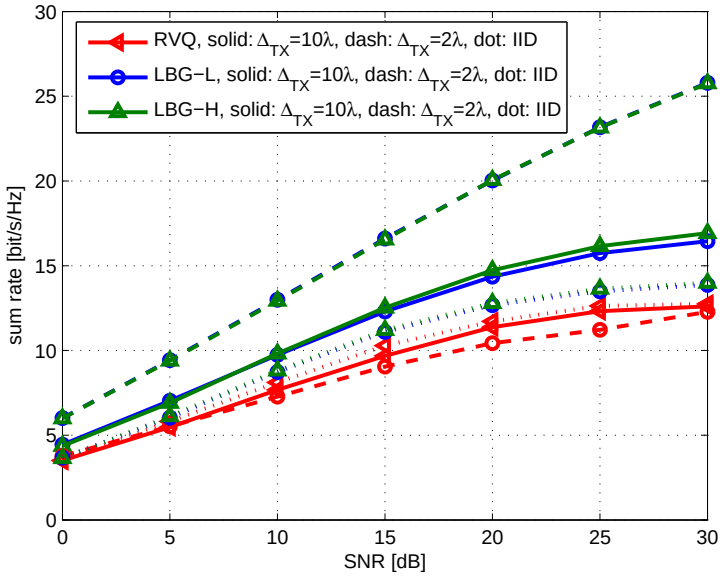

Figure 4.4: Comparison between different LBG based codebooks and RVQ. $M = 4$*,* $N = 2$*,* $K = 20$*,* $B = 3$*, MESC, i.i.d. Rayleigh fading channel model and SCM with* $\Delta_{TX} = 10, 2\ \lambda$

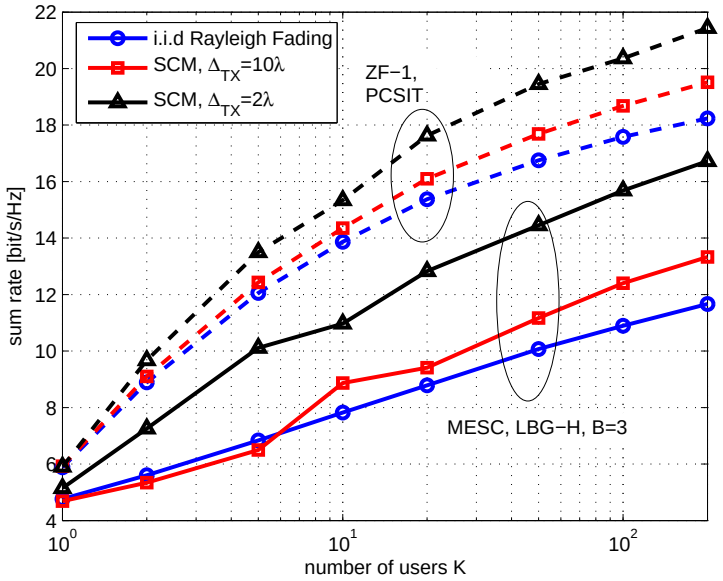

Figure 4.5: Effects of transmit antenna correlation and multiuser diversity. $M = 4$*,* $N = 2$*,* $SNR = 10$ *dB.*

of unitary matrices in the codebook, U-BF with the second proposed codebook mitigates the self-defeating effect of ii) and outperforms both PU2RC and U-BF with the first codebook. Moreover for spatially uncorrelated channels and low FB rates U-BF is generally preferable to ZF-BF thanks to a better estimation of the achievable SINR in the user selection process. We emphasize that all schemes benefit from the additional degrees of freedom provided by multiple

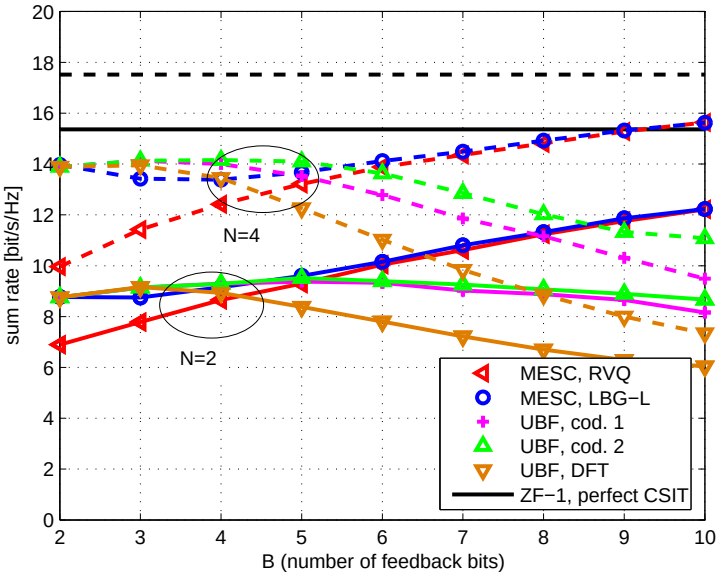

Figure 4.6: Comparison between ZF-MESC with LBG based codebooks or RVQ and Unitary BF. $M = 4$, $K = 20$, SNR $= 10$ dB, i.i.d. Rayleigh fading channel.

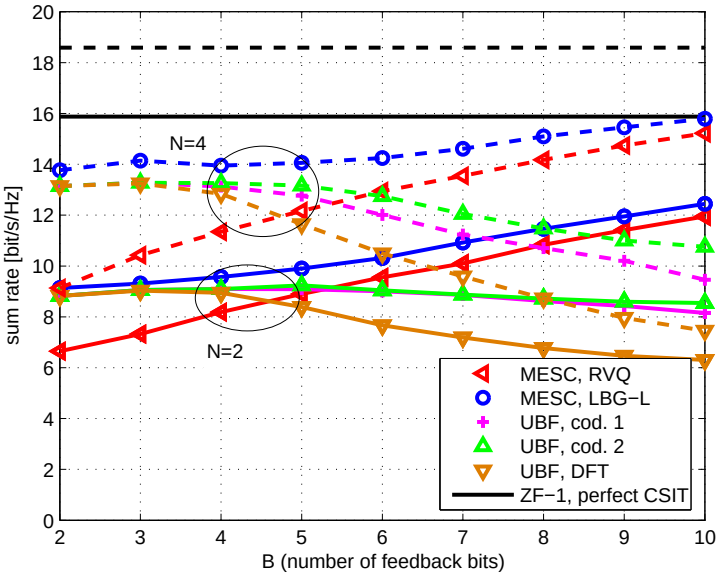

Figure 4.7: Comparison between ZF-MESC with LBG based codebooks or RVQ and Unitary BF. $M = 4$, $K = 20$, SNR $= 10$ dB, SCM with $\Delta_{TX} = 10\ \lambda$ and $\Delta_{RX} = 0.5\ \lambda$.

receive antennas thanks to MESC or MMSE receiver that attenuates multiuser interference.

We notice from Fig. 4.6 that all proposals are equivalent for $B = \log_2 M$. Indeed for both schemes the codebook is a unitary matrix and the combiner is computed as the MMSE combiner assuming the CDIs of the selected users form a set of M orthogonal vectors. We underline that among the unitary beamforming strategies best performance is achieved using

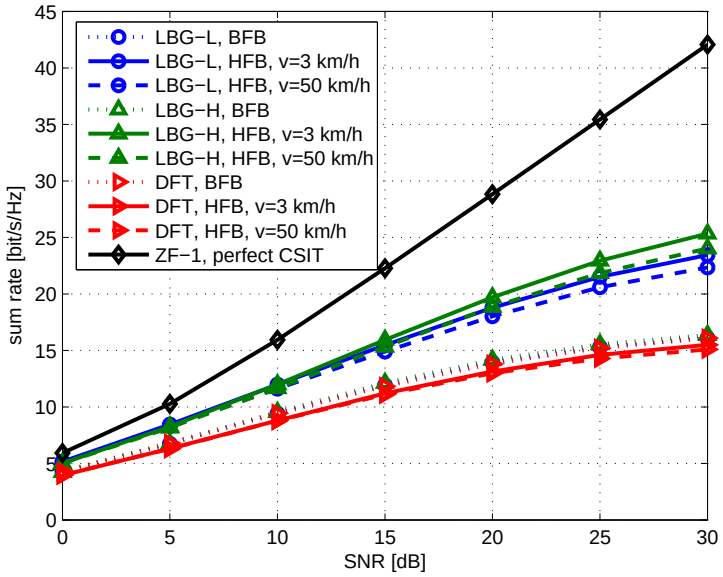

Figure 4.8: Comparison between hierarchical (HFB) and basic (BFB) feedback signalling. $M = 4$, $N = 2$, $K = 20$, $B = 3$ feedback bits per interval, $B_{max} = 12$, SCM with $\Delta_{TX} = 10\ \lambda$ and $\Delta_{RX} = 0.5\ \lambda$. Comparisons are between the LBG codebooks at low (L) and high (H) SNR and the DFT codebook given in [72], when using MESC.

the second codebook design algorithm (see Section 4.5.1).

Finally in Fig. 4.8 we show the performance impact of exploiting time correlation for LBG codebooks designed with a hierarchical structure. A codebook is designed with a depth of $B_{max} = 12$ levels and is indexed using $B = 3$ bits per interval. We compare the proposed solution using the MESC and a hierarchical LBG codebook with the hierarchical DFT codebook based on [72]. Since the channel is lightly correlated the performance of hierarchical DFT is poor while the proposed LBG codebooks used jointly with hierarchical feedback (HFB) provide significant gain with respect to the basic feedback strategy (BFB), even when the speed of the mobiles increases. Still we notice that LBG-H is preferable at high SNR. We observe that the hierarchical approach used with the DFT codebook results in a degradation of system performance. This is mainly due to the tree search applied with the new CDI quantization strategy that does not guarantee a maximum likelihood quantization as a brute force approach would do. Indeed since the DFT codebook does not effectively characterize the channel statistics, even increasing its effective size using the hierarchical feedback does not provide a sufficient gain in quantization accuracy to compensate the loss due to the suboptimum tree search.

4.7 Conclusions

This chapter considers the problems of i) transceiver design and ii) channel quantization in a multiuser MIMO downlink system with limited uplink FB where, differently from Chapter 3, users are equipped with multiple antennas. These provide additional degrees of freedom that are exploited to enhance the achievable throughput for a given feedback rate. We consider two

different solutions based on ZF and unitary beamforming, respectively, and allocating in both cases at most one stream per selected user.

Under ZF-BF the proposal jointly designs the receive combiner at each mobile and the feedback information in order to maximize the expected achievable SINR. For this proposal, known as the maximum expected SINR combiner (MESC), we provide an analytic characterization of the achievable throughput in the case of many users and show how additional receive antennas or higher multiuser diversity can reduce the required feedback rate to achieve a target throughput. Simulations results show that MESC generally outperforms other schemes known in literature. Under ZF-BF, codebook design is based on two new performance metrics derived for low and high SNR. The resulting quantization codebook exploits the spatial correlation of the MIMO channel and leads to a simple generalization of the FB strategies BFB and HFB introduced in Chapter 3. Numerical results show how antenna correlation in conjunction with multiuser diversity can provide performance improvement over uncorrelated channels in multiuser MIMO systems with limited feedback.

Under unitary BF users can have a perfect estimation of the achievable SINR adopting an MMSE combiner. This property is exploited in the proposal of two codebooks comprising an high number of tentative unitary beamformers as opposed to state of the art solutions.

U-BF has the advantage of requiring only common pilots for channel estimation and provides very competitive sum rate performance for low FB rates. Differently, ZF beamforming with MESC requires dedicated pilots for channel estimation, but when used with HFB can provide close to optimum performance even with small FB bits, especially for low mobility users.

Chapter 5

Multiuser MIMO downlink in a multi-cell cellular network

In parallel to the rapid development of understanding the fundamentals of information and communication theory of MU-MIMO, there is a need to apply these techniques to real-world systems. In contrast to isolated communication links where performance is typically limited by noise, the spectral efficiency of cellular networks is limited by co-channel interference generated by nearby cells. In this chapter we evaluate the performance of MU MIMO transmission strategies introduced in the previous chapters in the context of a multi-cell packet-based cellular network and using also SU MIMO as term of comparison. We consider two different network configurations i) TDD system with perfect CSI at transmitter and ii) FDD system where CSI at transmitter is provided through limited uplink FB from mobile terminals (MTs).

In the context of TDD systems we also investigate a novel class of techniques known as network MIMO [73, 74], which coordinates the transmissions among multiple base stations (BSs) for reducing interference. As a result of network MIMO, intercell interference can be eliminated among the coordinating bases, resulting in a significant improvement in system throughput. The tradeoff is that this technique requires user messages and channel state information to be shared among the coordinating bases, resulting in the need for enhanced backhaul capabilities. Performance evaluation of network MIMO often use an equal rate (as opposed to scheduled packet) criterion [73],[75] or use simplified cellular models in order to obtain analytical results [74]. Moreover full coordination over all bases in the network is often assumed. In this chapter we consider a more practical setting where coordination is among a limited set of BSs and multiuser proportional fair scheduling is adopted to guarantee fairness among users. With limited coordination we implicitly reduce the backhaul required for base station cooperation [76], that represents one of the main challenges in future deployments of network MIMO.

Moreover as a simple, alternative way to reduce inter-cell interference we investigate higher order sectorization, where parallel spatial channels are created physically rather than electronically through beamforming. As a final contribution, MU MIMO schemes introduced in Chapters 3 and 4 under the assumption of limited uplink FB are evaluated in a real multi-cell network, showing the potential gains of MU MIMO over SU MIMO even for FDD systems and

low FB rates.

The chapter is organized as follows. In Section 5.1 we establish the system model, and in Section 5.2 we describe the SU and MU-MIMO transmission strategies. In Section 5.3 we describe the cellular system methodology, including how MU-MIMO techniques are generalized to perform network coordination. We give numerical results in Section 5.4 and present conclusions in Section 5.5.

The material in this chapter has been in part published in [77], [66], [39] and [78].

5.1 System model

We generalize the system model in Section 2.1 considering a cellular network with N_c base stations and K users where each base and user are equipped with M and N antennas, respectively. The term “base station” or “base" will be used generically to refer to a sector, a cell, or a cluster of cells, depending on the context. Users are dropped uniformly in the network, and each is assigned to the base with maximum average SNR based on pathloss and shadowing as described in Section 5.3. We let $\mathcal{S}_b$ denote the set of users assigned to base b, with $b = 0, \ldots, N_c - 1$. We are interested in determining the throughput of base 0 in the presence of interference from the other $N_c - 1$ bases. For the kth user assigned to base $b = 0$, the received signal is:

$$\mathbf{y}_k = \mathbf{H}_{k,0}\mathbf{x}_0 + \sum_{b=1}^{N_c-1} \mathbf{H}_{k,b}\mathbf{x}_b + \mathbf{n}_k \tag{5.1}$$

where $\mathbf{H}_{k,b}$ is the $N \times M$ complex channel matrix between base b and user k, $\mathbf{x}_b$ is the M-dimensional transmitted signal from base b, and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is an additive complex white Gaussian noise vector with identity covariance matrix. Bases with indices $1, \ldots, N_c - 1$ correspond to the other bases in the network that cause interference to this user. We assume a block fading model for the channel so that it is static over one time slot. Moreover we assume an average sum power constraint (SPC) P for the M transmit antennas in each base, i.e.

$$\text{tr}(\text{E}[\mathbf{x}_b\mathbf{x}_b^H]) \leq P. \tag{5.2}$$

The transmitted signal $\mathbf{x}_b$ is a summation of the signals for users in $\mathcal{S}_b$ and, assuming linear precoding, the signal transmitted by base b is given by

$$\mathbf{x}_b = \sum_{j \in \mathcal{S}_b} \mathbf{G}_j \mathbf{s}_j \tag{5.3}$$

where d_j, $\mathbf{G}_j \in \mathbb{C}^{M \times d_j}$ and $\mathbf{s}_j = [s_j(1), \ldots, s_j(d_j)]^T$ are the number of transmitted streams, the precoding matrix and the information symbol vector for user j, respectively. We note that this model can also accommodate coordination among spatially separated bases by considering the desired base with index 0 to be a “super-base” that includes all bases within a coordinated cluster.

5.2 Transmission strategies

In order to reflect the operation of a possible next-generation packet-based cellular network, we assume that the average number of users per sector is much larger than the number of transmit antennas per sector. Due to the limited degrees of freedom, not all users can be served during each transmission slot. Therefore, we employ a scheduler that decides which users are served during each time slot, and at what rate. During the nth time slot, the scheduler generates a QoS weight $\alpha_k(n)$ for user k according to the multiuser proportional fair scheduling (MPFS) algorithm [23]. Let $\mathcal{S}(n)$ be the set of users scheduled at slot n (we have dropped the dependence on the cluster index b for convenience) and $R_k(n, \mathcal{S}(n))$ the rate scheduled to user k at slot n. For MPFS, the average throughput of MT k up to slot n is denoted as $T_k(n)$ and is updated as follows:

$$T_k(n+1) = \left(1 - \frac{1}{\tau}\right) T_k(n) + \frac{1}{\tau} R_k(n, \mathcal{S}(n)) \,, \tag{5.4}$$

where τ is a parameter related to the time over which fairness should be achieved. In [23] it has been shown that proportional fairness, maximizing $\sum_k \log_2 T_k(n)$, is achieved by scheduling users according to the following criterion:

$$\mathcal{S}(n) = \arg\max_{\mathcal{U}} \sum_{k \in \mathcal{U}} \log_2 \left(1 + \frac{R_k(n, \mathcal{U}(n))}{(\tau - 1) T_k(n-1)}\right) . \tag{5.5}$$

We observe that for $\tau >> 1$ we can approximate

$$\log_2 \left(1 + \frac{R_k(n, \mathcal{U}(n))}{(\tau - 1) T_k(n-1)}\right) \approx \frac{R_k(n, \mathcal{U}(n))}{(\tau - 1) T_k(n-1)} \tag{5.6}$$

and (5.5) boils down to the following criterion

$$\mathcal{S}(n) = \arg\max_{\mathcal{U}} \sum_{k \in \mathcal{U}} \alpha_k(n) R_k(n, \mathcal{U}(n)) \,. \tag{5.7}$$

with weights $\alpha_k(n) = T_k(n-1)^{-1}$. Therefore in each time slot, the scheduler computes $\alpha_k(n)$ for each user and the resource allocation algorithm determines the users $\mathcal{S}(n)$ and user rates to maximize the weighted sum rate (see also Fig. 5.1).

We consider both single user (SU) and multiuser (MU) MIMO transmission techniques. For SU-MIMO we adopt: 1) Transmit diversity using Alamouti space-time block coded (STBC) transmission and maximal ratio combining at the recever [79], 2) closed-loop BLAST (CLB) which requires CSI at the transmitter and serves as an upper bound on SU-MIMO performance [14, 15], and 3) Open-loop SM with linear minimum mean square error (MMSE) combiner as a relatively simple way to achieve SM gain without requiring CSI at the transmitter [80]. Since each BS transmits to only one user during a given slot, the set of selected users $\mathcal{S}(n)$ is simply the user with the largest weighted rate:

$$\mathcal{S}(n) = \{\tilde{k}\} = \arg\max_{k} \alpha_k(n) r_{SU,k}(\mathbf{H}_k(n), P) \tag{5.8}$$

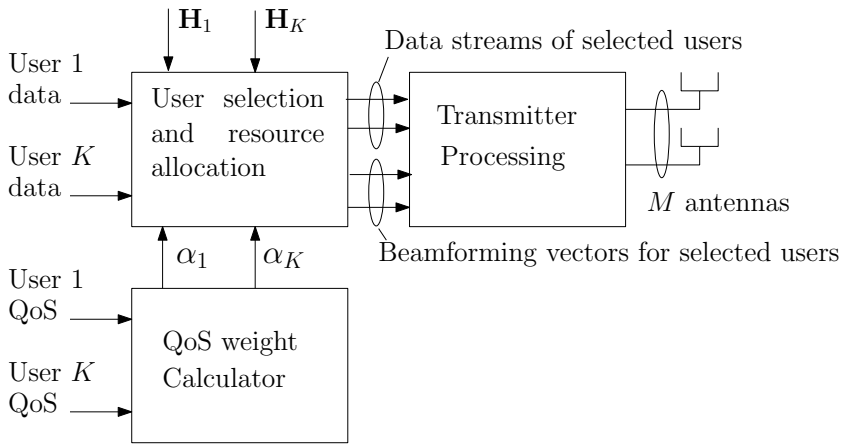

Figure 5.1: Detail of base station transmitter with scheduler which includes: QoS weight calculator, user selector, and resource allocator.

where $r_{SU,k}(\mathbf{H}_k(n), P)$ is the rate of user k which assumes the following values according to the specific transmission strategy [79, 14, 15, 80]:

$$r_{SU,k}(\mathbf{H}_k(n), P) = \begin{cases} \log_2\left(1 + \dfrac{P}{2}\displaystyle\sum_{m=1}^{2} |\underline{\mathbf{h}}_{k,m}(n)|^2\right) & \text{STBC} \\ \max\limits_{\mathbf{Q}\succeq\mathbf{0},\mathrm{tr}(\mathbf{Q})\leq\mathbf{P}} \log_2\left|\mathbf{I}_N + \mathbf{H}_k(n)\mathbf{Q}\mathbf{H}_k(n)^H\right| & \text{CLB} \\ \displaystyle\sum_{j=1}^{M} \log_2\left[1 + \underline{\mathbf{h}}_{k,j}(n)^H \mathbf{A}_{k,j}(m)^{-1}\underline{\mathbf{h}}_{k,j}(n)\right] & \text{MMSE} \end{cases} \tag{5.9}$$

where $\underline{\mathbf{h}}_{k,m}(n)$ is the mth column of matrix $\mathbf{H}_k(n)$, $\mathbf{Q}$ is the transmit covariance matrix and $\mathbf{A}_{k,j}(n) = \left(\frac{M}{P}\mathbf{I}_N + \sum_{i=1,i\neq j}^{M} \underline{\mathbf{h}}_{k,i}(n)\underline{\mathbf{h}}_{k,i}(n)^H\right)$. We underline that for STBC we consider only the case $M = 2$. The actual transmission rate during slot n is

$$R_{SU}(n) = r_{SU,\tilde{k}}(\mathbf{H}_{\tilde{k}}(n), P) \tag{5.10}$$

In case of MU MIMO we consider five different transmission techniques: three under the assumption of perfect CSI at transmitter and two adopting limited uplink feedback. In case of perfect CSIT we adopt: 1) Multiuser eigenmode transmission introduced in Chapter 2 (denoted as ZF-M) for which user selection follows (2.18) , 2) ZF-1 where only the dominant eigenmode of each user can be selected for transmission (see Chapter 2), and 3) the capacity achieving dirty paper coding (DPC) [9] for which users are selected according to (2.7). In case of limited uplink feedback we consider: 4) ZF beamforming with MESC and LBG-based codebook designed for high SNR, introduced in Chapter 4 (see (4.13) for user selection and computation of the weighted sum rate) , and 5) U-BF with MMSE receivers described in Chapter 4. For both 4) and 5) we consider the additional MMSE processing computed based on the effective set of selected users (see Section 4.2.3). For the specific computation of the optimum user set $\mathcal{S}(n)$ and the rates of selected users in case of MU MIMO strategies we remand to Chapters 2 and 4.

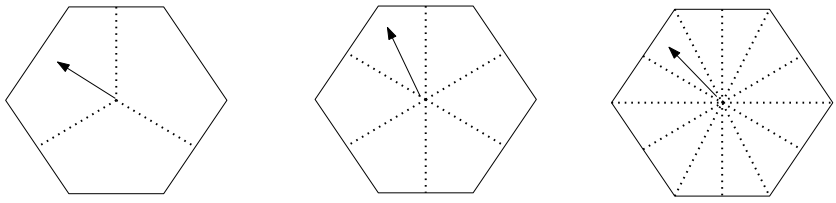

Figure 5.2: Sectorization with $S = 3, 6, 12$ sectors per cell. The arrow indicates the orientation of a representative sector antenna.

5.3 Cellular system simulation methodology

The channel coefficient between each transmit and receive antenna pair is a function of distance-based pathloss, shadow fading, and Rayleigh fading. We let the (n, m)th element ($n = 1, \ldots, N, m = 1, \ldots, M$) of the kth user's MIMO channel matrix $\mathbf{H}_{k,b}$ from base b be given by:

$$[\mathbf{H}_{k,b}]_{(n,m)} = \beta_{k,b}^{n,m} \sqrt{A(\theta_{k,b(m)}) \left[\mu_{k,b}/\mu_0\right]^{\gamma} \rho_{k,b} \Gamma} \tag{5.11}$$

where $\beta_{k,b}^{n,m}$ is independent Rayleigh fading, $\beta_{k,b}^{n,m} \sim \mathcal{CN}(0,1)$, $A(\theta_{k,b(m)})$ is the antenna element response as a function of the direction from the mth antenna of the bth base to the kth user, $\mu_{k,b}$ is the distance between the bth base and the kth user, μ_0 is a fixed reference distance, $\gamma = 3.5$ is the pathloss coefficient, and $\rho_{k,b}$ is the lognormal shadowing between the bth base and kth user with standard deviation $\sigma_\rho = 8$ dB. Since shadowing is caused by large scatterers we assume that antennas of the same cell are close enough to be characterized by the same shadowing effect. We assume universal frequency reuse, so that all bases transmit on the same frequency. The parameter Γ is the reference SNR defined as the SNR measured at the reference distance μ_0, assuming a single antenna at the base station transmitting at full power and accounting only for the distance-based pathloss. If we let μ_0 be the distance from the base station to the cell boundary, a reference SNR $\Gamma = 20$dB captures the various power and noise parameters associated with a typical cellular network operating in the interference-limited regime [75].

The antennas of each cell site are grouped and oriented so there are $S = 3, 6$ or 12 sectors per cell, with the orientations shown in Figure 5.2. The antennas are spaced sufficiently far apart so they are spatially uncorrelated. We model the antenna element response as an inverted parabola that is parameterized by the 3 dB beamwidth θ_{3dB} and the sidelobe power A_s measured in dB: $\left(A(\theta_{k,b(m)})\right)_{\text{dB}} = -\min\{12(\theta_{k,b(m)}/\Theta_{3dB})^2, A_s\}$ where $\theta \in [-\pi, \pi]$ is the direction of user k with respect to the broadside direction (given by the arrow in Fig. 5.2) of the mth antenna of base b. In the case of coordination, the broadside direction could be different for the coordinated antennas as we discuss later. As the sectorization order increases, the beamwidth decreases, and the physical width of each sector's antenna changes inversely proportionally to the beamwidth [81]. For $S = 3, 6, 12$, the corresponding parameters are $\Theta_{3dB} = (70/180)\pi$, $(35/180)\pi$, $(17.5/180)\pi$ and $A_s = 20, 23, 26$ dB, respectively. The antenna parameters as a function of S are summarized in Tab 5.1.

Number of sectors per cell (S)	Θ_{3dB}	A_s
3 sectors	$(70/180)\pi$	20 dB
6 sectors	$(35/180)\pi$	23 dB
12 sectors	$(17.5/180)\pi$	26 dB

Table 5.1: Antenna pattern parameters.

To provide fairness in the network we adopt the MPFS algorithm with fairness factor $\tau = 10$ time slots. A total of 60 users are uniformly dropped in each cell site, and users are assigned to the base (or more generally, the cell cluster) with the highest average SNR accounting for distance-based pathloss and shadowing. For each drop of users, the channel is modelled as $\mathbf{H}_{k,b} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and we either assume a TDD or FDD mode. In TDD mode we consider stationary users so that channel state information (CSI) at the transmitter is ideal. Perfect CSI is also assumed at the receiver. In FDD mode we assume either static or moving MTs and assume limited CSI from the receivers.

We consider four sets of results, each for a different network configuration. The first three sets consider a TDD system, while the last one assumes an FDD system. Moreover the first set of results compares SU-MIMO (diversity and SM) and MU-MIMO for three-sector cells. Here we assume a $N_c = 19$-cell network with the architecture A in Fig. 5.3 and $M = 2$ antennas per sector. This represents a configuration commonly found in current 3G system deployments. The second set of results considers the effects of high-order sectorization for SU-MIMO and MU-MIMO. Again we assume $N_c = 19$ (architecture A of Fig. 5.3) but the total number of antennas per cell is 12. There are $S = 3, 6, 12$ sectors per cell. This configuration is associated with near future cellular deployments. For the third set of results, we study the impact of base station coordination still adopting 12 antennas per cell. Because of the complexity of sharing data and CSI over the backhaul network, we consider coordination over limited clusters. We let C denote the number of cells in the coordination cluster and consider $C = 1, 3$, and 7. As a baseline, we consider the case of no coordination ($C = 1/S, N_c = 19$) with network architecture A. With this notation the number of coordinated antennas per cluster is $M \times L \times C$. We assume that the antenna elements are sectorized according to the parameters for $S = 3$ and the corresponding sector orientation in Figure 5.2. For $C = 1$, the 12 co-located antennas for each cell site are coordinated. The number of bases is $N_c = 19$, and the cell topology is given by architecture A of Figure 5.3. For $C = 3$ and $C = 7$, the cell topologies are given by architectures B and C, respectively, in Figure 5.3. There are a total of 36 and 84 antennas in each coordination cluster, respectively, and the number of cells per base (recall “base" refers to a cluster of coordinated cells) is $N_c = 7$ for both cases. [1]

For the final set of simulation results we compare MU MIMO strategies with limited FB in three-sectors cells and with no cell coordination, i.e. $N_c = 19$, $S = 3$ and $C = 1/3$. The total number of antennas per cell is 12, i.e. $M = 4$ antennas per sector, and CLB is included as term

[1] We note that in case of coordination between spatially separated antennas ($C = 3, 7$) it would be necessary to consider an average per-base power constraint instead of the SPC introduced in Section 2.1. Off-line analysis of the power allocation per base indicates that under SPC, the distribution of power is nearly the same for all bases. This observation indicates the marginal performance difference under a per-base constraint would be minimal.

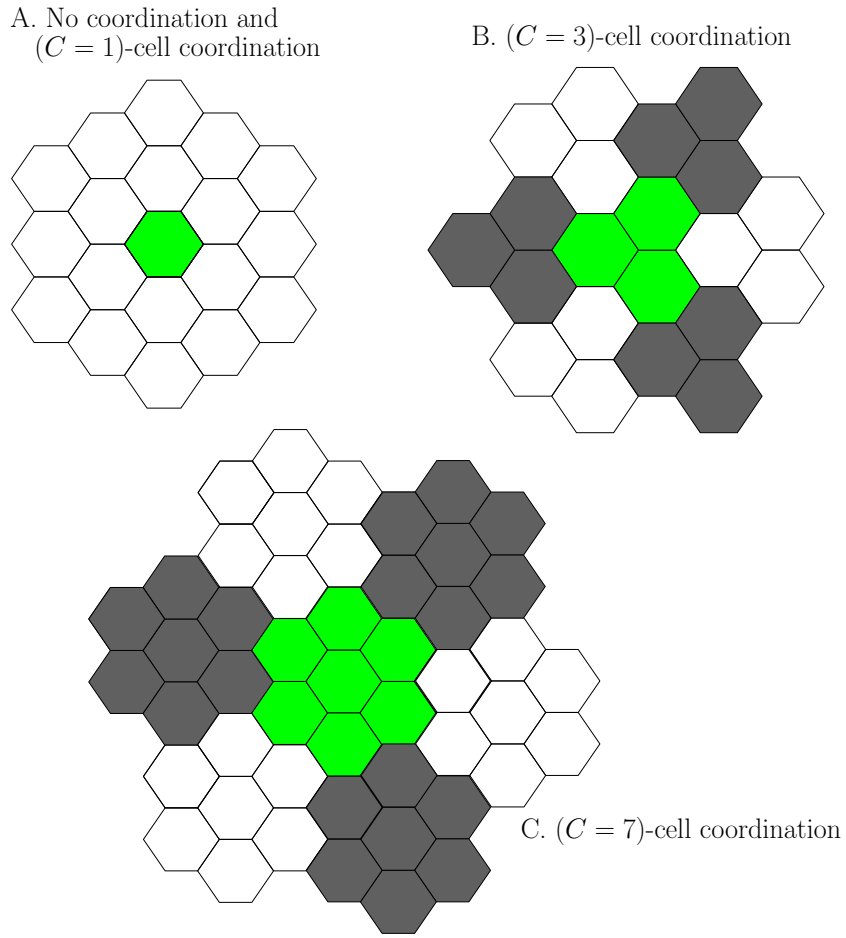

Figure 5.3: Network architectures for varying levels of base station coordination.

of comparison.

Colored intercell interference (or more generally, inter-cluster interference) is accounted for using a two-phase methodology. In the first phase, the resource allocation and transmit covariance calculations are performed assuming the intercell interference is spatially white and estimating the achievable SINR assuming all bases transmit at full power and accounting for path loss and shadowing. In the second phase, the actual achievable rates are computed assuming that the transmit covariances are colored according to sample covariances generated from the first phase. The assumption of spatially white noise in the first phase is the worst-case noise and results in a somewhat pessimistic rate. This methodology circumvents the problem of resource allocation when the statistics of the colored spatial noise are not known. In order to achieve the rates predicted in the presence of colored noise, we assume that fast incremental redundancy or some other higher level medium access protocol is employed to progressively adapt the rates.

Cell wraparound (or more generally, cluster wraparound) is used to prevent network edge effects by ensuring each cell (cluster) is surrounded by a sufficient number of interfering cells (clusters). For the case of no coordination and $C = 1$-cell coordination, wraparound is used so each cell is surrounded by two rings of cells. Each cell is at the center of its own network,

as shown in architecture A of Figure 5.3. Similarly, for the case of $C = 3$ and $C = 7$-cell coordination, cluster wraparound is used so that each cluster is surrounded by one ring of clustered cells. Even though the network topology changes with C, the comparisons are valid because at least 2 rings of interfering cells are always considered; considering more cells as source of interference would have a negligible effect on the user SINR statistics.

5.4 Numerical results

We present four sets of simulation results showing either i) the cumulative distribution function (CDF) of the cell throughput, ii) the mean cell throughput or iii) the CDF of the mean user rate.

5.4.1 Diversity, SU-MIMO, and MU-MIMO with no base coordination

We set $N_c = 19$, $S = 3$ and $C = 1/3$. Each sector has a $M, N = (2, 2)$ deployment. As shown in Fig. 5.4, because of the interference-limited nature of the network, the gains of SU-MIMO MMSE over less complex STBC with MRC are marginal (similar results have been shown in [82].) In the case of isolated cells or lower frequency reuse, intercell interference would decrease, and the throughput per sector would increase. However the overall throughput *per cell* would be reduced because of the reuse inefficiency [83]. If CSI is available at the transmitter, SU-MIMO with CLB provides an additional gain of 30% in median throughput compared to the MMSE receiver.

If CSI is available at the transmitter, then MU-MIMO ZF should be used since it provides a significant performance gain with comparable complexity with respect to SU-MIMO CLB. Under ZF-1, this gain is due to multiuser diversity and transmitting a single stream to two users simultaneously. Under ZF-M, there is the additional option of transmitting both streams to a single user, like SU-MIMO CLB. However, ZF-M has a negligible performance gain with respect to ZF-1 and the resource allocation for ZF-1 is also more robust in the presence of colored intercell interference. Using ZF-1 is further justified because it is less complex to implement compared than ZF-M, requiring less control signalling and feedback overhead. The rightmost curve in Fig. 5.4 indicates the potential gains under sophisticated dirty paper coding due to non-linear processing and multiuser interference pre-cancellation inside each sector.

5.4.2 Impact of sectorization

In the second set of results, we study the effects of sectorization on throughput for a fixed number of antennas per cell. The total number of antennas per cell is 12, so that as we vary the number of sectors per cell as $S = 3, 6, 12$, the number of antennas per sector is $M = 4, 2, 1$, respectively. Results are given for single antenna ($N = 1$) users and multiple antenna ($N = 2$) users in Figs. 5.5 and 5.6, respectively. To make a fair comparison, we calculate the throughput per cell.

We first consider CLB performance for $N = 1$ in Figure 5.5. With only a single receive antenna, no spatial multiplexing is possible. In going from $S = 3$ to 6, the diversity and combining order drops from $M = 4$ to 2. However, this drop in the diversity order *per sector*

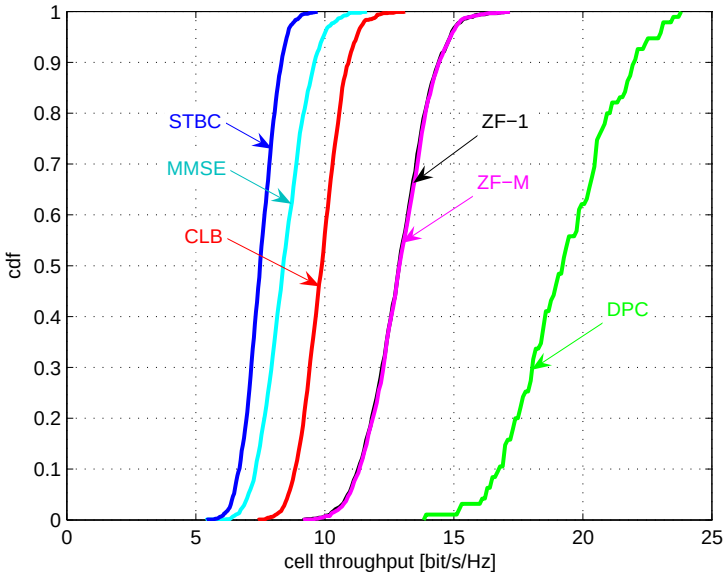

Figure 5.4: CDF of throughput per cell (bit/s/Hz), $S = 3$ sectors per cell, $M = 2$ antennas per sector, $N = 2$ antennas per user, no base coordination. SU-MIMO techniques (STBC, MMSE, CLB) are compared with MU-MIMO techniques (ZF-1, ZF-M, DPC).

is offset by the doubling in the number of sectors per cell. Overall, the median cell throughput increases by about 35%. A similar gain is observed for $N = 2$ in Figure 5.6 where multiple receive antennas allow for spatial multiplexing per user. Fixing the number of antennas *per sector* to $M = 2$ and referring to the CLB curve in Figure 5.4, the median throughput nearly doubles in going from $S = 3$ to 6 sectors for $N = 2$. These results highlight the potential gains of higher order sectorization.

Comparing CLB and ZF-1, CLB transmits to a single user using N streams whereas ZF-1 transmits a single stream to as many as M users. For the case of $S = 6, M = 2, N = 2$, even though CLB and ZF-1 have the same multiplexing order, the ZF-1 performance is superior because of the multiuser diversity advantage. For the other cases with $S = 3$ or 6, ZF-1 has a clear multiplexing advantage. For MU MIMO when $N = 2$ (see Fig. 5.6) we have the option of allocating multiple streams to a single user using ZF-M. We observe that the performance gain over the more restrictive ZF-1 is minimal, for the same reasons given in Section 5.4.1.

For both CLB and ZF, performance improves in going from 3 to 6 sectors. For CLB, the improvement is the result of higher order multiplexing. However for ZF, the maximum number of spatial channels per cell is fixed to 12, indicating that the spatial channels formed by sectorization are more effective than those formed by ZF beamforming. For both CLB and ZF, the throughput is further improved in going from $S = 6$ to 12 sectors. In this case, since there is only $M = 1$ antenna per sector, no spatial multiplexing can be achieved, and the CLB and ZF techniques are equivalent. The superior performance of $S = 12$ comes at the expense of larger antenna elements, as mentioned in Section 5.3.

Regarding DPC, in the case of single antenna users, the opposite trend regarding sectorization is observed. In other words, the throughput CDF slightly shifts to the left as the number

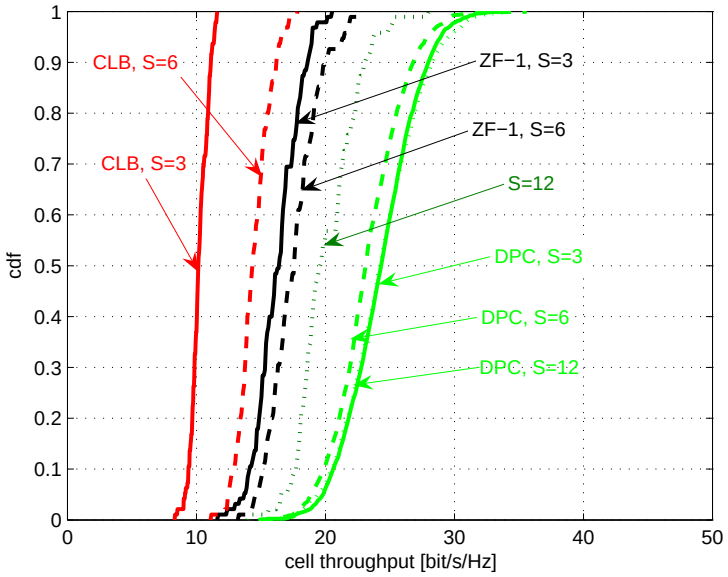

Figure 5.5: CDF of throughput per cell (bit/s/Hz) for fixed number of antennas per cell. $S = 3, 6, 12$ *sectors per cell, 12 antennas per cell,* $N = 1$ *antenna per user, no base coordination.*

of sectors goes from $S = 3$ to 6. The reason is that the spatial channels formed with DPC are more effective than those formed by sectorization. On the other hand, for $N = 2$ under DPC, the throughput performance increases as S increases. The reason is that the transmit covariances during the first phase of the simulation methodology are created assuming spatially white interference while throughput performance is measured in the presence of colored interference. Therefore with higher order sectorization, inter-cell interference appears more spatially white and there is a lower performance loss when the throughput is actually computed.

5.4.3 Impact of base coordination

In this section, we set $S = 3$ and consider the impact of coherent network MIMO as a function of coordination cluster size, designated by the number of coordinating cells C. We consider coordinated transmission using ZF-1 and DPC and include CLB with no coordination as a reference term. For a fair comparison between different coordination orders the results are given in terms of throughput per cell.

In going from no coordination up to $C = 7$-cell coordination for ZF-1, the median throughput increases by about 60% for single antenna users (see Fig. 5.7) and 50% for multiple antenna users (see Fig. 5.8). The largest gains are achieved by moving from $C = 1/3$ to $C = 1$ when the intersector interference within a cell is mitigated. Diminishing returns occur as the coordination cluster size increases, indicating that interference mitigation is not effective once the interference power is equal or below that of the receiver noise. Therefore network coordination gains are higher for higher transmit powers (in other words, higher cell edge SNR). This observation was made for uplink network coordination in [75]. Note that the ZF-1 performance with $C = 1$-cell coordination is about the same as $S = 12$ sectors. This option presents a

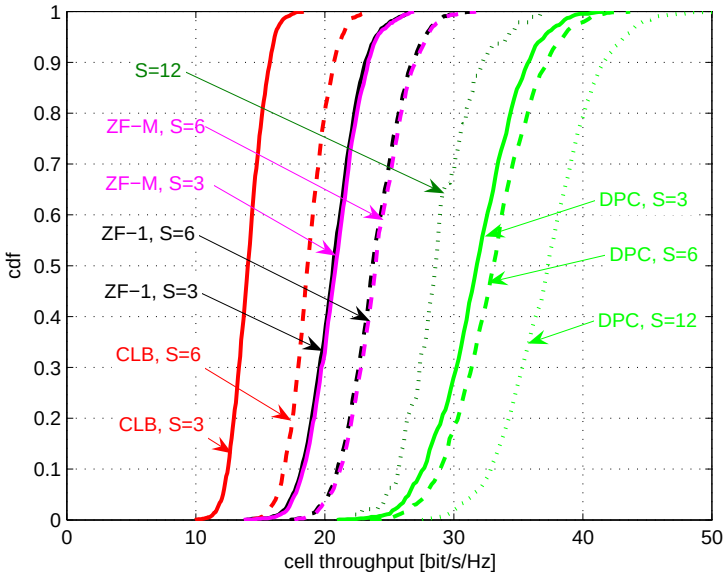

Figure 5.6: CDF of throughput per cell (bit/s/Hz) for fixed number of antennas per cell. $S = 3, 6, 12$ *sectors per cell, 12 antennas per cell,* $N = 2$ *antennas per user, no base coordination.*

favorable performance-complexity tradeoff since it can be implemented with a much smaller antenna array and minimally complex coordination among co-located antennas. The gains of coordination for DPC are much higher where, with $C = 3$-cell coordination, the throughput is almost double the case of no coordination. If we consider CLB with $S = 3$ sectors and $M = 4$ antennas per sector as a baseline, then ZF-1 with $C = 7$-cell coordination gives an approximate 2.5-fold improvement in median cell throughput for both $N = 1$ and 2.

5.4.4 Limited feedback transmission strategies

For the last set of numerical results the network is configured as in architecture A of Fig. 5.3, i.e. $N_c = 19$, $C = 1/S$. There are $S = 3$ sectors per cell with $M = 4$ antennas each and MTs are equipped with either $N = 1$, 2 or 4 antennas.

Fig. 5.9 considers a static spatially uncorrelated channel and compares in terms of mean cell throughput as a function of the number of feedback bits B: i) ZF-BF with MESC and LBG-based codebook, ii) U-BF with codebook described in Section 4.5.1 (U-BF, cod. 1) iii) U-BF with codebook described in Section 4.5.1 (U-BF, cod. 2) and iv) PU2RC [22]. For all schemes we consider a BFB strategy. Interesting all transmission strategies benefit from the additional degrees of freedom provided by multiple receive antennas thanks to MMSE receiver that attenuates multiuser interference. While cell throughput for ZF-BF increases with B the performance of U-BF and PU2RC degrades with high feedback rate. This recalls similar results obtained in Chapter 4 for users with equal average SNR.

Both U-BF and PU2RC seem to outperform ZF-BF in terms of mean cell-throughput for small B but this loss is traded for an higher fairness of the system. Indeed for $B = 4$, Fig. 5.10 shows that the CDF of the individual user rate has a steeper behaviour for ZF-BF assuring more

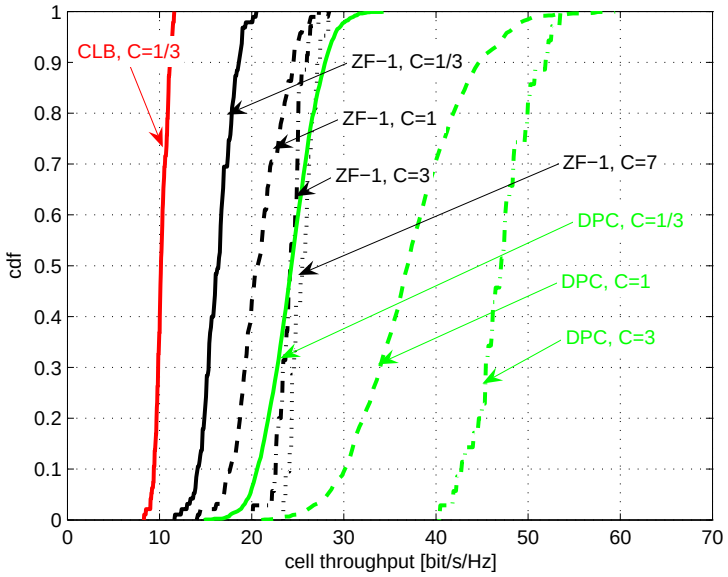

Figure 5.7: CDF of throughput per cell (bit/s/Hz), 12 antennas per cell, $C = 1/3, 1, 3, 7$ cell cluster coordination, $N = 1$ antenna per user, $S = 3$.

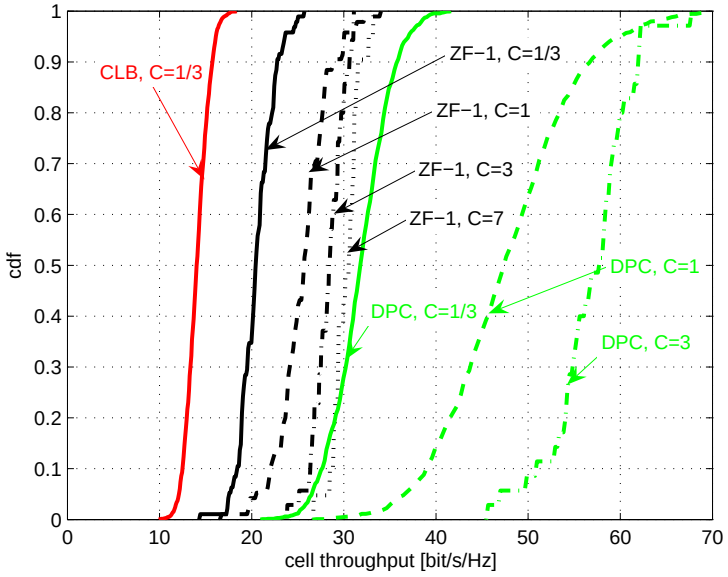

Figure 5.8: CDF of throughput per cell (bit/s/Hz), 12 antennas per cell, $C = 1/3, 1, 3, 7$ cell cluster coordination, $N = 2$ antennas per user, $S = 3$.

fairness among users. Interesting for $B = 4$, U-BF with cod. 2 achieves both an higher cell throughput and an higher user fairness with respect to other unitary beamforming strategies.

Finally Fig. 5.11 compares i) ZF-BF with BFB, ii) U-BF, cod. 1, with BFB, iii) ZF-BF with HFB, iv) ZF-BF with perfect CSIT (ZF-BF-PCSIT) and v) CLB, in terms of cell throughput for different speeds of the mobile terminals. Spatial and time correlation of the channel is modelled

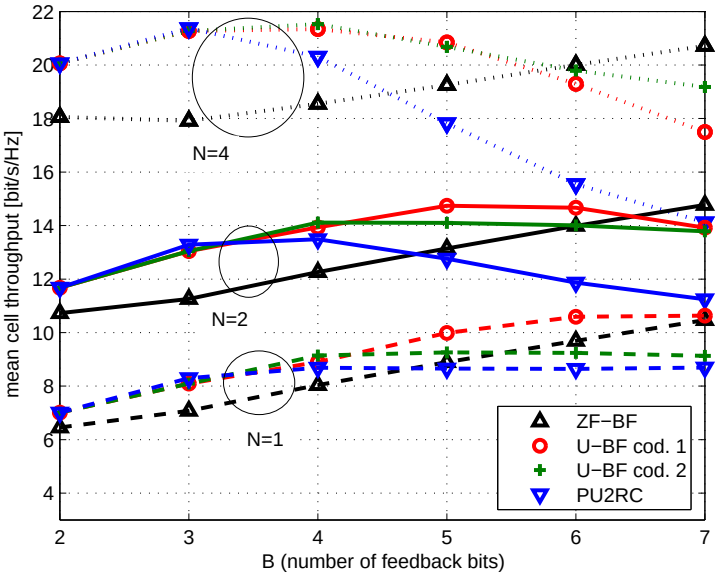

Figure 5.9: Mean cell throughput of U-BF, cod. 1, U-BF, cod. 2, ZF-BF and PU2RC as a function of the FB rate B. All schemes adopt the BFB strategy. $M = 4$, $K = 20$ users per sector on average. Spatially uncorrelated Rayleigh fading channel, $v = 0$ km/h.

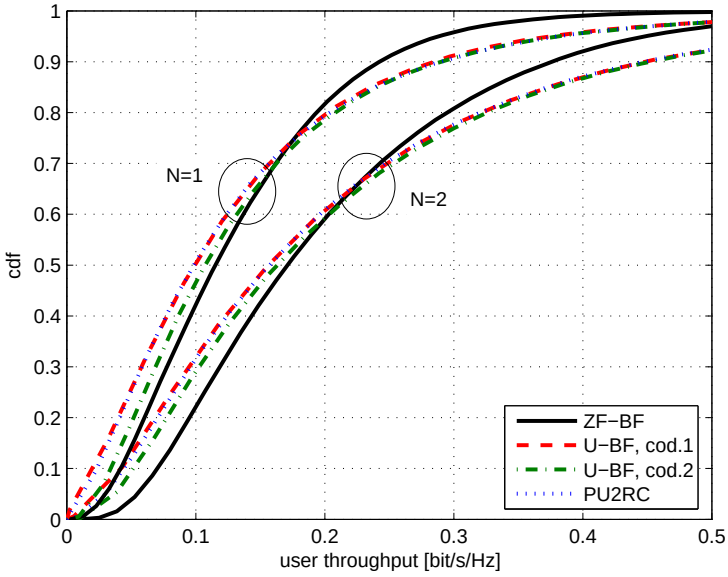

Figure 5.10: CDF of user rate for U-BF, cod. 1, U-BF, cod. 2, ZF-BF and PU2RC. All schemes adopt the BFB strategy. $M = 4$, $K = 20$ users per sector on average, $N = 1, 2$, $B = 4$, spatially uncorrelated Rayleigh fading channel, $v = 0$ km/h.

as in [60] assuming transmit antennas spaced $\Delta_{Tx} = 10\ \lambda$, with λ the transmission wavelength, $f_c = c/\lambda = 2$ GHz the carrier frequency and $c = 3 \times 10^8$ m/s the speed of light. The time slot is $T = 0.5$ ms and mobile terminals speed is either $v = 3$, or 50 km/h. All limited feedback

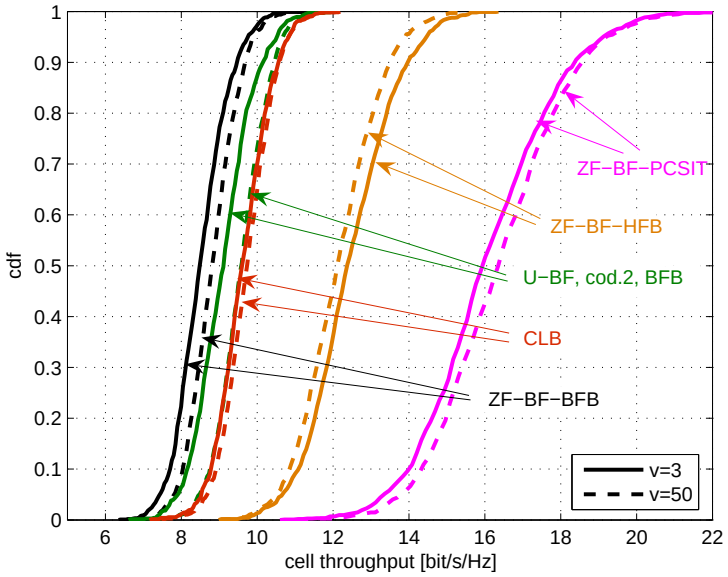

Figure 5.11: CDF of the cell throughput for U-BF, cod. 2 with BFB, ZF-BF-BFB, ZF-BF-HFB and CLB. $M = 4$, $K = 20$ users per sector on average, $N = 1$, $B = 4$, spatial channel model with $\Delta_{Tx} = 10\lambda$.

schemes use $B = 4$ bits and $B_{max} = 12$ is the maximum number of levels in the LBG-tree codebook. While ZF-BF-BFB and U-BF, cod. 1, with BFB have comparable performance, ZF-BF-HFB is able to exploit the time correlation of the channel and thanks to hierarchial indexing achieves approximately a 50% improvement in median cell throughput over the other schemes. Moreover ZF-BF-HFB significantly outperforms CLB with PCSIT, emphasizing the potential gains of MU MIMO even with few FB bits. Numerical simulations verified that all unitary BF schemes (U-BF, cod. 1, U-BF, cod. 2 and PU2RC) have practically the same performance for $B = 4$. Interestingly, increasing mobile terminals speed, the proportional fair algorithm benefits from time diversity providing higher cell throughput for ZF-BF-BFB, U-BF cod. 1 with BFB, CLB and ZF-BF-PCSIT. Differently HFB is less efficient with increasing v because quantization accuracy degrades. Numerical simulations verified that similar arguments hold for $N > 1$ as well.

5.5 Conclusions

This chapter evaluates the downlink throughput of SU and MU MIMO techniques under a unified simulation environment that models a multi-cell system serving a dense population of stationary users.

In case of perfect CSI at transmitter, a reasonable assumption in TDD systems, for a given sectorization order, MU-MIMO outperforms SU-MIMO because of greater spatial multiplexing. For a fixed number of antennas per cell, throughput increases as the number of sectors per cells increases, resulting in larger antennas. Coordinating transmissions among antennas at one

or more cells improves throughput by mitigating interference but requires additional backhaul resources and higher computational complexity.

Compared to a SU-MIMO baseline with $S = 3$ sectors per cell and $M = 4$ antennas per sector, a MU-MIMO with the same antenna architecture but with coordination only among 12 co-located antennas in a cell (effectively eliminating the notion of sectors) increases the median throughput by a factor of 2. Similar performance is achievable with the configuration using the maximum sectorization order but the first solution requires a much smaller antenna array. By coordinating a cluster of 7 cells (a total of 84 antennas), the throughput increases by a factor of 2.5.

In case of no coordination and limited uplink FB from users, a suboptimum MU-MIMO strategy based on ZF with MESC and HFB shows a significant performance gain with respect to SU MIMO with perfect CSIT, emphasizing the potential gains of MU MIMO even in FDD systems.

Chapter 6

Multiuser MIMO-OFDM with limited feedback

In Chapters 2-5 we considered single carrier flat fading MIMO transmissions with particular interest in uplink FB optimization for FDD MU MIMO downlink systems.

Anyway FB optimization might become even more relevant in broadband systems where the channel is dispersive. In this context, orthogonal frequency division multiplexing (OFDM) is considered a good candidate as modulation scheme for the LTE of 3G cellular systems [84], since it converts a frequency selective fading channel into a number of parallel flat fading subchannels that allow an efficient and simple equalization.

In this chapter we consider a MU MIMO-OFDM downlink system with beamforming and address the problems of channel quantization and FB optimization. We underline that, differently from most contributions in literature [85, 86, 61], the proposed solutions are explicitly designed for a multiuser environment. In order to reduce control overhead and signal processing complexity we resort to an approach vastly adopted in the standardization of 4th generation wireless communication systems, e.g. LTE [84], where the available bandwidth is divided into resource blocks (RBs), each comprising a number of adjacent subcarriers [87]. We notice that as the transmitter is equipped with multiple antennas, in general the channel relative to a RB is represented by a space-frequency matrix. Moreover the spectral width of a RB is chosen so that almost independent fading realizations are experienced in adjacent RBs, motivating an independent channel quantization per RB. The RB approach reduces the complexity of i) channel quantization at receivers, i.e. how to select a suitable codematrix for a given RB channel matrix, and ii) user selection at the transmitter.

The chapter contains two main contributions: i) the proposal and comparison of techniques for RB channel matrix (space-frequency) quantization and ii) the optimization of FB bit allocation across the RBs in a specific multiuser environment. Within the problem of RB channel matrix quantization we propose a new simplified metric for designing the quantization codebook based on the Lloyd-Max algorithm [20]. This metric is derived from the system achievable throughput and is a non trivial extension of the approach followed in Chapter 3 for flat fading MIMO channels. As main result on RB channel quantization we provide joint conditions on the channel coherence bandwidth and the FB rate per RB that allow for an approximation of the

channel matrix (space-frequency) as a simpler channel vector (space). This approach further simplifies quantization at receivers and user selection at transmitter, with the additional benefit of reducing the complexity of beamformer design at transmitter.

As second contribution we consider an optimization problem that has received less attention in literature but might lead tremendous performance improvement in a MU environment. In fact, most state-of-the-art techniques use the available FB bits to quantize the complete channel frequency response of a user [85, 88, 89], to exploit frequency correlation. This might not be the best approach in a MU environment where multiuser diversity can be exploited. To this purpose we study whether it's better to i) *concentrate* all the available FB bits in quantizing only one selected RB for each user or ii) *distribute* the available FB bits among more RBs. We show analytically that when the number of users is large, the first approach is preferable and through simulations we validate this finding even for a practical number of users in the system. This result emphasizes the importance of an accurate channel knowledge in case of limited uplink feedback and recalls a similar finding obtained in [48] for flat MIMO channels, where a *high-rate/high-quality feedback* from a *small number* of randomly selected users is proved to be preferable than a *low-rate feedback* from a *large number* of users.

The chapter is organized as follows. Section 6.1 presents the system model, Section 6.2 addresses the problem of codebook design for RB channel matrix quantization and Section 6.3 compares both analytically and by numerical simulations the RB channel matrix and the RB channel vector approximation models. Section 6.4 describes two FB rate allocation strategies that exploit frequency/multiuser diversity and numerical simulations are provided in Section 6.5. Finally, Section 6.6 concludes the paper summarizing the main results.

The material of this chapter has been in part published in [90] and [91].

6.1 System model

We generalize the system model in Section 2.1 and consider the downlink of a cellular OFDM system with N_C subcarriers. The transmitter has M transmit antennas and K users have one antenna each. The available bandwidth is divided into N_R RBs [84] each comprising L adjacent subcarriers, and both FB signalling and user selection are performed on a RB-basis. The channel is assumed frequency selective and the $1 \times M$ channel vector of user k relative to subcarrier ℓ of RB n is denoted with $\mathbf{h}_{k,n}(\ell)$. The $L \times M$ channel matrix relative to RB n is denoted with $\mathbf{H}_{k,n} = [\mathbf{h}_{k,n}(0)^T, \ldots, \mathbf{h}_{k,n}(L-1)^T]^T$. We assume that the channel is quasi static, i.e., it can be considered invariant for the duration of one OFDM symbol and vectors $\mathbf{h}_{k,n}(\ell)$ are assumed to be uncorrelated across users. We consider an FDD system and in each OFDM symbol (slot) users feed back a partial CSI, which is used by the transmitter to schedule downlink transmissions and perform beamforming.

Let $\mathcal{S}_n = \{s_n^{(1)}, \ldots, s_n^{(|\mathcal{S}_n|)}\}$ be the set of scheduled users receiving data on RB n and $\mathbf{x}_n(\ell)$ the $1 \times M$ transmitted symbol vector on subcarrier $\ell = 0, \ldots, L-1$, which is related to the information symbols $\{d_{j,n}(\ell)\}$ for user $j \in \mathcal{S}_n$ via linear beamforming, i.e.

$$\mathbf{x}_n(\ell) = \sum_{j \in \mathcal{S}_n} \mathbf{g}_{j,n}(\ell) d_{j,n}(\ell) \,, \tag{6.1}$$

with $\{\mathbf{g}_{j,n}(\ell)\}$ $1 \times M$ beamforming vectors. The signal received by user $k \in \mathcal{S}_n$ on subcarrier ℓ can be written as

$$\begin{aligned} y_{k,n}(\ell) &= \mathbf{h}_{k,n}(\ell)\mathbf{x}_n^T(\ell) + n_{k,n}(\ell) \\ &= \left[\mathbf{h}_{k,n}(\ell)\mathbf{g}_{k,n}^T(\ell)\right] d_{k,n}(\ell) + \sum_{j\in\mathcal{U}(n), j\neq k} \left[\mathbf{h}_{k,n}(\ell)\mathbf{g}_{j,n}^T(\ell)\right] d_{j,n}(\ell) + n_{k,n}(\ell) \end{aligned} \quad (6.2)$$

where $n_{k,n}(\ell)$ is the additive complex Gaussian noise with zero mean and unit variance and the summation term in the second line of (6.2) accounts for multiuser interference on user k.

The transmit signal is subject to the average power constraint

$$\mathrm{E}\left[\sum_{n=1}^{N_R}\sum_{\ell=0}^{L-1} ||\mathbf{x}_n(\ell)||^2\right] \leq P\,, \quad (6.3)$$

where P is the available power at transmitter. We assume equal power allocation across active RBs and the corresponding subcarriers, hence the power allocated on subcarrier ℓ of RB n is $P(\ell) = P/(L\bar{N}_R)$ where $\bar{N}_R$ is the number of active RBs. Moreover, $P(\ell)$ is uniformly distributed across the users selected on subcarrier ℓ and the power allocated to the kth selected user is

$$P_k(\ell) = \frac{P(\ell)}{|\mathcal{S}_n|} = \frac{P}{L\bar{N}_R|\mathcal{S}_n|}\,. \quad (6.4)$$

From (6.3) and assuming that the channel has unitary average gain, the average system signal to noise ratio per subcarrier is defined as SNR $= P/(LN_R)$.

6.1.1 Finite rate feedback strategies

As in Chapters 3-4, we assume that user k perfectly estimates its frequency selective channel frequency response $\bar{\mathbf{H}}_k = [\mathbf{H}_{k,1}^T, \ldots, \mathbf{H}_{k,N_R}^T]^T$ and feeds back in each slot a quantized version utilized at transmitter for both user selection and beamformer design. In this paper we propose two different strategies for allocating the available FB bits among the N_R RBs, denoted as *distributed feedback* (DFB) and *best RB feedback* (BeFB).

In DFB a set of D RBs is randomly selected by the transmitter for each user, irrespective of channel conditions, with the aim of assuring that CSI relative to at least one user is available for each RB. Then user k feeds back two information for each assigned RB: 1) the channel direction information (CDI) given by a quantized version $\hat{\mathbf{H}}_{k,n} = [\hat{\mathbf{h}}_{k,n}(0)^T, \ldots, \hat{\mathbf{h}}_{k,n}(L-1)^T]^T$ of the normalized channel matrix $\tilde{\mathbf{H}}_{k,n} = [\tilde{\mathbf{h}}_{k,n}(0)^T, \ldots, \tilde{\mathbf{h}}_{k,n}(L-1)^T]^T$, with $\tilde{\mathbf{h}}_{k,n}(\ell) = \mathbf{h}_{k,n}(\ell)/||\mathbf{h}_{k,n}(\ell)||$ and 2) an analog channel quality indicator (CQI) related to the estimated achievable throughput on the RB. The total FB bits B_{CDI} for CDI quantization are uniformly distributed among the D RBs leading to $b_{CDI} = B_{CDI}/D$ FB bits per RB. This FB bit allocation includes as limit situations: i) a uniform bit distribution of B_{CDI} among all the RBs, when $D = N_R$ and ii) the use of B_{CDI} bits for quantizing only the RB channel matrix selected by the transmitter, when $D = 1$. We notice that the optimum value of D should be chosen by the transmitter as a function of both the user selection algorithm and the number of users in the system and then conveyed to the users via a feed-forward link. We refer to Section 6.5 for

	Num. of selected RBs	**CQI**		**CDI**		
		tot. bits B_{CQI}	**bits per RB** b_{CQI}	**tot. bits** B_{CDI}	**bits per RB** b_{CDI}	**bits RB indexing**
DFB	D	D	1	$B-D$	$B/D-1$	0
BeFB	1	1	1	$B-1$	$B-1-\log_2(N_R)$	$\log_2(N_R)$

Table 6.1: Allocation of the available FB bits in DFB and BeFB

more details on the optimization of D.

Similarly to DFB with $D=1$, BeFB still uses all the available FB bits for characterizing only the channel matrix of one RB, but differs from the aforementioned strategy in the choice of the RB. In this case, selection is done by the user, which feeds back information about the RB providing the highest estimated achievable throughput. Note that with BeFB the user needs to feed back the index of the selected RB along with CDI and CQI. Indeed, $\log_2(N_R)$ bits are reserved as RB index and the remaining $B_{CDI}-\log_2(N_R)$ bits are used for RB channel quantization.

With regard to CQI FB, we notice that BeFB requires only one analog value, while DFB needs D values per slot, which may become a significant FB overhead. In the following, we optimistically assume that by exploiting the time correlation of the MIMO channel, each CQI can be updated with an incremental approach using only one FB bit per slot, similarly to power control schemes in cellular systems [92][1]. As a consequence, DFB needs at least D FB bits whereas BeFB requires only one bit per slot. Adopting this model, if B bits are available, DFB keeps $B_{CQI}=D$ bits to update the CQIs and $B_{CDI}=B-D$ bits for CDI quantization. Differently, BeFB uses only one bit for CQI FB and $B_{CDI}=B-1$ bits for CDI quantization. Notice that we are implicitly favoring DFB when $D>1$, because in time variant channels one bit per slot might not be enough for an accurate CQI update. As a consequence, more FB bits should be allocated for this task, leaving fewer bits for CDI quantization. The allocation of the available FB bits according to the two strategies is summarized in Tab. 6.1.

6.1.2 CDI quantization and CQI feedback

We consider two approaches for CDI quantization: RB channel matrix quantization (RBCM-Q) and RB channel vector quantization (RBCV-Q). In RBCM-Q each RB is represented by L frequency components per transmit antenna and we make use of a codebook $\mathcal{C}_{RBCM-Q}$

[1]This assumption needs a comment for BeFB. If the channel is slowly time variant, we can expect the value of CQI FB, as the maximum value of N_R CQIs, to change slowly from slot to slot, and hence one FB bit per slot may suffice to describe its variation, even if the maximum value may occur for a different RB. Indeed, denote with $\gamma_{k,n}(t)$ (see also Section 6.1.2) the CQI for user k in RB n at time slot t. In BeFB the CQI is chosen as the maximum among N_R CQIs. If the best RB $\overline{n}$ changes in subsequent time slots, i.e $\overline{n}=i$ at time slot t and $\overline{n}=j$, $j\neq i$ in time slot $t+1$, it means that $\gamma_{k,j}(t)<\gamma_{k,i}(t)=\gamma_k(t)^{\max}$ where $\gamma_k(t)^{\max}$ denotes the maximum CQI for user k in time slot t, but $\gamma_k(t+1)^{\max}=\gamma_{k,j}(t+1)>\gamma_{k,i}(t+1)$. As a consequence it's easy to show that one of the following inequalities must hold: either i) $|\gamma_k(t+1)^{\max}-\gamma_k(t)^{\max}|\leq|\gamma_{k,j}(t+1)-\gamma_{k,j}(t)|$ or ii) $|\gamma_k(t+1)^{\max}-\gamma_k(t)^{\max}|\leq|\gamma_{k,i}(t+1)-\gamma_{k,i}(t)|$. Therefore in slowly time variant channels if we can assume, as in DFB, a 1-bit incremental quantization for the CQI of a generic RB (e.g. i or j), this is reasonable also in BeFB because in each time slot the difference $|\gamma_k(t+1)^{\max}-\gamma_k(t)^{\max}|$ is always smaller than the difference between subsequent CQIs relative to at least one RB.

of $2^{b_{CDI}}$, $L \times M$ complex matrices $\{\mathbf{C}_i\}$. Differently, with RBCV-Q we look for a single $1 \times M$ space vector that approximates the entire RB channel matrix, assuming channel vectors across the frequency domain are similar. In this case we use a codebook $\mathcal{C}_{RBCV-Q}$ of $2^{b_{CDI}}$, $1 \times M$ complex vectors $\{\mathbf{c}_i\}$. RBCM-Q allows for a better characterization of the RB channel matrix in case of highly frequency selective fading channels, but when b_{CDI} is large the memory required for codebook storage may be prohibitive. On the other hand, RBCV-Q requires less memory but might become less effective when the coherence bandwidth of the channel diminishes. A comparison between these two strategies is proposed in Section 6.3.

In this paper the system performance is evaluated in terms of achievable throughput. We recall that user k does not have *a priori* knowledge of the CDIs of other selected users, therefore it does neither know the beamforming vectors $\{\mathbf{g}_{j,n}\}$ nor the multiuser interference. Nevertheless, generalizing the approach in Chapter 3, under the approximation of assuming M orthogonal selected users and equal power allocation, the achievable throughput on the generic RB n of user k is estimated as

$$R(\mathbf{H}_{k,n}, \hat{\mathbf{H}}_{k,n}) = \sum_{\ell=0}^{L-1} \log_2 \left(1 + \gamma_{k,n}(\ell)\right) \ , \tag{6.5}$$

where

$$\gamma_{k,n}(\ell) = \frac{\rho ||\mathbf{h}_{k,n}(\ell)||^2 |\tilde{\mathbf{h}}_{k,n}(\ell)\ \hat{\mathbf{h}}_{k,n}(\ell)^H|^2}{1 + \rho ||\mathbf{h}_{k,n}(\ell)||^2 (1 - |\tilde{\mathbf{h}}_{k,n}(\ell)\ \hat{\mathbf{h}}_{k,n}(\ell)^H|^2)} \tag{6.6}$$

is the achievable signal-to-interference plus noise ratio (SINR) of user k on subcarrier ℓ of RB n and $\rho = P/(MLN_R)$ is the power allocated to each user. Interestingly, (6.6) can be shown to provide a tight approximation of the *expected* achievable SINR even in case of *nearly* orthogonal selected users as done in Chapter 3 for flat MIMO channels. We adopt (6.5) as metric for RBCM-Q, therefore as *quantization rule* we select the codematrix $\hat{\mathbf{H}}_{k,n}$ that maximizes the estimated achievable throughput,

$$\text{Quantization rule} \qquad \hat{\mathbf{H}}_{k,n} = \underset{\mathbf{C}_i \in \mathcal{C}_{RBCM-Q}}{\arg\max} \ R(\mathbf{H}_{k,n}, \mathbf{C}_i) \ , \tag{6.7}$$

where $\mathbf{C}_i = \left[\mathbf{c}_i(0)^H, \mathbf{c}_i(1)^H, \ldots, \mathbf{c}_i(L-1)^H\right]^H$ with $||\mathbf{c}_i(\ell)|| = 1$, $\ell = 0, \ldots, L-1$, is the generic codematrix. In particular RBCV-Q can be considered a special case with $\mathbf{c}_i(\ell) = \mathbf{c}_i$, for $\ell = 0, \ldots, L-1$. For both RBCM-Q and RBCV-Q the codebook is designed off-line and known to both transmitter and users a priori. The actual codebook design is discussed later in Section 6.2.

With regard to CQI, based on (6.7) and extending the approach of Chapter 3 to OFDM, the value of CQI feedback for RB n is given by

$$\gamma_{k,n} = 2^{R\left(\mathbf{H}_{k,n}, \hat{\mathbf{H}}_{k,n}\right)/L} - 1 \ . \tag{6.8}$$

We notice that by performing channel coding across the L subcarriers of RB n, the achievable rate on the RB is $R_{k,n} = L \log_2(1 + \gamma_{k,n})$. Therefore (6.8) can be seen as the equivalent SINR on RB n that assures rate $R_{k,n}$.

6.1.3 User selection

The transmitter performs user selection and linear beamforming based on the CDIs (6.7) and CQIs (6.8) received from the users and supports up to M users on each RB. As user selection algorithm we propose a simple generalization of the semi-orthogonal user selection (SUS) scheme [42] described in Chapter 3 that operates on a RB basis. Define the initial user set $\mathcal{A}_n^{(0)}$ containing all the indexes of users that fed back CDI and CQI for RB n. The first selected user is

$$u_n^{(1)} = \arg\max_{k \in \mathcal{A}_n^{(0)}} \gamma_{k,n} \, . \tag{6.9}$$

After selecting i users, the $(i+1)$th user is chosen within the set

$$\mathcal{A}_n^{(i)} = \left\{ k \in \mathcal{A}_n^{(i-1)} : \left(\frac{1}{L} \sum_{\ell=0}^{L-1} |\hat{\mathbf{h}}_{k,n}(\ell) \, \hat{\mathbf{h}}_{u_n^{(j)},n}^H(\ell)|^2 \right)^{1/2} \leq \epsilon, \quad j = 1, \ldots, i \right\} \tag{6.10}$$

as

$$u_n^{(i+1)} = \arg\max_{k \in \mathcal{A}_n^{(i)}} \gamma_{k,n} \, , \tag{6.11}$$

where ϵ is a design parameter setting the maximum correlation allowed between the quantized channel matrices of the selected users[2].

We note that in DFB with $D = N_R$, $\mathcal{A}_n^{(0)} = \{1, \ldots, K\}$, while for both BeFB and DFB with $D < N_R$, $\mathcal{A}_n^{(0)} \subseteq \{1, \ldots, K\}$. Moreover the SUS algorithm described in Chapter 3 becomes a special case of the proposed algorithm when used with RBCV-Q.

6.1.4 Zero-forcing beamforming (ZF-BF)

Let $\hat{\mathbf{H}}_n(\ell) = [\hat{\mathbf{h}}_{s_n^{(1)}}(\ell)^T, \ldots, \hat{\mathbf{h}}_{s_n^{(|\mathcal{S}_n|)}}(\ell)^T]^T$ be the concatenated unit-norm quantized channel vectors of the selected users $\mathcal{S}_n$ on subcarrier ℓ of RB n. By denoting with $\mathbf{F}_n(\ell) = [\mathbf{f}_{k,n}^T(0), \ldots, \mathbf{f}_{k,n}^T(|\mathcal{S}_n|)]$ the right pseudo-inverse of the quantized channel, the ZF beamforming matrix is given by

$$\begin{aligned} \mathbf{G}_n(\ell) &= [\mathbf{g}_{k,n}^T(0), \ldots, \mathbf{g}_{k,n}^T(|\mathcal{S}_n|)] = \mathbf{F}_n(\ell) \text{diag}(\mathbf{P}_n(\ell))^{1/2} \\ &= \hat{\mathbf{H}}_n(\ell)^H \left(\hat{\mathbf{H}}_n(\ell) \hat{\mathbf{H}}_n(\ell)^H \right)^{-1} \text{diag}(\mathbf{P}_n(\ell))^{1/2} \, , \end{aligned} \tag{6.12}$$

where $\mathbf{P}_n(\ell)$ is the vector of power normalization coefficients imposing the power constraint (6.4), with entries $P_{k,n}(\ell) = [\mathbf{P}_n(\ell)]_k = \frac{P}{L\bar{N}_R |\mathcal{S}_n| \|\mathbf{f}_k(\ell)\|^2}$.

We note that when RBCM-Q is used we need to design one beamformer per subcarrier. Differently, in RBCV-Q the same beamformer is used for all the subcarriers of the same RB, reducing the design complexity by a factor L.

[2] We underline that the optimization of ϵ depends on both K and the quantization codebook. For example, if we define $\alpha = \min_{i,j=1,\ldots,2^{b_{CDI}}, i \neq j} \left(\frac{1}{L} \sum_{\ell=0}^{L-1} |\mathbf{c}_i(\ell) \mathbf{c}_j(\ell)^H|^2 \right)^{1/2}$ as the minimum correlation between codematrices in $\mathcal{C}_{RBCM-Q}$, we have to set $\epsilon > \alpha$, otherwise only one user is selected by the transmitter.

6.2 Codebook design for RBCM-Q and RBCV-Q

In this section we address the design of the quantization codebook for RBCV-Q and RBCM-Q, which in both cases is done off-line for a given channel statistic and is saved in the memory of both transmitter and receivers. For ease of notation we drop both indexes n and k of the RB and user.

The RBCV model does not characterize the channel variability across the L subcarriers of a RB and the codebook can be simply designed using algorithms proposed for *flat* MIMO channels, either uncorrelated [46, 42] or correlated [47, 93, 51]. In particular we adopt the codebook generation technique used in Chapter 3 based on the LBG algorithm.

Differently, quantization in the RBCM model is performed jointly for the L subcarriers forming a RB, whose correlation depends on the power delay profile of the MIMO channel. As a consequence, the codebook for RBCM-Q should exploit correlation both in frequency and across transmit antennas (spatial). To this aim we consider the LBG algorithm generalizing the approach used in Chapter 3 for flat MIMO channels.

For RBCM, the performance metric (6.5) is fairly complex to be used in (3.31). In fact, codebook optimization could be done only by numerical methods, with a very high computation complexity and no guarantee of convergence. It is shown in Appendix A.3 that a suboptimum performance metric, derived by applications of Jensen's inequality to (6.5) in both low and high-SNR regimes, is given by

$$\mu(\mathbf{H}, \mathbf{C}_i) = \sum_{\ell=0}^{L-1} |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2 \ . \tag{6.13}$$

Now, codebook design according to (3.31) with the performance metric (6.13) leads to optimizing separately each row of $\mathbf{C}_i$. Indeed given a partition region $\mathcal{R}_i$ of the training sequence, the corresponding optimum codematrix $\mathbf{C}_i$ is given by

$$\mathbf{C}_i = \underset{\substack{\mathbf{C}\in\mathbb{C}^{L\times M} \\ ||\mathbf{c}(\ell)||^2=1,\ \ell=0,\ldots,L-1}}{\arg\max} \sum_{\mathbf{H}\in\mathcal{R}_i} \mu(\mathbf{H}, \mathbf{C}) \ , \tag{6.14}$$

where from (6.13)

$$\begin{aligned} \sum_{\mathbf{H}\in\mathcal{R}_i} \mu(\mathbf{H}, \mathbf{C}) &= \sum_{\ell=0}^{L-1} \sum_{\mathbf{H}\in\mathcal{R}_i} |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}(\ell)^H|^2 \\ &= \sum_{\ell=0}^{L-1} \mathbf{c}(\ell) \left(\sum_{\mathbf{H}\in\mathcal{R}_i} \tilde{\mathbf{h}}(\ell)^H \tilde{\mathbf{h}}(\ell) \right) \mathbf{c}(\ell)^H \ . \end{aligned} \tag{6.15}$$

The maximization of (6.15) can be achieved independently for each row. In particular the ℓth row of the optimum codematrix $\mathbf{C}_i$ is the dominant eigenvector of the matrix $\sum_{\mathbf{H}\in\mathcal{R}_i} \tilde{\mathbf{h}}(\ell)^H \tilde{\mathbf{h}}(\ell)$, normalized to unit norm.

6.3 Comparison between RBCM-Q and RBCV-Q

In this section we compare analytically RBCV-Q and RBCM-Q, using as metric an equivalent time domain (TD) representation of (6.13). Analytic results are then validated using numerical simulations to compare the two quantization schemes in terms of achievable throughput (6.5), using (6.7) as quantization rule.

In details, the discrete inverse Fourier transform of RBCM $\mathbf{H}$ provides

$$\mathbf{p}(i) = \sum_{\ell=0}^{L-1} \mathbf{h}(\ell) e^{+j2\pi\ell\frac{i}{L}} \,, \quad i = 0, \ldots, L-1 \,, \tag{6.16}$$

where $\{\mathbf{p}(i)\}$ are space vectors in the TD. We underline that for the RBCV model vectors $\{\mathbf{p}(i)\}$ are of the type

$$\begin{cases} \mathbf{p}(i) = 0 \quad i = 0, \ldots, j-1, j+1, \ldots, L-1 \\ \mathbf{p}(j) \neq 0 \end{cases} \tag{6.17}$$

for a suitable value of $j \in \{0, \ldots, L-1\}$.

Assuming that vectors $\{\mathbf{p}(i)\}$ are statistically independent, optimum quantization is obtained by quantizing each vector separately. In particular, let $\hat{\mathbf{p}}(i)$ be the quantized version of $\tilde{\mathbf{p}}(i) = \mathbf{p}(i)/||\mathbf{p}(i)||$. With this approach in the RBCM model the available FB bits b_{CDI} are distributed among the L TD vectors while in RBCV model the FB bits are used to quantize only one TD vector.

Here we derive the optimum bit allocation across the L TD vectors $\{\mathbf{p}(i)\}$. Let's introduce $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_L]$, where $\beta_i \geq 0$, $i = 0, \ldots, L-1$, $\sum_{i=0}^{L-1} \beta_i = 1$, and say $b_{CDI}\beta_i$ the number of bits assigned to the quantization of $\mathbf{p}(i)$. Quantities $\{\beta_i\}$ are chosen to maximize the expectation of the performance metric (6.13) with respect to both channel statistics and codebook realizations. In the analysis, for the sake of simplicity, we assume that different $\mathbf{p}(i)$ are quantized with independently generated codebooks. Using the inverse of (6.16) and after some algebra, the expectation of (6.13) can be expressed in terms of the TD vectors as

$$\mathrm{E}\left[\mu(\mathbf{H}, \hat{\mathbf{H}})\right] \;=\; \mathrm{E}\left[\sum_{\ell=0}^{L-1} \frac{1}{||\mathbf{h}(\ell)||^4} \sum_{i=0}^{L-1} ||\mathbf{p}(i)||^4 |\tilde{\mathbf{p}}(i)\hat{\mathbf{p}}(i)|^2\right] + C \,, \tag{6.18}$$

where C is a constant not involved in the optimization. From the independence between the norm of a TD vector and its direction and defining $z(i) = \mathrm{E}\left[||\mathbf{p}(i)||^4 \sum_{\ell=0}^{L-1} \frac{1}{||\mathbf{h}(\ell)||^4}\right]$, from (6.18) we have

$$\mathrm{E}\left[\mu(\mathbf{H}, \hat{\mathbf{H}})\right] = \sum_{i=0}^{L-1} z(i)\, \mathrm{E}\left[|\tilde{\mathbf{p}}(i)\hat{\mathbf{p}}(i)|^2\right] + C \,. \tag{6.19}$$

Optimization of (6.19) with respect to quantization bits yields the following constrained problem[3]

[3]Quantizing a unit norm vector according to the minimum chordal distance [50], i.e. the maximum inner

$$\max_{\boldsymbol{\beta}} \sum_{i=0}^{L-1} z(i)\, \mathrm{E}\left[|\tilde{\mathbf{p}}(i)\hat{\mathbf{p}}(i)|^2\right] \tag{6.20a}$$

$$\beta_i \geq 0\,, i = 0, \ldots, L-1 \tag{6.20b}$$

$$\textstyle\sum_{i=0}^{L-1} \beta_i \leq 1\,. \tag{6.20c}$$

We focus on the special case in which the channel vectors $\mathbf{h}(\ell)$ have i.i.d. complex Gaussian entries with zero mean and unit variance. Moreover we adopt the quantization upper bound (QUB) [88] based on the assumption that each quantization cell is a Voronoi region of a spherical cap with surface area $2^{-b_{CDI}\beta_i}$ of the total surface area of the unit sphere. With these assumptions we obtain

$$\mathrm{E}\left[|\tilde{\mathbf{p}}(i)\hat{\mathbf{p}}(i)|^2\right] = 1 - \left(\frac{M-1}{M}\right) 2^{-\frac{b_{CDI}\beta_i}{M-1}}\,. \tag{6.21}$$

Using (6.21) the optimization problem (6.20) can be rewritten as the convex problem

$$\min_{\boldsymbol{\beta}} \sum_{i=0}^{L-1} z(i)\, 2^{-\frac{b_{CDI}\beta_i}{M-1}} \tag{6.22a}$$

$$\beta_i \geq 0\,, i = 0, \ldots, L-1 \tag{6.22b}$$

$$\textstyle\sum_{i=0}^{L-1} \beta_i \leq 1\,. \tag{6.22c}$$

The following Lemma provides a sufficient condition for the optimality of RBCV-Q.

Lemma 2 *Under the assumptions of i) independent TD vectors $\{\mathbf{p}(i)\}$ and ii) QUB approximation, the solution of (6.22) is of the type $\beta_j = 1$ and $\beta_i = 0$, $i \neq j$, i.e. RBCV-Q is optimum according to the performance metric (6.13) (see also (6.17)), when*

$$b_{CDI} \lessapprox (M-1)\log_2\left(\frac{E[||\mathbf{p}(j)||^4]}{E[||\mathbf{p}(i)||^4]}\right)\,, \quad \forall\, i \neq j\,. \tag{6.23}$$

Proof: The Lagrangian for the convex problem (6.22) is given by

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \eta) = \sum_{i=0}^{L-1} z(i)\, 2^{-\frac{b_{CDI}\beta_i}{M-1}} + \eta\left(\sum_{i=0}^{L-1}\beta_i - 1\right) - \sum_{i=0}^{L-1}\lambda_i\beta_i \tag{6.24}$$

where $\boldsymbol{\lambda} = [\lambda_0, \ldots, \lambda_{L-1}]^T$ comprises the dual variables for (6.22b) and η is the dual variable for (6.22c). Computing the derivatives of the Lagrangian with respect to both $\{\beta_i\}$ and dual

product, causes an uncertainty of a common phase rotation of all the elements of the vector, i.e. $|\tilde{\mathbf{p}}(i)\hat{\mathbf{p}}(i)|^2 = |\tilde{\mathbf{p}}(i)\hat{\mathbf{p}}(i)e^{j\phi_i}|^2$, $\forall\phi_i \in [0, 2\pi)$. In the time domain quantization strategy the transmitter needs to reconstruct the RB channel matrix $\bar{\mathbf{H}}$ from the quantized TD vectors $\{\mathbf{p}(i)\}$ to perform beamforming. Therefore the phase ϕ_i along with the vector norm $||\mathbf{p}(i)||$ are essential and a fraction of the available FB bits should be used to characterize these quantities for each quantized tap $i \neq j$. In the following we neglect this problem because we are only interested in deriving a bound for RBCV-Q optimality when the channel matrix is quantized with only one vector.

variables leads to the the Karush-Kuhn-Tucker (KKT) conditions

$$\frac{b_{CDI} z(i)}{(M-1)} \, 2^{-\frac{b_{CDI}\beta_i}{M-1}} = \eta - \lambda_i \tag{6.25a}$$

$$\textstyle\sum_{i=0}^{L-1} \beta_i = 1 \,, \quad \beta_i \geq 0 \,, \quad i = 0, \ldots, L-1 \tag{6.25b}$$

$$\lambda_i \geq 0 \,, \quad i = 0, \ldots, L-1 \tag{6.25c}$$

$$\lambda_i \beta_i = 0 \,. \tag{6.25d}$$

In particular, a solution of the type $\beta_j = 1$ and $\beta_i = 0$, $i \neq j$ (i.e. RBCV-Q is optimum), is obtained when

$$b_{CDI} < (M-1) \log_2 \left(\frac{z(j)}{z(i)} \right) \quad \forall i \neq j \,. \tag{6.26}$$

Finally, using the approximation $z(i) \simeq \mathrm{E}[||\mathbf{p}(i)||^4]$, validated by simulations, leads to (6.23). □

As an example we consider a system with $M = 4$ transmit antennas, $N_C = 256$ subcarriers and $L = 12$ adjacent subchannels per RB. We use a *frequency selective* Rayleigh fading MIMO *channel* (FSC) with an exponential power delay profile and independent channel taps. The two strategies RBCV-Q and RBCM-Q are compared in Fig. 6.1 for different values of the root mean square delay spread normalized with respect to the sampling period, τ_{rms}. We recall that the boundary (6.23) for RBCV-Q optimality was achieved using the average performance metric (6.13). Differently, in numerical simulations, we compare RBCV-Q and RBCM-Q performance adopting the achievable throughput (6.5) as performance metric with quantization rule (6.7). Moreover for RBCM-Q a different codebook was generated for each value of τ_{rms} using the LBG algorithm with the performance metric (6.13), while for RBCV-Q the same codebook is used for all channels. We observe that the simpler RBCV-Q approach is optimal, and hence RBCM-Q is not needed, for channels with low τ_{rms}, while RBCM-Q is preferable for channels with a smaller coherence bandwidth or with a higher FB rate. Nevertheless, in practical cellular environments [60] we usually have $\tau_{rms} \leq 4$, therefore RBCV-Q is the best approach in practical channel conditions even with high FB rates.

It's worth underlining that although the theoretical boundary (6.23) was derived under particular conditions, it still gives an indication (pessimistic) of the optimality of RBCV-Q. Unfortunately, due to high complexity of RBCM-Q with high FB rate, the comparison between RBCM-Q and RBCV-Q could be verified only for $b_{CDI} \leq 11$. Nevertheless we can trivially argue that if $\tau_{rms} \to 0$, RBCV-Q becomes optimum for any FB rate because the channel on each RB becomes flat. This consideration is validated by the behaviour of the theoretical boundary (6.23) in Fig. 6.1 when $\tau_{rms} \to 0$.

6.4 DFB vs BeFB

In this section we compare DFB with BeFB, generalizing the analysis in [42] to multiuser MIMO-OFDM. Initially we use an *ideal* i.i.d. MIMO *channel* model (Id-C) where the frequency response is constant within a RB and independent across different RBs and resort to

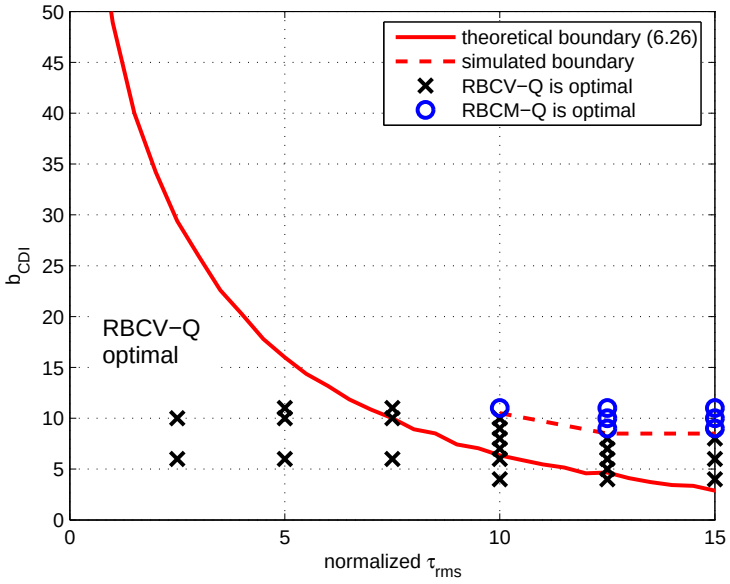

Figure 6.1: Comparison between RBVC-Q and RBCM-Q in different channel conditions. $M = 4$ and FSC model.

the QUB approximation for channel quantization. This provides an approximation of a real channel when each RB has a spectral width comparable to the channel coherence bandwidth. However, in Section 6.5 we provide numerical simulations for more realistic channel models, validating the analytical results achieved with the Id-C. Under Id-C conditions the user selection algorithm (6.11) simplifies to the SUS algorithm while the quantization encoding rule (6.7) is equivalent to the maximum inner product [42]. Moreover the CQI for user k in all subcarriers of RB n simplifies to $\gamma_{k,n}(\ell) = \gamma_{k,n}$ computed with $\hat{\mathbf{h}}_{k,n}(\ell) = \hat{\mathbf{h}}_{k,n}$, $\ell = 0, \ldots, L-1$. We claim, as in [42], that even if the exact SINR for user k after SUS is unknown at the transmitter, $\gamma_{k,n}$ provides a close approximation of the SINR for small values of ϵ in (6.10).

To find an expression of the achievable throughput, we need the distribution of $\gamma_{k,n}$ for a *selected* user which depends on both the FB scheme (see Tab. 6.1) and the user selection algorithm. We recall that according to the SUS algorithm, the $(i+1)$th selected user on RB n has the highest $\gamma_{k,n}$ among $|\mathcal{A}_n^{(i)}|$ users with independent channels and the same average SNR. Moreover in DFB the statistic of $\gamma_{k,n}$ is different from BeFB due to the maximization performed across N_R i.i.d RB channels. As a consequence, let $N_n^{(i)}$ be the number of i.i.d. random variables among which the maximum $\gamma_{k,n}$ is chosen at step $i+1$ of the SUS algorithm. In DFB $N_n^{(i)} = |\mathcal{A}^{(i)}(n)|$, while in BeFB $N_n^{(i)} = N_R|\mathcal{A}^{(i)}(n)|$, because we account for the maximization performed across the N_R RBs of each user.

Let $\bar{\gamma}_{i:U,n}$ be the ith largest order statistics among U i.i.d random variables $\{\gamma_{k,n}\}$. The user selection rule (6.11) can be seen as the selection of the $(i+1)$th largest order statistics in a set with $\bar{N}_n^{(i)}$ elements all having the same statistics, where $\bar{N}_n^{(i)} = N_n^{(i)} + 1 = |\mathcal{A}^{(i)}(n)| + i$ in DFB, $\bar{N}_n^{(i)} = N_n^{(i)} + 1 = N_R|\mathcal{A}^{(i)}(n)| + i$ in BeFB and i accounts for the number of users already selected. An approximated expression for the achievable throughput per RB is given

by

$$\mathrm{E}[R] \simeq \mathrm{E}\left\{ L \sum_{i=1}^{M} \log_2 \left(1 + \bar{\gamma}_{i:\bar{N}_n^{(i-1)},n} \right) \right\} , \tag{6.27}$$

where the approximation takes into account that $\gamma_{k,n}$ is the SINR after beamforming in case M users with orthogonal CDIs are selected at transmitter.

6.4.1 Asymptotic analysis of DFB and BeFB

In this section we derive an approximation of (6.27) in case of many users, i.e. large K. First we note that $|\mathcal{A}_n^{(i)}|$ is in general a random variable depending on the selection of the RB channels to be fed back by each user. Focusing on $\mathcal{A}_n^{(0)}$, in DFB, user k belongs to $\mathcal{A}_n^{(0)}$ only if RB n has been selected for transmission. Since we assume the selection to be independent of the user channel conditions, each RB has the same probability of being selected for FB. Therefore the probability of feeding back CSI of RB n is D/N_R and applying the law of large numbers, when K is large, we approximate the cardinality of the initial user set as $|\mathcal{A}_n^{(0)}| \simeq (KD)/N_R$. Differently, in BeFB, user k belongs to $\mathcal{A}_n^{(0)}$ only if $\gamma_{k,n}$ is the maximum among the N_R i.i.d. RB channels. A further application of the law of large numbers yields $|\mathcal{A}_n^{(0)}| \simeq K/N_R$.

Let α_i be the probability that a user belongs to $\mathcal{A}_n^{(i)}$. An approximation of $|\mathcal{A}_n^{(i)}|$ can be derived from $|\mathcal{A}_n^{(0)}|$ by applying the law of large numbers on K, i.e. $|\mathcal{A}_n^{(i)}| \simeq \left(|\mathcal{A}_n^{(0)}| - i \right) \alpha_i$, and we finally get

$$\text{DFB} \qquad \bar{N}_n^{(i)} = |\mathcal{A}_n^{(i)}| + i \simeq \left(\frac{KD}{N_R} - i \right) \alpha_i + i = \frac{KD}{N_R} \alpha_i + O(1) , \tag{6.28}$$

$$\text{BeFB} \qquad \bar{N}_n^{(i)} = N_R |\mathcal{A}_n^{(i)}| + i \simeq (K - iN_R)\alpha_i + i = K\alpha_i + O(1) . \tag{6.29}$$

From [42, Theorem 1], (6.28) and (6.29) an approximation of (6.27) in case of many users is

$$\mathrm{E}[R] \simeq L \sum_{i=1}^{M} \log_2 \left(1 + \rho \log \frac{2^{b_{CDI}} U \alpha_{i-1}}{\rho^{M-1}} \right) , \tag{6.30}$$

where $U = (KD)/N_R$ for DFB while $U = K$ for BeFB. The term $\Delta = \log \frac{2^{b_{CDI}} U \alpha_{i-1}}{\rho^{M-1}}$ in (6.30) can be interpreted as the SNR variation, which includes the effects of quantization error, frequency and multiuser diversity. From Tab. 6.1, it is straightforward to prove that for large K, exploiting multiuser diversity, (6.30) is maximized for DFB when $D = 1$, i.e. when all the available FB bits are used to characterize only one RB. This behaviour does not change if a larger fraction of the available FB bits is left for CQI updating. Interestingly, for a given SNR variation Δ that assures a constant gap from ZF beamforming with perfect CSIT, B and K should scale with P and N_R as

$$\text{DFB} \qquad \tfrac{B}{D} + \log_2(KD) = (M-1)\log_2 P + \log_2 N_R + c , \tag{6.31a}$$

$$\text{BeFB} \qquad B + \log_2 K = (M-1)\log_2 P + \log_2 N_R + c , \tag{6.31b}$$

where c depends on Δ. It is worth noticing that DFB with $D = 1$ gives the same scaling law as BeFB. Since in BeFB a portion $\log_2(N_R)$ of the available FB bits is used to index the selected RB, this suggests that the same performance of DFB with $D = 1$ can be achieved exploiting frequency diversity and using a smaller codebook, hence requiring less memory. Finally, DFB with $D > 1$ requires almost D times FB bits to achieve the same throughput of BeFB.

It is important to observe that (6.30)-(6.31) are valid only in a *large user regime* as $K \to \infty$. For finite K, N_R and B, if P is too large, the system enters the *interference-limited regime*. Following similar arguments applied in [42], an approximated expression for the achievable throughput per RB becomes

$$\mathrm{E}[R] \simeq \frac{L\,M}{M-1}(b_{CDI} + \log_2 U) + \frac{L\sum_{i=1}^{M}\log_2 \alpha_{i-1}}{M-1}\,. \tag{6.32}$$

Again DFB with $D = 1$ and BeFB achieve the highest achievable throughput, with BeFB requiring less memory. Finally for finite K and N_R, if either B is too large or P is too small, the system falls in the *high resolution or noise-limited region* and the achievable throughput may be approximated as

$$\mathrm{E}[R] \simeq L\sum_{i=1}^{M}\log_2(1 + \rho\log_2 U\alpha_{i-1})\,. \tag{6.33}$$

In this case, the best choice for DFB becomes $D = N_R$, because quantization noise does not significantly limit the achievable throughput and having an higher multiuser diversity per RB is preferable. Still, BeFB achieves the highest achievable throughput, as DFB with $D = N_R$, because the loss in multiuser diversity is compensated by the gain in frequency diversity.

6.5 Simulation results

In this section we present numerical results to validate the asymptotic analysis of Section 6.4.1. We assume a transmitter equipped with $M = 4$ antennas and $\epsilon = 0.35$ is the correlation parameter in (6.10). The OFDM system has $N_C = 256$ subcarriers and $N_R = 8$ RBs with $L = 12$ subcarriers each. The channel is assumed spatially uncorrelated and we adopt two different channel models: i) the Id-C introduced in Section 6.4 and ii) the FSC with $\tau_{rms} = 2.5$ introduced in Section 6.3. In the Id-C we use QUB approximation while in the FSC we perform RBCV-Q using an LBG-based codebook with (6.7) as quantization rule.

In Fig. 6.2 we set $B = 12$ as the total amount of FB bits per user and consider $K = 200$ users in the system. We compare DFB and BeFB in terms of the achievable throughput as a function of the average SNR. As term of comparison we also include the quantization scheme proposed in [89] that performs channel quantization in the time domain using a mean square error (MSE) criterion. Since all channel taps are independent across both space and time domains and circularly symmetric, quantization is performed independently on each component of the channel taps using an uniform quantizer. We denote this scheme as *time-domain uniform quantization* (TD-UQ). We underline that to minimize the MSE the quantization bits are distributed across the different channel taps according to the power delay profile of the channel,

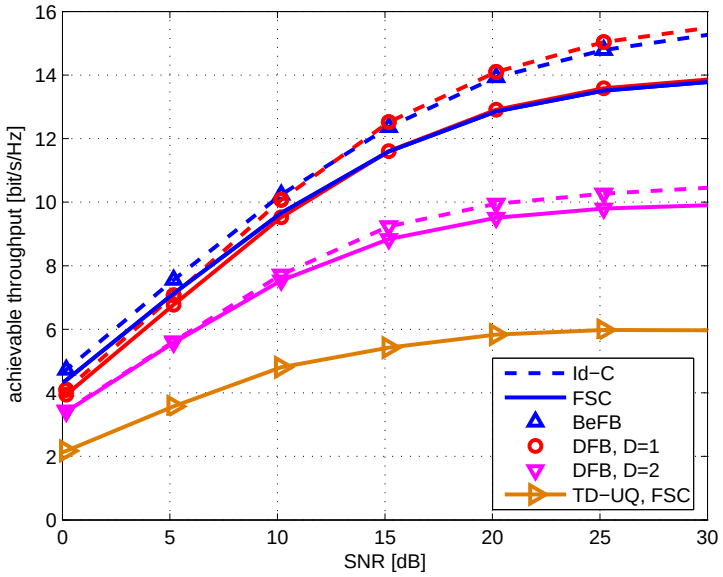

Figure 6.2: Comparison between DFB, BeFB and TD-UQ in terms of achievable throughput vs. average SNR. $M = 4$, $K = 200$ and $B = 12$. Two spatially uncorrelated channels are considered: i) Id-C, ii) FSC with $\tau_{rms} = 2.5$.

using bit allocation strategies proposed in [94].

We observe how BeFB and DFB with $D = 1$ yield very close performance in the high SNR region while BeFB is preferable at low-SNR thanks to its ability of exploiting frequency diversity as predicted in Section 6.4.1. Eventually, for increasing SNR the system becomes interference limited as observed in (4.28). Moreover both BeFB and DFB with $D = 1$ outperform DFB with $D > 1$, thanks to a better policy in channel quantization and feedback signalling. We notice that TD-UQ allocates the available FB bits across all RBs similarly to DFB with $D = N_R$ and therefore it has very poor performance with respect to BeFB and DFB with $D = 1$ when K is large. Similar considerations hold even when the uniform quantizer is substituted by the optimum scalar quantizer. Interestingly, the relative performance of the FB strategies are similar in both types of channels (Id-C and FSC). Only a small degradation in the achievable throughput is observed for the FSC, mainly due to the frequency variability of the channel within a RB.

Fig. 6.3 compares the achievable throughput of DFB and BeFB as a function of the number of users K in the system for an average SNR $= 20$ dB. From the asymptotic analysis of Section 6.4.1 for fixed B and P the system gets into the large-system regime as $K \to \infty$, and in these conditions BeFB and DFB with $D = 1$ significantly outperforms DFB with $D > 1$. In practice Fig. 6.3, besides investigating the gap between the proposed FB schemes, shows that the asymptotic results give useful indications even for practical values of K. Indeed, although the optimum value of D for DFB depends on K, DFB with $D = 1$ outperforms DFB with $D > 1$ already for $K = N_R$. Interestingly TD-UQ is preferable for lower values of K but becomes a bad choice when K increases. We underline that for $K < N_R$ there is a non zero probability that some subcarriers are not used for transmission in BeFB and DFB with small D.

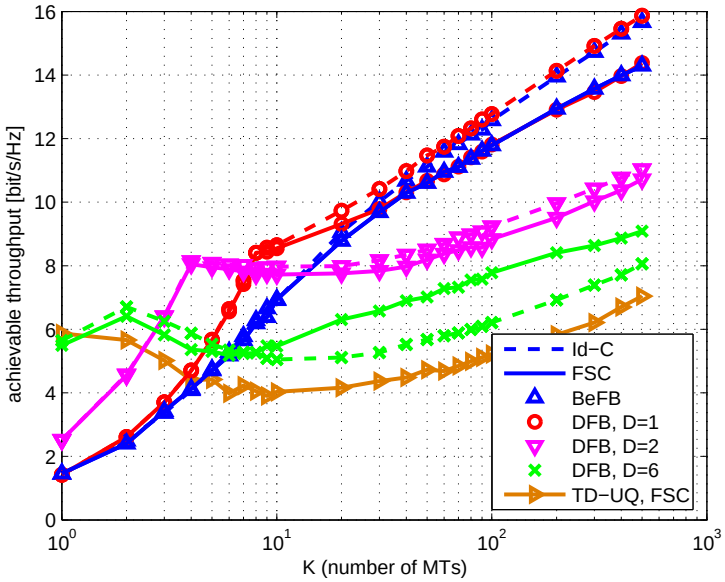

Figure 6.3: Comparison between DFB, BeFB and TD-UQ in terms of achievable throughput vs number of users K. $M = 4$, $B = 12$ and $SNR = 20$ dB. Two spatially uncorrelated channels are considered: i) Id-C and ii) FSC with $\tau_{rms} = 2.5$.

Finally in Fig. 6.4, for the Id-C and considering $K = 512$ users, we show how scaling the number of FB bits B according to (6.31b) with $c = -6$, both BeFB and DFB with $D = 1$ can assure a constant gap from the curve of perfect CSIT. Differently, in DFB with $D > 1$ scaling the FB rate as in (6.31b) is not sufficient to guarantee a constant SNR variation; this would be achieved only by scaling B according to (6.31a). The same behaviour has been observed even under the FSC for practical values of $B \leq 12$.

6.6 Conclusions

The chapter considers the problem of channel quantization and FB optimization in multiuser MIMO-OFDM downlink systems. As in current standard proposals for next generation wireless networks, to reduce the control overhead and the signal processing complexity the available bandwidth is divided into RBs whose spectral width reflects the coherence bandwidth of the channel.

From both the analysis and the numerical results of the chapter we can derive two main contributions. Firstly, we show that quantizing the RBCM by a single space vector is optimum for practical values of the FB rate and in typical dispersive channels, i.e. RBCV-Q is to be preferred to RBCM-Q. In this context we also derive a new simplified performance metric for codebook design in RB channel quantization, which is related to the system achievable throughput.

As a second contribution, an asymptotic analysis of the system throughput derived for a large number of users reveals that allocating all the available FB bits to quantize only the

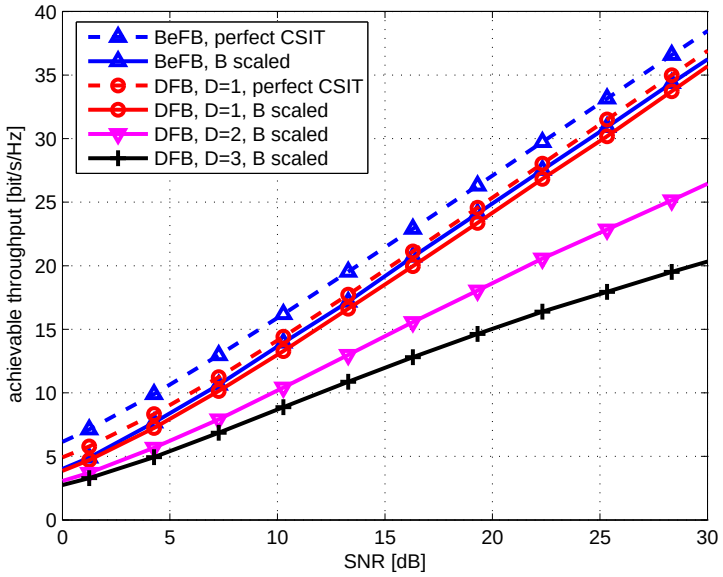

Figure 6.4: Comparison between DFB and BeFB in terms of achievable throughput vs average SNR. QUB is adopted scaling the FB rate B with P as in (6.31b). $K = 512$ and Id-C model.

channel relative to one RB provides a significant gain in achievable throughput over a more classic distributed FB approach and simulations validate these considerations even for a moderate number of users in the network. We conclude that, in case of limited uplink feedback, for a practical number of users in the network, accurate channel knowledge is preferable to high frequency/multiuser diversity.

Chapter 7

On state estimation in networked control systems

Recent research activities and technological progress in communication theory, DSP capabilities and computing are revolutionizing our capabilities to build distributed sensor networks [95] that, by offering access to unprecedented quality and quantity of information can improve tremendously our possibilities and abilities of monitoring and controlling the environment. In these networked control systems (NCSs) [17] the aim is to estimate or control one or more dynamical systems, using multiple sensors, actuators and controllers that are not physically co-located and need to exchange information via a wireless digital communication network. Typical applications vary from environmental control [96], vehicular networks and traffic control [97], surveillance systems, tracking in warehouse [98] and military scenarios [99].

In NCSs measurements and control packets are subject to random delay and loss, [18]. Moreover, as to each component is effectively allocated only a small portion of the available bandwidth, measurements and control information need to be quantized and this affects the stability of the system, [19]. This suggests that a cross-layer design of communication and estimation/control systems might provide significant performance improvement over a separated approach [100], leading to a generalization of classical control theory to account for the stochastic nature of the communication channel.

In this chapter we propose a framework for state estimation in NCSs where observations from multiple sensors are subject to random delays and packet losses, generalizing previous contributions [18, 101]. We derive the minimum error covariance estimator which, differently from standard Kalman filtering [102], is strongly time variant, stochastic and jumps between different estimation strategies as a function of received measurements. As a low-complexity solution we design the optimum estimator with constant gains which depends on the packet arrival probabilities. In case of a stable system we generalize the proposed solutions to account for the effects of measurements quantization and limited transmission bandwidth. Assuming a simple scalar system we show how the proposed framework can give useful indications in the design of NCSs in the presence of i) low-cost sensors using a fix modulation, and ii) long-term future sensors capable of rate adaptation. As examples of applications we investigate issues within two network set-ups: i) cross-layer optimization of quantization processes and network

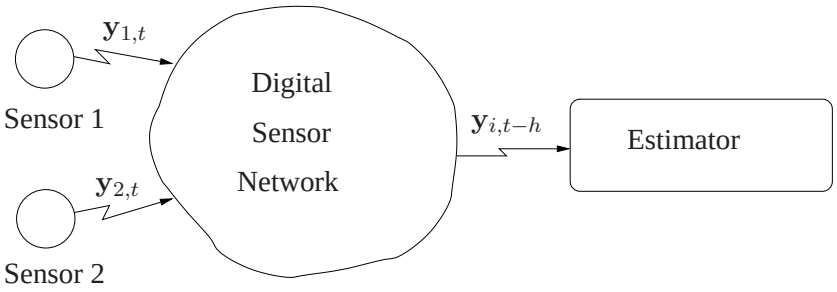

Figure 7.1: System setting.

resource allocation and ii) comparison between single-hop and multi-hop communication protocols.

The chapter is organized as follows. The problem of state estimation is introduced in Section 7.1, the minimum error covariance estimator is derived in Section 7.2 and a suboptimum estimator with constant gains is proposed in Section 7.3. Section 7.4 introduces the problem of measurements quantization and Section 7.5 present some examples of applications of the proposed framework for stable scalar systems. Finally Section 7.6 summarizes the main results of the chapter.

Part of the material in this chapter has been published in [103].

7.1 Problem formulation

We consider a general multi-input multi-output (MIMO) discrete time linear system and partition the observation vector into two parts, modelling the observation of the state by two different sensors. The system has the following dynamics:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{w}_k \tag{7.1}$$

$$\begin{bmatrix} \mathbf{y}_{1,k} \\ \mathbf{y}_{2,k} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} \mathbf{v}_{1,k} \\ \mathbf{v}_{2,k} \end{bmatrix} \tag{7.2}$$

where $\mathbf{x}_k, \mathbf{w}_k \in \mathbb{R}^n$ are the state of the system and the system noise at instant k and $\mathbf{y}_{1,k}, \mathbf{v}_{1,k} \in \mathbb{R}^{m_1}$, $\mathbf{y}_{2,k}, \mathbf{v}_{2,k} \in \mathbb{R}^{m_2}$, are the measurements and measurement noises at time k for sensor 1 and 2, respectively. Moreover $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{C}_1 \in \mathbb{R}^{m_1\times n}$, $\mathbf{C}_2 \in \mathbb{R}^{m_2\times n}$ and $(\mathbf{x}_0, \mathbf{w}_k, \mathbf{v}_{1,k}, \mathbf{v}_{2,k})$ are uncorrelated Gaussian random vectors with mean $(\bar{\mathbf{x}}_0, \mathbf{0}, \mathbf{0}, \mathbf{0})$ and covariance matrix given by $(\mathbf{P}_0, \mathbf{Q}, \mathbf{R}_1, \mathbf{R}_2)$. Furthermore we define $\mathbf{C} = \left[\mathbf{C}_1^T \mathbf{C}_2^T\right]^T$ and the covariance matrix of the total noise $\mathbf{v}_k = \left[\mathbf{v}_{1,k}^T \mathbf{v}_{2,k}^T\right]^T$ as $\mathbf{R} = \text{diag}\,(\mathbf{R}_1, \mathbf{R}_2)$. We also assume that the pair $(\mathbf{A}, \mathbf{C})$ is observable and $(\mathbf{A}, \mathbf{Q}^{1/2})$ is controllable.

The measurements outputs $\mathbf{y}_{1,k}$, $\mathbf{y}_{2,k}$ are encoded separately by two sensors, time-stamped, and transmitted through a digital sensor network (DSN) whose goal is to deliver packets from multiple sources (sensors) to a destination (estimator) possibly introducing random delays (see also Fig. 7.1). Time-stamping of measurements are necessary to reorder packets at the receiver in case they arrive out of order. We assume that transmitted packets incorporates error detection coding [20], therefore the estimator knows perfectly weather packets received from sensors

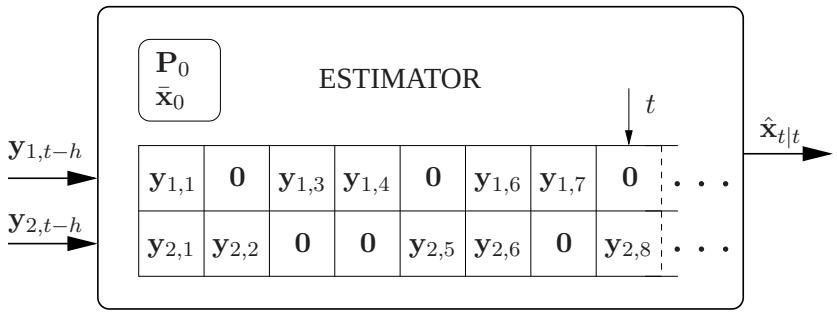

Figure 7.2: Estimator with infinite buffers.

contain errors or not. At estimator side all packets correctly delivered are collected in a queue with two infinite buffers, one for each sensor, while faulty packets are discarded (see also Fig. 7.2). The arrival process is modelled by the random variable $\gamma_{i,k}^t$, with $i \in \{1, 2\}$ denoting the sensor, in the following way[1]

$$\gamma_{i,k}^t = \begin{cases} 1 & \mathbf{y}_{i,k} \text{ has been received at } t \geq k \\ 0 & \text{otherwise} \end{cases} \tag{7.3}$$

From (7.3) it follows that if a packet is present in the buffer at time t then it will be present for all future times, i.e. if $\gamma_{i,k}^t = 1$ then $\gamma_{i,k}^{t+h} = 1, \forall h \in \mathbb{N}$. We also define the packet delay $\tau_{i,k}$, $i = 1, 2$, for observation $\mathbf{y}_{i,k}$ as follows:

$$\tau_{i,k} = \begin{cases} \infty & \gamma_{i,k}^t = 0, \forall t \geq k \\ t_{i,k} - k & \text{otherwise} \end{cases} \tag{7.4}$$

where $t_{i,k} \triangleq \min\{t | \gamma_{i,k}^t = 1\}$ denotes the arrival time of observation $\mathbf{y}_{i,k}$ at the estimator. Since the packet delay can be random, observation measurements can arrive out of order at the estimator, moreover we assume a packet is lost when $\tau_{i,k} = \infty$.

At first we do not consider the effects of measurement quantization, assuming that the distortion introduced by data encoding/decoding is negligible with respect to measurement noise. An extension of the proposed model to cope with quantization effects is investigated later in Section 7.4.

The value stored in the kth slot of the estimator buffer i at time t can be written as

$$\tilde{\mathbf{y}}_{i,k}^t = \gamma_{i,k}^t \mathbf{y}_{i,k} = \gamma_{i,k}^t \mathbf{C}_i \mathbf{x}_k + \gamma_{i,k}^t \mathbf{v}_{i,k}\,, \quad i = 1, 2 \tag{7.5}$$

where we assume a zero [18] is stored in slot k if observation $\mathbf{y}_{i,k}$ has not arrived at time t.

[1] With this model we are assuming that the channel coherence time is comparable to the length of the packet, therefore packet arrivals in subsequent time slots can be approximated as independent. Moreover the use of a Bernoulli random variable implies a "hard" detection/decoding strategy [20] where packets are simply dropped if they contain errors. Generalizations to correlated packet arrivals [104] and more powerful "soft" detection/decoding strategies, where detected bits are associated with reliability values [20], are left as future works.

7.2 Minimum error covariance estimator

In this section we derive the optimal mean square error estimator $\hat{\mathbf{x}}_{t|t} \triangleq \mathrm{E}\left[\mathbf{x}_t|\mathcal{F}_t^t, \bar{\mathbf{x}}_0, \mathbf{P}_0\right]$ where $\mathcal{F}_h^t = \left\{\widetilde{\mathbf{y}}_{1,h}^t, \widetilde{\mathbf{y}}_{2,h}^t, \boldsymbol{\gamma}_{1,h}^t, \boldsymbol{\gamma}_{2,h}^t\right\}$, $\widetilde{\mathbf{y}}_{i,h}^t = \left(\tilde{\mathbf{y}}_{i,1}^t, \ldots, \tilde{\mathbf{y}}_{i,h}^t\right)$, $\boldsymbol{\gamma}_{i,h}^t = \left(\gamma_{i,1}^t, \ldots, \gamma_{i,h}^t\right)$, $i = 1, 2$. Let $\mathbf{e}_{t|t} \triangleq \mathbf{x}_t - \hat{\mathbf{x}}_{t|t}$, estimator $\hat{\mathbf{x}}_{t|t}$ is optimal in the sense it minimizes the error covariance $\mathbf{P}_{t|t} \triangleq \mathrm{E}\left[\mathbf{e}_{t|t}\mathbf{e}_{t|t}^T|\mathcal{F}_t^t, \bar{\mathbf{x}}_0, \mathbf{P}_0\right]$. First let's introduce the following quantities

$$\hat{\mathbf{x}}_{k|h}^t \triangleq \mathrm{E}\left[\mathbf{x}_k|\mathcal{F}_h^t, \bar{\mathbf{x}}_0, \mathbf{P}_0\right] \tag{7.6}$$

$$\mathbf{P}_{k,h}^t \triangleq \mathrm{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|h}^t\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|h}^t\right)^T|\mathcal{F}_h^t, \bar{\mathbf{x}}_0, \mathbf{P}_0\right] \tag{7.7}$$

where with a little abuse of notation we can say $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t}^t$ and $\mathbf{P}_{t|t} = \mathbf{P}_{t|t}^t$.

Generalizing results from [18, Theorem 1] and [101], the optimal estimator can be derived from the following theorem whose proof is given in Appendix B.1.

Theorem 3 *Let's consider the stochastic linear system (7.1)-(7.2), the packet arrival process defined in (7.3) and the mean square error estimator $\hat{\mathbf{x}}_{t|t}$. Then the following facts hold:*

(a) *The optimal estimator is given by $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t}^t$ with:*

$$\begin{aligned}\hat{\mathbf{x}}_{k|k}^t = {} & \mathbf{A}\hat{\mathbf{x}}_{k-1|k-1}^t + \\ & \gamma_{1,k}^t\gamma_{2,k}^t\mathbf{K}_k^t\left(\tilde{\mathbf{y}}_k^t - \mathbf{C}\mathbf{A}\hat{\mathbf{x}}_{k-1|k-1}^t\right) + \\ & \gamma_{1,k}^t(1-\gamma_{2,k}^t)\mathbf{K}_{1,k}^t\left(\tilde{\mathbf{y}}_{1,k}^t - \mathbf{C}_1\mathbf{A}\hat{\mathbf{x}}_{k-1|k-1}^t\right) + \\ & (1-\gamma_{1,k}^t)\gamma_{2,k}^t\mathbf{K}_{2,k}^t\left(\tilde{\mathbf{y}}_{2,k}^t - \mathbf{C}_2\mathbf{A}\hat{\mathbf{x}}_{k-1|k-1}^t\right)\end{aligned} \tag{7.8}$$

$$\begin{aligned}\mathbf{P}_{k+1|k}^t = {} & \mathbf{A}\mathbf{P}_{k|k-1}^t\mathbf{A}^T + \mathbf{Q} + \\ & -\gamma_{1,k}^t\gamma_{2,k}^t\mathbf{A}\mathbf{K}_k^t\mathbf{C}\mathbf{P}_{k|k-1}^t\mathbf{A}^T + \\ & -\gamma_{1,k}^t(1-\gamma_{2,k}^t)\mathbf{A}\mathbf{K}_{i,k}^t\mathbf{C}_1\mathbf{P}_{k|k-1}^t\mathbf{A}^T + \\ & -(1-\gamma_{1,k}^t)\gamma_{2,k}^t\mathbf{A}\mathbf{K}_{2,k}^t\mathbf{C}_2\mathbf{P}_{k|k-1}^t\mathbf{A}^T,\end{aligned} \tag{7.9}$$

where $\hat{\mathbf{x}}_{0|0}^t = \bar{\mathbf{x}}_0$, $\mathbf{P}_{1,0}^t = \mathbf{P}_0$ are the initialization conditions and

$$\mathbf{K}_k^t = \mathbf{P}_{k|k-1}^t\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k|k-1}^t\mathbf{C}^T + \mathbf{R})^{-1} \tag{7.10}$$

$$\mathbf{K}_{i,k}^t = \mathbf{P}_{k|k-1}^t\mathbf{C}_i^T(\mathbf{C}_i\mathbf{P}_{k|k-1}^t\mathbf{C}_i^T + \mathbf{R}_i)^{-1}, \quad i = 1, 2 \tag{7.11}$$

are the Kalman gains.

(b) *The optimal estimator $\hat{\mathbf{x}}_{t|t}$ can be computed iteratively using a buffer of finite length N if and only if $\gamma_{i,k}^t = \gamma_{i,k}^{t-1}$, $i = 1, 2$, $\forall k \geq 1$, $\forall t \geq k + N$. In case this property is satisfied then $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t}^t$ where $\hat{\mathbf{x}}_{t|t}^t$ is given by (7.8)-(7.9) for $t = 1, \ldots, N$ and as follows for*

$t > N$:

$$\hat{\mathbf{x}}^t_{t-N|t-N} = \hat{\mathbf{x}}^{t-1}_{t-N|t-N} \quad (7.12)$$

$$\mathbf{P}^t_{t-N+1|t-N} = \mathbf{P}^{t-1}_{t-N+1|t-N} \quad (7.13)$$

$$\textit{Equations (7.8)-(7.9)} \qquad k = t-N+1, \ldots, t \quad (7.14)$$

□

We observe that there are two major differences with respect to the standard Kalman filter [102]. First the optimal estimator described by (7.8)-(7.9) jumps between different estimation strategies according to the values assumed by $\gamma^t_{i,k}$. In detail we have 1) an open loop estimate when $\gamma^t_{1,k} = \gamma^t_{2,k} = 0$, 2) a closed loop estimate when $\gamma^t_{1,k} = \gamma^t_{2,k} = 1$, for which state estimation evolution is the same as the classical Kalman filter, 3) estimation evolution as if $y_{1,k}$ were the only observation for $\gamma^t_{1,k} = 1$, $\gamma^t_{2,k} = 0$, and 4) estimation evolution as if $y_{2,k}$ were the only observation for $\gamma^t_{1,k} = 0$, $\gamma^t_{2,k} = 1$. Therefore, the optimal estimator and its error covariance are strongly time-variant and stochastic.

The second difference is that standard Kalman filter only needs the current measurement $\mathbf{y}_t$, the previous state estimate $\hat{\mathbf{x}}_{t-1|t-1}$ and the past error covariance $\mathbf{P}_{t|t-1}$ to compute the current state estimate $\hat{\mathbf{x}}_{t|t}$. Differently, in case of random delays and packet drops, for two sensors the optimal estimator needs i) to store all past packets and ii) to invert up to t matrices (see (7.8) and (7.9)) at any time instant t, with a linear increase of complexity as time progresses. Moreover, a buffer of finite length N can be used if and only if all packets correctly delivered arrive at estimator within N time steps (Theorem 3(b)). We note that in DSNs transmission protocols usually drop from transmission buffer packets that are very old, e.g. older than N time slots. Therefore the problem of computing the optimal estimator with a finite buffer is quite common in DSNs.

7.3 Optimum estimator with constant gains

The great complexity of the optimal estimator motivates the investigation of suboptimum but more practical solutions for state estimation. In this section we study a filter with constant gains and a buffer with finite dimension N, i.e. $\mathbf{K}^t_{t-h} = \mathbf{K}_h$, $\mathbf{K}^t_{i,t-h} = \mathbf{K}_{i,h}$, $i = 1, 2$, $\forall t \in \mathbb{N}$ and $h = 0, \ldots, N-1$. The gains for the different state evolution scenarios are designed to achieve the smallest error covariance at steady-state, assuming stationary i.i.d. arrival processes.

We model the packet arrival process at estimator relative to sensor i with the probability function

$$\mathrm{P}[\tau_{i,t} \leq h] = \lambda_{i,h} \quad (7.15)$$

where $t \geq 0$ and $0 \leq \lambda_{i,h} \leq 1$ is clearly non decreasing in $h \geq 0$. Moreover we define the packet loss probability as

$$\lambda_{i,loss} \triangleq 1 - \sup\{\lambda_{i,h} | h \geq 0\} . \quad (7.16)$$

Finally we denote the maximum delay of arrived packets relative to sensor i as $\tau_{i,max}$ and define $\tau_{max} \triangleq \max_i\{\tau_{i,max}\}$. We underline that the arrival processes relative to the two sensors are

independent and described by possibly different probability functions.

We consider a constant-gains estimator $\tilde{\mathbf{x}}_{t|t} = \tilde{\mathbf{x}}^t_{t|t}$ with finite buffer of dimension N, where $\tilde{\mathbf{x}}^t_{t|t}$ is recovered iteratively from $\tilde{\mathbf{x}}^t_{t-h|t-h}$, $h = 0, \ldots, N-1$ as $\hat{\mathbf{x}}^t_{t|t}$ in (7.12)-(7.14), but using constant gains, i.e. $\mathbf{K}^t_{t-h} = \mathbf{K}_h$, $\mathbf{K}^t_{i,t-h} = \mathbf{K}_{i,h}$, $i = 1, 2$, $\forall t \in \mathbb{N}$ and $h = 0, \ldots, N-1$.

Let $\tilde{\mathbf{e}}^t_{t-k+1|t-k} = \mathbf{x}_{t-k+1} - \mathbf{A}\tilde{\mathbf{x}}^t_{t-k|t-k}$ be the prediction error and consider the prediction error covariance matrix $\overline{\mathbf{P}}^t_{t-k+1|t-k} = \mathrm{E}\left[\tilde{\mathbf{e}}^t_{t-k+1|t-k}(\tilde{\mathbf{e}}^t_{t-k+1|t-k})^T\right]$ where expectation is computed with respect to both initial conditions and arrival processes. Moreover define the following operator:

$$\begin{aligned}\Phi_{\lambda_1,\lambda_2}(P) &= \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{Q} - \lambda_1\lambda_2\mathbf{A}\mathbf{K}_P\mathbf{C}\mathbf{P}\mathbf{A}^T + \\ &\quad -\lambda_1(1-\lambda_2)\mathbf{A}\mathbf{K}_{1,P}\mathbf{C}_1\mathbf{P}\mathbf{A}^T + \\ &\quad -(1-\lambda_1)\lambda_2\mathbf{A}\mathbf{K}_{2,P}\mathbf{C}_2\mathbf{P}\mathbf{A}^T \end{aligned} \tag{7.17}$$

with the gains $\mathbf{K}_P = \mathbf{P}\mathbf{C}^T(\mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R})^{-1}$, $\mathbf{K}_{i,P} = \mathbf{P}\mathbf{C}_i^T(\mathbf{C}_i\mathbf{P}\mathbf{C}_i^T + \mathbf{R}_i)^{-1}$, $i = 1, 2$. The following theorem, whose proof is outlined in Appendix B.2, derives the optimal constant-gains estimator that minimizes the steady state error covariance.

Theorem 4 *Let us consider the stochastic linear system given in (7.1)-(7.2), where i)* $(\mathbf{A}, \mathbf{C})$ *is observable and* $(\mathbf{A}, \mathbf{Q}^{1/2})$ *is controllable, ii) arrival processes are defined at the beginning of the section and iii) estimators have constant gains* $\{\mathbf{K}_k\}_{k=0}^{N-1}$, $\{\mathbf{K}_{i,k}\}_{k=0}^{N-1}$, $i = 1, 2$. *The following statements hold:*

- ***(a)*** *If* $\mathbf{A}$ *is not strictly stable there exists an instability region for the couple* $(1-\lambda_{1,loss}, 1-\lambda_{2,loss})$ *for which no stable estimator with constant gains exists. More generally a different stability region can be found for each value of* N. *Hence given* $0 \leq \lambda_{1,N-1} \leq 1$ *there exists* $\lambda^c_{2,N-1}$ *such that a stable estimator exists if and only if* $\lambda_{2,N-1} > \lambda^c_{2,N-1}$. *Both a lower bound* $\underline{\lambda}^c_{2,N-1}$ *and an upper bound* $\overline{\lambda}^c_{2,N-1}$ *can be derived for* $\lambda^c_{2,N-1}$, *i.e.* $\underline{\lambda}^c_{2,N-1} \leq \lambda^c_{2,N-1} \leq \overline{\lambda}^c_{2,N-1}$, *[101]. Similarly a bound* $\lambda^c_{1,N-1}$ *can be derived for* $\lambda_{1,N-1}$ *when* $\lambda_{2,N-1}$ *is fixed.*

- ***(b)*** *Let* N *such that* $(\lambda_{1,N-1}, \lambda_{2,N-1})$ *belongs to the stable region, the optimal gains* $\{\mathbf{K}^N_k\}_{k=0}^{N-1}$, $\{\mathbf{K}^N_{i,k}\}_{k=0}^{N-1}$, $i = 1, 2$ *are defined as follows:*

$$\mathbf{K}^N_k = \mathbf{V}^N_k\mathbf{C}^T(\mathbf{C}\mathbf{V}^N_k\mathbf{C}^T + \mathbf{R})^{-1}, \tag{7.18}$$

$$\mathbf{K}^N_{i,k} = \mathbf{V}^N_k\mathbf{C}_i^T(\mathbf{C}_i\mathbf{V}^N_k\mathbf{C}_i^T + \mathbf{R}_i)^{-1}, \tag{7.19}$$

$$\mathbf{V}^N_{N-1} = \Phi_{\lambda_{1,N-1},\lambda_{2,N-1}}(\mathbf{V}^N_{N-1}) \tag{7.20}$$

$$\mathbf{V}^N_k = \Phi_{\lambda_{1,k},\lambda_{2,k}}(\mathbf{V}^N_{k+1}),\ k = N-2, \ldots, 0 \tag{7.21}$$

Moreover $\lim_{t\to\infty}\overline{\mathbf{P}}^t_{t-k+1|t-k} = \mathbf{V}^N_k$, *independently of the initial condition* $(\mathbf{P}_0, \bar{\mathbf{x}}_0)$. *Also* $\mathbf{V}^{N+1}_0 \leq \mathbf{V}^N_0$. *Finally, let* $\tau_{max} = \max_i\{\tau_{i,max}\}$, *if* $\tau_{max} < \infty$, *then* $\mathbf{V}^N_0 = \mathbf{V}^{\tau_{max}}_0$ *for all* $N \geq \tau_{max}$.□

Theorem 4 shows how the optimal estimator with constant gains can be derived solving an algebraic matrix equation (7.20) and then iterating $N-1$ times operator (7.21) having the same

structure of (7.20) but with different values of $\lambda_{i,k}$. In case the system is unstable a stable estimator exists if and only if the packet loss probabilities for the two links are sufficiently small or more precisely if and only the couple $(\lambda_{1,N-1}, \lambda_{2,N-1})$ belongs to the stability region which is a function of N and system parameters. Interestingly a stable estimator with constant gains exists if and only if the optimal estimator with constant gains exists. Furthermore estimator stability does not depend on the packet delay τ_{max} as long as a sufficient number of packets eventually arrive. Another important result is that the performance of the estimator improves as the buffer length N increases but at the same time a longer buffer implies higher computation complexity. Therefore a natural trade-off arises between estimation performance and complexity. However if the maximum packet delay for both sensors is finite, i.e. $\tau_{max} < \infty$, then the performance of the estimator does not improve for $N > \tau_{max}$.

7.4 Quantization processes and transmission strategies

In this section we show how the proposed framework can be adapted to account for the effects of measurements quantization and limited transmission bandwidth. We refer to a simple scalar system but the technique can be generalized to more complex MIMO systems.

Let us consider a scalar, stable system with $A < 1$, $C_1 = C_2 = 1$, $R_1, R_2 > 0$, i.e. the system is observed by two sensors with possible different features. Let $\hat{y}_{i,k}$ be the quantized version of $y_{i,k}$. We model the effects of measurements quantization (source coding) as an additive noise, i.e. $y_{i,k} = \hat{y}_{i,k} + n_{i,k}(B_i)$, whose distribution depends on both the quantization strategy and the number of quantization bits B_i. We note that the measurement can be modelled as a Gauss-Markov process $y_{i,k+1} = Ay_{i,k} + \nu_{i,k}$ where $\nu_{i,k} = w_{k-1} + (1-A)v_{i,k}$ is a Gaussian random variable with zero mean and variance $\sigma_\nu^2 = Q + (1 - A^2)R_i$. Moreover y_k results a Gaussian source with zero mean and variance $\sigma_y^2 = \frac{\sigma_\nu^2}{1-A^2}$. In case of small distortion and uniform quantization $n_{i,k}(B)$ is white noise with zero mean and variance $\sigma_n^2(B_i) = 3\sigma_y^2 2^{-2B_i}$ [94]. The effects of measurement quantization can be included in the proposed framework considering the equivalent measurement noise $\bar{v}_{i,k} = v_{i,k} + n_{i,k}(B_i)$ having zero mean and variance $\bar{R}_i = R_i + \sigma_n^2(B_i)$.

We adopt a time division multiple access (TDMA) as medium access control (MAC) strategy where sensor i is allocated a portion T_i of the available time slot T. We assume a total transmission bandwidth W_B and a block fading model with independent Rayleigh fading realizations for all links, i.e. each radio link gain is modelled as $\sqrt{\Gamma_i} h_{i,k}$ with $h_{i,k} \sim \mathcal{CN}(0, 1)$ and Γ_i, $i \in \{1, 2\}$, denoting the average link SNR. Moreover we assume that only the average SNR is available at the two transmitters. The probability of packet loss depends on channel conditions, number of quantization bits, modulation and coding strategy. We consider two different sensor deployments. The first set up adopts low-cost sensors with PSK (BPSK or QPSK) and no channel coding. Assuming independent transmitted symbols, the arrival probability for a packet composed by $B_i = T_i W_B$ bits in a flat Rayleigh fading channel is given by [1]

$$\lambda_{i,0} = \begin{cases} \left[\frac{1}{2}\left(1 + \sqrt{\frac{\Gamma_i}{\Gamma_i+1}}\right)\right]^{B_i} & \text{BPSK} \\ \left(\sqrt{\frac{\Gamma_i}{\Gamma_i+2}}\right)^{\frac{B_i}{2}} & \text{QPSK} \end{cases} \tag{7.22}$$

Differently, the second set up considers a long-term future sensor deployment where sensors might perform adaptive rate allocation (ad-RA) choosing among a large set of modulation and coding rate modes similarly to the IEEE 802.11 wireless LAN [105]. In this case, assuming Gaussian coding the probability of packet loss might be approximated with the outage probability [1]

$$p_{\text{out},i}(B_i, T_i) = 1 - \exp \frac{-\left(2^{B_i/(W_B T_i)} - 1\right)}{\Gamma_i} \tag{7.23}$$

with $\sum_{i=1}^{2} T_i = T$. From (7.23) we have

$$\lambda_{i,0} = 1 - p_{\text{out},i}(B_i, T_i), \quad i = 1, 2\,. \tag{7.24}$$

We underline that $\lambda_{i,0}$ provides an upper bound for the arrival probability in case of long packets [1]. Moreover this bound becomes tight when the transmitter has the possibility of choosing among a large set of modulation and coding rate modes and the channel code is almost capacity achieving.

7.5 Examples of applications

The framework introduced in Sections 7.3 and 7.4 can give useful directions on the design of NCSs. In particular as examples of applications, we investigate issues within two network setups: i) cross-layer optimization of quantization processes and resource allocation with $N = 1$, and ii) comparison between single-hop and multi-hop communication protocols in case the system state is observed by a single sensor.

7.5.1 Cross-Layer optimization of quantization processes and resource allocation with $N = 1$

As in Section 7.4 we consider a simple scalar, stable system with $A = 0.9$, $R_1 = R_2 = 10^{-3}$ and set the length of the receive buffer $N = 1$. Moreover we assume a simple *single-hop* communication protocol with *no* packet *retransmission* (SH-nR) where each sensor measurement is transmitted directly to the estimator, with no retransmission in case of packet loss. Under this setting we propose a cross-layer optimization of the communication parameters $\{B_i\}$ and $\{T_i\}$ for the minimization of the error covariance matrix V_0^1. We recall that for BPSK or QPSK, B_i depends linearly on T_i as $B_i = W_B T_i$. Therefore we can consider the following constraint optimization problems:

$$\text{BPSK/QPSK} \qquad \min_{T_i \geq 0,\, \sum_{i=1}^{2} T_i \leq T} V_0^1 \tag{7.25}$$

$$\text{ad-RA} \qquad \min_{B_i, T_i \geq 0,\, \sum_{i=1}^{2} T_i = T} V_0^1 \tag{7.26}$$

Although the communication parameters $\{B_i\}$ and $\{T_i\}$ are optimized minimizing V_0^1, to better *represent* the results we introduce a more effective cost function: $\Delta_1 = V_0^1/\sigma_x^2$ where $\sigma_x^2 = \frac{Q}{1-A^2}$ is the variance of x_k.

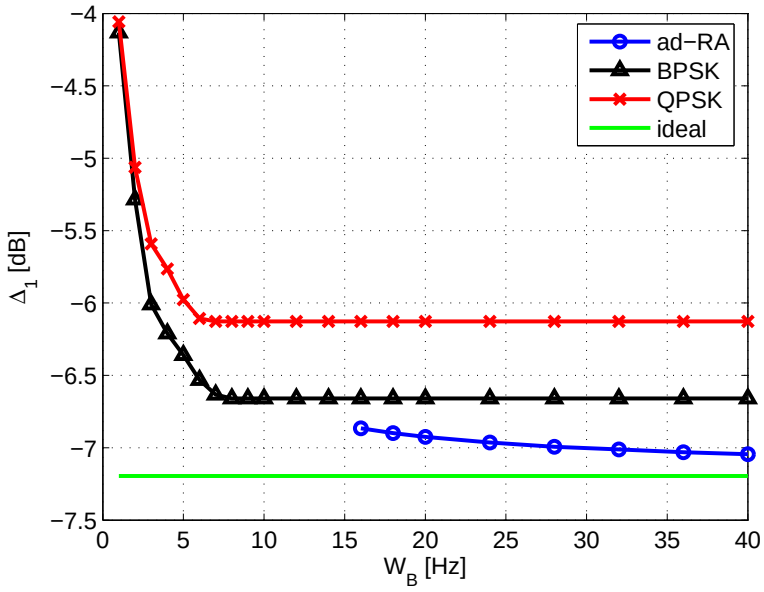

Figure 7.3: Δ_1 *as a function of the available bandwidth* W_B. $T = 1$ *s,* $A = 0.9$, $Q = 0.1$, $R_1 = R_2 = 10^{-3}$, $\Gamma_1 = 5$ *dB,* $\Gamma_2 = 0$ *dB.*

In Figs. 7.3 and 7.4 we assume $T = 1$ s, $Q = 0.1$, $\Gamma_1 = 5$ dB and $\Gamma_2 = 0$ dB and for both BPSK/QPSK and ad-RA we represent, respectively, Δ_1 and T_i as a function of W_B. In Fig. 7.3 as term of comparison we also include the *ideal scenario* with no packet loss and no signal quantization. We can see that Δ_1 decreases with W_B because additional bandwidth can only be beneficial, nevertheless good performance is achievable even with a sufficiently small bandwidth, e.g. $10 \leq W_B \leq 40$, and there is no significant gain in increasing the bandwidth further, since we are already very close to *ideal scenario* performance.

We observe how ad-RA can improve system performance especially for large W_B and approaches the lower bound given by the *ideal scenario* for $W_B \to \infty$. Furthermore BPSK performs very close to ad-RA and is preferable to QPSK. This means that it's preferable to have an higher arrival probability as in BPSK than more quantization bits and QPSK. For the same reason we notice in Fig. 7.4 that for BPSK and QPSK the total portion of T allocated for transmissions, i.e. $\sum_{i=1}^{2} T_i$, decreases with W_B because above a certain threshold additional quantization bits only cause a degradation of the packet arrival probability without improving the estimation process, as $\sigma_n^2(B_i)$ becomes negligible with respect to R_i. Differently, in case of ad-RA we always have $\sum_{i=1}^{2} T_i = 1$ because the additional bandwidth can be used to increase $\lambda_{i,0}$ with a proper choice of the modulation and coding rate mode.

Interestingly, for ad-RA the entire time slot is always allocated to the sensor with highest SNR, i.e. $T_1 = 1$ s. Differently, for BPSK/QPSK this is optimum only for small transmission bandwidth and $T_1 > T_2 > 0$ for larger W_B.

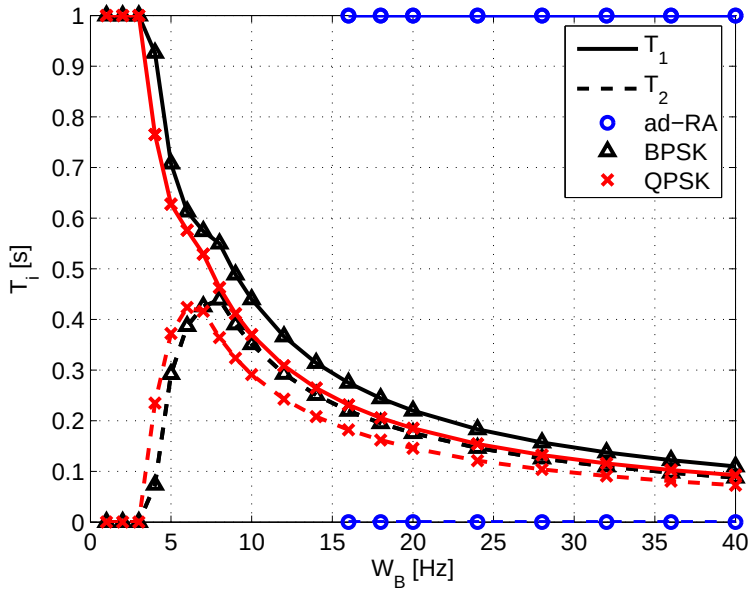

Figure 7.4: T_i as a function of the available bandwidth W_B. $T = 1$ s, $A = 0.9$, $Q = 0.1$, $R_1 = R_2 = 10^{-3}$, $\Gamma_1 = 5$ dB, $\Gamma_2 = 0$ dB.

7.5.2 Single-hop vs multi-hop communication protocols for single sensor measurements

We assume a scalar, stable system observed by a single sensor with $C_1 = 1$. Using BPSK we consider three different transmission protocols: 1) SH-nR introduced in Section 7.5.1, 2) *single-hop* communication with packet *retransmission* (SH-R), 3) *multi-hop* communication with packet *retransmissions* (MH-R). As for SH-nR, in SH-R each measurement is transmitted directly to the estimator (there is only $N_h = 1$ hop), but differently from SN-nR, SH-R employs packet retransmission in case of packet loss. Moreover we assume that SH-R uses orthogonal resources (either in frequency or time domain) for retransmitted packets so that arrival probabilities for the different measurements are independent. Based on these assumptions SH-R requires a larger transmission bandwidth than SH-nR also due to packet acknowledgements from the estimator. Finally in MH-R we assume $N_h > 1$ hops and packet retransmission. Again the different packets follow independent paths (either in frequency, time or space domain). Considering only packet loss and Rayleigh fading the SNR in each hop is $\Gamma_{1,N_h} = N_h^2 \Gamma_1$ because we assume each hop covers the same distance. Under these assumptions the arrival probabilities for SH-R and MH-R are [106]

$$\lambda_{1,h} = (\lambda_{1,0})^{N_h} \sum_{j=0}^{h-N_h} \binom{N_h + j - 1}{j} (1 - \lambda_{1,0})^j \tag{7.27}$$

where $\lambda_{1,0} = \left[\frac{1}{2} \left(1 + \sqrt{\frac{\Gamma_{1,N_h}}{\Gamma_{1,N_h}+1}} \right) \right]^{B_i}$.

In Fig. 7.5 we set $T = 1$ s, $A = 0.9$, $R_1 = 10^{-3}$, $\Gamma_1 = -5$ dB and compare SH-nR, SH-R

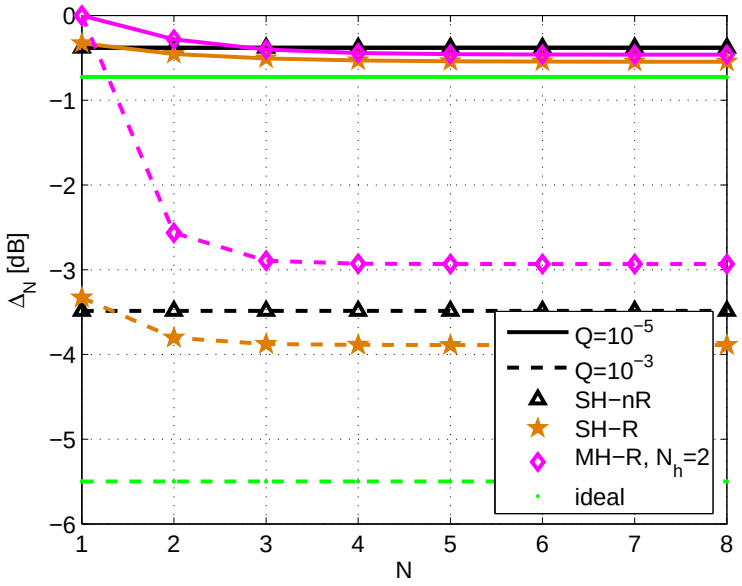

Figure 7.5: Δ_N as a function of the buffer length N. $T = 1$ s, $A = 0.9$, $R_1 = 10^{-3}$, $\Gamma_1 = -5$ dB.

and MH-R in terms of $\Delta_N = V_0^N/\sigma_x^2$ as a function of the receive buffer length N, for different values of Q. We notice that for each strategy we consider the number of quantization bits B_1 that minimizes Δ_N for the maximum buffer length $N_{\max} = 8$. As predicted by Theorem 4 the performance of both SH-R and MH-R improves with N and interestingly SH-R is the best strategy for N large enough. SH-nR performs very close to SH-R, especially for increasing Q, because x_k becomes fast time variant and old measurements (coming from retransmitted packets) do not provide significant information for the estimation process. Differently, as Q decreases, x_k becomes slowly time variant and even old observations are useful to improve the estimation process as in SH-R and MH-R. We notice that even if a smaller Q leads to a reduction of V_0^N, the cost Δ_N and Q relate in a different way. Generally MH-R besides increasing the complexity of network design and packet routing, does not provide gain over SH communication protocols because it increases the packet delay, which is a major drawback in applications dealing with state estimation. Numerical simulations reveal that MH-R might be useful only for small Q and very low Γ_1. Similar considerations hold even using ad-RA.

7.6 Conclusions

We derive the minimum error covariance estimator and the optimum estimator with constant gains for a NCS where observations from multiple sensors are subject to random delays and packet losses. The effects of measurements quantization are investigated for a stable system. For a scalar stable system, simple BPSK and single-hop communication protocols provide close to optimum estimation performance. This supports the use of low-cost sensors in environment monitoring applications.

Chapter 8

Cross-layer design of networked control systems

In Chapter 7 we dealt with state estimation in networked control systems (NCSs) accounting for wireless link inefficiencies. In this chapter we consider the more general problem of state control.

Most research activities in NCSs either focus on the problem of packet loss assuming no quantization [107, 101, 108] or study measurements and control signal quantization but assuming perfect packet delivery [109, 19, 110]. Considering a general unstable system and i.i.d Bernoulli packet drops, [108] derives conditions on the arrival probabilities of measurements and control packets that guarantee closed loop stability of a discrete time system. The more specific problem of optimum state estimation is investigated in [101] for measurements coming from multiple sensors and in [104] for Markovian packet loss. Differently, under the ideal assumption of no packet loss, [109, 19] derive the minimum transmission rate that guarantees closed-loop system stability of a discrete time unstable system.

In real sensor networks packet losses and signal quantization/encoding are not separate problems but intimately correlated. In this work we study the problem of system controllability in NCSs with multiple sensors, accounting as in [100], for both packet drops and signal quantization at sensors and controller. Assuming a TCP-like protocol [111] between controller and actuator, we solve the problem of optimum Linear Quadratic Gaussian (LQG) control [102] around a target state for a stable system and provides a generalization for unstable systems in case of a large transmission bandwidth. We show how the separation principle [102] of classical control theory does not hold in general because the optimum estimator depends on the quantizer used at controller. Moreover we characterize the limiting behavior of both estimator and controller in the infinite horizon and derive bounds for the mean square distance of the state from the target.

Since sensors, controller and actuator share the same transmission medium the available bandwidth needs to be allocated to the different wireless links, a problem that has received less attention in literature. We propose a cross-layer optimization of quantization processes and resource allocation in order to minimize a final performance metric, e.g. the mean square distance of the state from the target. As an example of application the optimization is proposed

for a simple scalar system using the framework derived for state estimation and control in the infinite horizon, nevertheless the technique is general and can be extended even to MIMO systems. As case studies we consider two different sensor deployments: i) contemporary low-cost sensors with PSK (BPSK or QPSK) modulation and no channel coding, ii) long-term future sensor deployment capable of adaptive rate allocation. We show that even with a small bandwidth transmission and simple BPSK modulation we can reach performance close to the optimum state control achievable with no signal quantization and no packet loss. This supports the widespread use of low cost sensors for these applications.

The chapter is organized as follows. Section 8.1 introduces the problem of state control for a stable system, Section 8.2 derives the optimum state estimator in case of packet loss and quantized signals at both sensors and controller and Section 8.3 solves the problem of optimum control around a target state in case of a TCP-like protocol between controller and actuator. Then in Section 8.4 we characterize the properties of both estimator and controller in the infinite horizon and give generalizations for unstable systems. Sections 8.5 and 8.6 propose a cross layer optimization of i) quantization processes and ii) resource allocation for a scalar system and numerical simulations are provided in Section 8.7. Finally, Section 7.6 concludes the chapter summarizing the main findings and proposing future research activities.

The material in this chapter has been in part published in [112] and [113].

8.1 Problem formulation

We generalize the system model introduced in Chapter 7 and consider a general multi-input multi-output (MIMO) discrete time linear system with control packets and partition the observation vector into two parts, modelling the observation of the state by two different sensors. The system has the following dynamics

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k^a + \mathbf{w}_k \tag{8.1}$$

$$\begin{bmatrix} \mathbf{s}_{1,k} \\ \mathbf{s}_{2,k} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} \mathbf{v}_{1,k} \\ \mathbf{v}_{2,k} \end{bmatrix} \tag{8.2}$$

where $\mathbf{x}_k, \mathbf{w}_k \in \mathbb{R}^n$ are the state and noise of the system at instant k, respectively, $\mathbf{u}_k^a \in \mathbb{R}^d$ is the control signal applied by actuator and $\mathbf{s}_{1,k}, \mathbf{v}_{1,k} \in \mathbb{R}^{m_1}$, $\mathbf{s}_{2,k}, \mathbf{v}_{2,k} \in \mathbb{R}^{m_2}$, are the measurements (or sensor observations) and measurement noises at time k for sensor 1 and 2, respectively. Moreover $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times d}$, $\mathbf{C}_1 \in \mathbb{R}^{m_1\times n}$, $\mathbf{C}_2 \in \mathbb{R}^{m_2\times n}$ and $(\mathbf{x}_0, \mathbf{w}_k, \mathbf{v}_{1,k}, \mathbf{v}_{2,k})$ are uncorrelated Gaussian random vectors with mean $(\bar{\mathbf{x}}_0, \mathbf{0}, \mathbf{0}, \mathbf{0})$ and covariance matrix $(\mathbf{P}_0, \mathbf{Q}, \mathbf{R}_1, \mathbf{R}_2)$, respectively. Furthermore we define $\mathbf{C} = \left[\mathbf{C}_1^T \mathbf{C}_2^T\right]^T$ and the covariance matrix of the total measurement noise $\mathbf{v}_k = \left[\mathbf{v}_{1,k}^T \mathbf{v}_{2,k}^T\right]^T$ as $\mathbf{R} = \text{diag}\,(\mathbf{R}_1, \mathbf{R}_2)$.

Measurements $\mathbf{s}_{1,k}$, $\mathbf{s}_{2,k}$ are quantized and encoded separately by two sensors and transmitted through a digital communication network (DCN) whose goal is to deliver packets from multiple sources (sensors) to a destination (estimator/controller) with possible packet drops. For simplicity each link is modelled as a single hop transmission with no packet retransmissions. Indeed both retransmissions and multi-hop DCN would possibly introduce random de-

lays in packet delivery which complicates the state estimation process [18, 114]. Moreover, packet retransmission is not guaranteed to provide performance gain in NCSs with stringent delay constraints as shown in Chapter 7 and this motivates our simplified approach. Similarly, the control signal computed by the estimator/controller is quantized, encoded and sent to the actuator through a wireless link. However over this link we assume a TCP-like protocol, so that successful or unsuccessful packet delivery at receiver (actuator) is acknowledged to the sender (estimator/controller) within the same sampling time period [111][1]. As in Chapter 7, we assume that estimator (actuator) perfectly knows whether packets received from sensors (controller) contain errors or not. Accounting for both the effects of signal encoding and packet loss, the received signals at estimator and actuator are modelled respectively as (see also Fig. 8.1)

$$\mathbf{y}_{i,k} = \gamma_{i,k}(\mathbf{s}_{i,k} + \mathbf{z}_{i,k}) = \gamma_{i,k}(\mathbf{C}_i\mathbf{x}_k + \underbrace{\mathbf{v}_{i,k} + \mathbf{z}_{i,k}}_{\bar{\mathbf{v}}_{i,k}}) \tag{8.3}$$

$$\mathbf{u}_k^a = \nu_k \underbrace{(\mathbf{u}_k^c + \mathbf{n}_k)}_{\bar{\mathbf{u}}_k^c} + (1-\nu_k)\mathbf{u}_\infty \tag{8.4}$$

where $\gamma_{1,k}$, $\gamma_{2,k}$ and ν_k are i.i.d. Bernoulli random variables which model the wireless links between sensors-controller and controller-actuator, with arrival probabilities $\bar{\gamma}_1$, $\bar{\gamma}_2$ and $\bar{\nu}$, respectively. Signal u_k^c is the control signal computed by the controller, $\mathbf{u}_\infty$ is a constant control signal applied by the actuator in case of packet loss and $\mathbf{z}_{1,k}$, $\mathbf{z}_{2,k}$, $\mathbf{n}_k$ are quantization errors for the sensor observations and the control signal, respectively. They are modelled as zero-mean, independent noises with covariance matrices $\mathbf{R}_{Z_1}$, $\mathbf{R}_{Z_2}$ and $\mathbf{R}_N$ whose values depend on the quantization strategy and the number of quantization bits. We underline that both $\mathbf{n}_k$ and $\mathbf{z}_{i,k}$ are modelled as independent of the source signals, which becomes a reasonable approximation for most practical quantization strategies as the number of quantization bits increases. In (8.4) for ease of notation we define $\bar{\mathbf{u}}_k^c = \mathbf{u}_k^c + \mathbf{n}_k$ as the quantized control input and $\bar{\mathbf{v}}_{i,k} = \mathbf{v}_{i,k} + \mathbf{z}_{i,k}$, $i = 1, 2$, as the equivalent observation noise after measurement quantization, assumed to be Gaussian with zero mean and covariance $\bar{\mathbf{R}}_i = \mathbf{R}_i + \mathbf{R}_{Z_i}$. Moreover we denote $\mathbf{y}_k = [\mathbf{y}_{1,k}^T, \mathbf{y}_{2,k}^T]^T$.

As a consequence of possible packet drops, the estimator/controller might have a complete ($\gamma_{i,k} = 1$, $i = 1, 2$), partial ($\gamma_{1,k} \neq \gamma_{2,k}$) or no ($\gamma_{i,k} = 0$ $i = 1, 2$) observation of the system state. Differently, if the control packet is correctly received ($\nu_k = 1$) the actuator applies the quantized control law $\mathbf{u}_k^a = \bar{\mathbf{u}}_k^c$, while if the packet is lost ($\nu_k = 0$) a constant control signal $\mathbf{u}_k^a = \mathbf{u}_\infty$ is applied.

We consider a stable system, i.e. $|\lambda_i(\mathbf{A})| < 1$, $i = 1, \ldots, \rho(\mathbf{A})$, with $\{\lambda_i(\mathbf{A})\}$ and $\rho(\mathbf{A})$ denoting the eigenvalues and the rank of matrix $\mathbf{A}$, respectively. Possible generalizations of the proposed framework for unstable systems are investigated in Section 8.4.1.

We notice that in (8.4), in case of packet loss, the actuator applies the constant control signal $\mathbf{u}_\infty$ whose value can be optimized according to the specific performance metric. This

[1] A TCP-like protocol between sensors and estimator is not considered because we focus on energy-efficient, low-cost sensors that simply transmit an observation of the system state, therefore packet acknowledgement would not be useful. Differently, packet acknowledgement would be helpful for more expensive sensors, with higher computational resources, that transmit state estimates rather than simple raw observations [18].

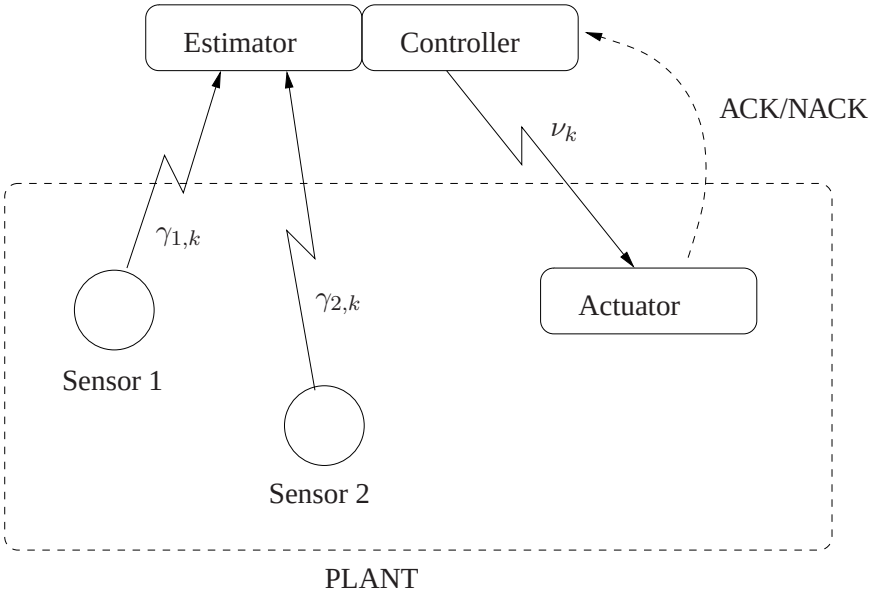

Figure 8.1: System setting.

is a more general approach with respect to the control law proposed in [108] that keeps the actuator idle ($\mathbf{u}_\infty = 0$) in case of packet drop. We underline that (8.4) is only an option, indeed the optimum strategy in case of packet drop is not known in general.

Let us define $\overline{\mathbf{u}}^{N-1} = [\mathbf{u}^c_{N-1}, \ldots, \mathbf{u}^c_1]$, $\bar{x}$ as the target system state and $\mathcal{G} = \{\mathbf{x}_0, \mathbf{P}_0, \bar{\mathbf{x}}, \mathbf{u}_\infty\}$. We consider the following cost function[2]

$$J_N(\overline{\mathbf{u}}^{N-1}, \mathcal{G}) = \mathrm{E}\Big[(\mathbf{x}_N - \bar{\mathbf{x}})^T \mathbf{W}_N (\mathbf{x}_N - \bar{\mathbf{x}}) + \sum_{k=0}^{N-1} \Big((\mathbf{x}_N - \bar{\mathbf{x}})^T \mathbf{W}_k (\mathbf{x}_N - \bar{\mathbf{x}}) + (\mathbf{u}^a_k)^T \mathbf{U}_k \mathbf{u}^a_k\Big) \Big| \overline{\mathbf{u}}^{N-1}, \mathcal{G}\Big] . \tag{8.5}$$

where $\mathbf{W}_k$ and $\mathbf{U}_k$ are positive semi-definite matrices. Notice that we weight the difference between the current system state and the target value, i.e. $(\mathbf{x}_k - \bar{\mathbf{x}})$, and the control input only when it is applied at actuator, i.e. $\mathbf{u}_\infty \neq 0$.

Let us define the information set $\mathcal{I}_k = \left\{\overline{\mathbf{y}}^k, \boldsymbol{\gamma}^k_1, \boldsymbol{\gamma}^k_2, \boldsymbol{\nu}^{k-1}\right\}$ where $\overline{\mathbf{y}}^k = [\mathbf{y}_k, \ldots, \mathbf{y}_1]$, $\boldsymbol{\gamma}^k_i = [\gamma_{i,k}, \ldots, \gamma_{i,1}]$ and $\boldsymbol{\nu}^{k-1} = [\nu_{k-1}, \ldots, \nu_1]$. We search for the control input sequence $\overline{\mathbf{u}}^{*N-1}$ as a function of the information set $\mathcal{I}_k$, the target state $\bar{\mathbf{x}}$ and the control signal $\mathbf{u}_\infty$, i.e. $\mathbf{u}^c_k = g_k(\mathcal{I}_k, \bar{\mathbf{x}}, \mathbf{u}_\infty)$, that minimizes the functional defined in (8.5), namely

$$J^*_N(\mathcal{G}) \triangleq \min_{\mathbf{u}^c_k = g_k(\mathcal{I}_k, \bar{\mathbf{x}}, \mathbf{u}_\infty),\, k=0\ldots,N-1} J_N(\overline{\mathbf{u}}^{N-1}, \mathcal{G}) \tag{8.6}$$

[2] A quadratic cost is reasonable and largely adopted in literature [102] because it induces a high penalty for large deviations of the state from the target but a relatively small penalty for small deviations. Moreover it leads to an analytical solution.

8.2 Estimator design under TCP like protocols

In this section we derive the equations for optimal state estimation when measurements come from two spatially distributed sensors and are subject to quantization and possible packet drops. In the derivation we use similar arguments that apply in standard Kalman filtering [102]. First we define the followings:

$$\begin{aligned}\hat{\mathbf{x}}_{k|k} &\triangleq \mathrm{E}[\mathbf{x}_k|\mathcal{I}_k] && (8.7)\\ \mathbf{e}_{k|k} &\triangleq \mathbf{x}_k - \hat{\mathbf{x}}_{k|k} && (8.8)\\ \mathbf{P}_{k|k} &\triangleq \mathrm{E}[\mathbf{e}_{k|k}\mathbf{e}_{k|k}^T|\mathcal{I}_k] && (8.9)\end{aligned}$$

For the optimal estimator the prediction step is given by:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1|k} &= \mathrm{E}[\mathbf{x}_{k+1}|\nu_k,\mathcal{I}_k] = \mathbf{A}\hat{\mathbf{x}}_{k|k} + \nu_k\mathbf{B}\mathbf{u}_k^c + (1-\nu_k)\mathbf{B}\mathbf{u}_\infty && (8.10)\\ \mathbf{e}_{k+1|k} &= \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} = \mathbf{A}\mathbf{e}_{k|k} + \nu_k\mathbf{B}\mathbf{n}_k + \mathbf{w}_k && (8.11)\\ \mathbf{P}_{k+1|k} &= \mathrm{E}[\mathbf{e}_{k+1|k}\mathbf{e}_{k+1|k}^T|\nu_k,\mathcal{I}_k] = \mathbf{A}\mathbf{P}_{k|k}\mathbf{A}^T + \nu_k\mathbf{B}\mathbf{R}_N\mathbf{B}^T + \mathbf{Q} && (8.12)\end{aligned}$$

where we used the independence between $\mathbf{w}_k$, $\mathbf{n}_k$ and $\mathcal{I}_k$ and the requirement that $\mathbf{u}_k$ is a deterministic function of $\mathcal{I}_k$. Exploiting the independence between $\mathbf{y}_{i,k+1}$, $\gamma_{i,k+1}$, $\mathbf{w}_k$, $\mathbf{n}_k$ and $\mathcal{I}_k$ and applying the theory of time variant Kalman filtering [18, Theorem 1],[101], the update step is given by [101]:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \gamma_{1,k+1}\gamma_{2,k+1}\mathbf{K}_{k+1}\left(\mathbf{y}_{k+1} - \mathbf{C}\hat{\mathbf{x}}_{k+1|k}\right) + \\ &\quad \gamma_{1,k+1}(1-\gamma_{2,k+1})\mathbf{K}_{1,k+1}\left(\mathbf{y}_{1,k+1} - \mathbf{C}_1\hat{\mathbf{x}}_{k+1|k}\right) + \\ &\quad (1-\gamma_{1,k+1})\gamma_{2,k+1}\mathbf{K}_{2,k+1}\left(\mathbf{y}_{2,k+1} - \mathbf{C}_2\hat{\mathbf{x}}_{k+1|k}\right) && (8.13)\\ \mathbf{P}_{k+1|k+1} &= \mathbf{P}_{k+1|k} - \gamma_{1,k+1}\gamma_{2,k+1}\mathbf{K}_{k+1}\mathbf{C}\mathbf{P}_{k+1|k} + \\ &\quad -\gamma_{1,k+1}(1-\gamma_{2,k+1})\mathbf{K}_{1,k+1}\mathbf{C}_1\mathbf{P}_{k+1|k} + \\ &\quad -(1-\gamma_{1,k+1})\gamma_{2,k+1}\mathbf{K}_{2,k+1}\mathbf{C}_2\mathbf{P}_{k+1|k} && (8.14)\end{aligned}$$

where

$$\begin{aligned}\mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k}\mathbf{C}^T\left(\mathbf{C}\mathbf{P}_{k+1|k}\mathbf{C}^T + \bar{\mathbf{R}}\right)^{-1} && (8.15)\\ \mathbf{K}_{i,k+1} &= \mathbf{P}_{k+1|k}\mathbf{C}_i^T\left(\mathbf{C}_i\mathbf{P}_{k+1|k}\mathbf{C}_i^T + \bar{\mathbf{R}}_i\right)^{-1}, \quad i=1,2 && (8.16)\end{aligned}$$

are the Kalman gains. We notice two main differences with respect to standard Kalman filtering [102]. Firstly, as in Chapter 7, the optimal estimator described by (8.10)-(8.16) jumps between different estimation strategies according to the values assumed by $\gamma_{i,k}$. In fact we have 1) an open loop estimate when $\gamma_{1,k} = \gamma_{2,k} = 0$, 2) a closed loop estimate when $\gamma_{1,k} = \gamma_{2,k} = 1$, for which state estimation evolution is the same as the classical Kalman filter, 3) estimation evolution as if $y_{1,k}$ were the only observation for $\gamma_{1,k} = 1$, $\gamma_{2,k} = 0$, 4) estimation evolution as if $y_{2,k}$ were the only observation for $\gamma_{1,k} = 0$, $\gamma_{2,k} = 1$. Secondly, both the optimal Kalman gains (8.15)-(8.16) and the error covariance matrices (8.12), (8.14) are strongly time-variant

and stochastic as they depend on the arrival sequences $\{\gamma_{i,k}\}$ and $\{\nu_k\}$.

8.3 Optimal control under TCP-like protocols

Evaluation of the optimal control policy and the corresponding value for the objective function will be derived following the dynamic programming approach based on the cost-to-go iterative procedure [102], which decomposes the minimization problem (8.6) into a sequence of much simpler minimizations.

Define the optimal value function $V_k(\mathbf{x}_k)$ as follows:

$$V_N(\mathbf{x}_N) \triangleq \mathrm{E}[(\mathbf{x}_N-\bar{\mathbf{x}})^T\mathbf{W}_N(\mathbf{x}_N-\bar{\mathbf{x}})|\mathcal{I}_N,\bar{\mathbf{x}},\mathbf{u}_\infty] \tag{8.17a}$$

$$\begin{aligned} V_k(\mathbf{x}_k) \triangleq \min_{\mathbf{u}_k^c} \mathrm{E}\big[&(\mathbf{x}_k-\bar{\mathbf{x}})^T\mathbf{W}_k(\mathbf{x}_k-\bar{\mathbf{x}})+(\mathbf{u}_k^a)^T\mathbf{U}_k\mathbf{u}_k^a \\ &+V_{k+1}(\mathbf{x}_{k+1})|\mathcal{I}_k,\bar{\mathbf{x}},\mathbf{u}_\infty\big] \end{aligned} \tag{8.17b}$$

where $k=N-1,\ldots,1$. Using dynamic programming theory [102], one can show that

$$J_N^*(\mathcal{G}) = V_0(\mathbf{x}_0)\,. \tag{8.18}$$

Under TCP-like protocols lemma 3 holds, whose proof can be found in Appendix B.3.

Lemma 3 *The value function $V_k(x_k)$ defined in (8.17) for the system dynamics (8.1)-(8.3) under TCP-like protocols can be written for $k=N,\ldots,0$ as:*

$$V_k(\mathbf{x}_k) = E\big[\mathbf{x}_k^T\mathbf{S}_k\mathbf{x}_k - 2\bar{\mathbf{x}}^T\mathbf{T}_k\mathbf{x}_k + 2(1-\bar{\nu})u_\infty^T\mathbf{B}^T\mathbf{Z}_k\mathbf{A}\mathbf{x}_k|\mathcal{I}_k,\bar{\mathbf{x}},\mathbf{u}_\infty\big]+c_k\,, \tag{8.19}$$

where the matrices $\mathbf{S}_k$, $\mathbf{T}_k$, $\mathbf{Z}_k$ and the scalar c_k can be computed recursively as follows:

$$\mathbf{S}_k = \mathbf{A}^T\mathbf{S}_{k+1}\mathbf{A}+\mathbf{W}_k-\bar{\nu}\mathbf{A}^T\mathbf{S}_{k+1}\mathbf{B}^T(\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{B}+\mathbf{U}_k)^{-1}\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{A} \tag{8.20}$$

$$\mathbf{T}_k = \mathbf{W}_k+\mathbf{T}_{k+1}\left(\mathbf{I}-\bar{\nu}\mathbf{B}(\mathbf{U}_k+\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{S}_{k+1}\right)\mathbf{A} \tag{8.21}$$

$$\mathbf{Z}_k = \mathbf{S}_{k+1}+\mathbf{Z}_{k+1}\mathbf{A}\left(\mathbf{I}-\bar{\nu}\mathbf{B}(\mathbf{U}_k+\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{S}_{k+1}\right) \tag{8.22}$$

$$\begin{aligned} c_k = {}& \mathrm{trace}\left(\left(\mathbf{A}^T\mathbf{S}_{k+1}\mathbf{A}+\mathbf{W}_k-\mathbf{S}_k\right)\mathbf{P}_{k|k}\right)+\mathrm{trace}(\mathbf{S}_{k+1}\mathbf{Q})+ \\ &\bar{\nu}\,\mathrm{trace}((\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{B}+\mathbf{U}_k)\mathbf{R}_N)+\bar{\mathbf{x}}^T\mathbf{M}_{1,k}\bar{\mathbf{x}}+(1-\bar{\nu})\mathbf{u}_\infty^T\mathbf{M}_{2,k}\mathbf{u}_\infty+ \\ &2(1-\bar{\nu})\mathbf{u}_\infty^T\mathbf{B}^T\mathbf{M}_{3,k}\mathbf{T}_{k+1}^T\bar{\mathbf{x}}+E[c_{k+1}|\mathcal{I}_k,\bar{\mathbf{x}},\mathbf{u}_\infty] \end{aligned} \tag{8.23}$$

where

$$\mathbf{M}_{1,k} = \mathbf{W}_k-\bar{\nu}\mathbf{T}_{k+1}\mathbf{B}(\mathbf{U}_k+\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{T}_{k+1}^T \tag{8.24}$$

$$\begin{aligned} \mathbf{M}_{2,k} = {}& \mathbf{U}_k+\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{B}+(1-\bar{\nu})\mathbf{B}^T\left(\mathbf{Z}_{k+1}\mathbf{A}+\mathbf{A}^T\mathbf{Z}_{k+1}^T\right)\mathbf{B}+ \\ &-\bar{\nu}(1-\bar{\nu})\mathbf{B}^T\mathbf{Z}_{k+1}\mathbf{A}\mathbf{B}(\mathbf{U}_k+\mathbf{B}^T\mathbf{S}_{k+1}B)^{-1}\mathbf{B}^T\mathbf{A}^T\mathbf{Z}_{k+1}^T\mathbf{B} \end{aligned}$$

$$\mathbf{M}_{3,k} = \bar{\nu}\mathbf{Z}_{k+1}\mathbf{A}\mathbf{B}(\mathbf{U}_k+\mathbf{B}^T\mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T-\mathbf{I} \tag{8.25}$$

and the initial values are $\mathbf{S}_N = \mathbf{W}_N$, $\mathbf{T}_N = \mathbf{W}_N$, $\mathbf{Z}_N = \mathbf{0}$ *and* $c_N = \bar{\mathbf{x}}^T \mathbf{W}_N \bar{\mathbf{x}}$. *Moreover the optimal control input is given by:*

$$\begin{aligned} \mathbf{u}_k^c &= -(\mathbf{B}^T \mathbf{S}_{k+1} \mathbf{B} + \mathbf{U}_k)^{-1} \mathbf{B}^T (\mathbf{S}_{k+1} \mathbf{A} \hat{\mathbf{x}}_{k|k} - \mathbf{T}_{k+1}^T \bar{\mathbf{x}} + (1-\bar{\nu}) \mathbf{A}^T \mathbf{Z}_{k+1}^T \mathbf{B} \mathbf{u}_\infty) \\ &\triangleq \mathbf{L}_k \hat{\mathbf{x}}_{k|k} - \mathbf{F}_k \bar{\mathbf{x}} + \mathbf{G}_k \mathbf{u}_\infty \end{aligned} \tag{8.26}$$

□

From (8.18) and Lemma 3 it follows that the cost function for the optimal LQG controller using TCP-like protocols is given by:

$$J_N^* = J_N^{(1)} \left(\left\{ \mathrm{E}_{\gamma_i,\nu}[\mathbf{P}_{k|k}] \right\}_{k=0}^{N-1} \right) + J_N^{(2)} \tag{8.27}$$

with

$$J_N^{(1)} \left(\left\{ \mathrm{E}_{\gamma_i,\nu}[\mathbf{P}_{k|k}] \right\}_{k=0}^{N-1} \right) = \sum_{k=0}^{N-1} \mathrm{trace} \left(\left(\mathbf{A}^T \mathbf{S}_{k+1} \mathbf{A} + \mathbf{W}_k - \mathbf{S}_k \right) E_{\gamma_i,\nu}[\mathbf{P}_{k|k}] \right) \tag{8.28}$$

$$\begin{aligned} J_N^{(2)} &= \bar{\mathbf{x}}_0^T \mathbf{S}_0 \bar{\mathbf{x}}_0 + \mathrm{tr}(\mathbf{S}_0 \mathbf{P}_0) - 2\bar{\mathbf{x}}^T \mathbf{T}_0 \bar{\mathbf{x}}_0 + \bar{\mathbf{x}}^T \mathbf{W}_N \bar{\mathbf{x}} + 2(1-\bar{\nu}) \mathbf{u}_\infty^T \mathbf{B}^T \mathbf{Z}_0 \mathbf{A} \bar{\mathbf{x}}_0 + \\ &\sum_{k=0}^{N-1} \left[\mathrm{tr}(\mathbf{S}_{k+1} \mathbf{Q}) + \bar{\nu}\, \mathrm{trace}((\mathbf{B}^T \mathbf{S}_{k+1} \mathbf{B} + \mathbf{U}_k) \mathbf{R}_N) \right] + \sum_{k=0}^{N-1} \bar{\mathbf{x}}^T \mathbf{M}_{1,k} \bar{\mathbf{x}} + \\ &\sum_{k=0}^{N-1} (1-\bar{\nu}) \mathbf{u}_\infty^T \mathbf{M}_{2,k} \mathbf{u}_\infty + \sum_{k=0}^{N-1} 2(1-\bar{\nu}) \mathbf{u}_\infty^T \mathbf{B}^T \mathbf{M}_{3,k} \mathbf{T}_{k+1}^T \bar{\mathbf{x}} \end{aligned} \tag{8.29}$$

where $\mathrm{E}_{\gamma_i,\nu}[\cdot]$ explicitly indicates that the expectation is calculated with respect to the arrival sequences $\{\gamma_{i,k}\}$ and $\{\nu_k\}$. We notice that the cost function depends on the target state $\bar{\mathbf{x}}$ and on the steady control law $\mathbf{u}_\infty$ that can be optimized for the minimization of (8.27). We could either derive the optimal $\mathbf{u}_\infty$ for any specific time step N. Here $\mathbf{u}_\infty$ is derived in the infinite horizon for $N \to \infty$ (see Lemma 5).

With respect to standard LQG control theory we have two main observations:

1. The separation principle between estimation and control does not strictly holds because optimal Kalman filtering depends on the adopted control strategy through the covariance quantization noise $\mathbf{R}_N$, as it can be inferred from (8.12). Even if under different assumptions, this observation recall the certainty equivalence and quasi separation theorem derived in [19].

2. Certainty equivalence [102] holds and the optimal control law is a linear function of the estimated state $\hat{\mathbf{x}}_{k|k}$, i.e. $\mathbf{u}_k^c = \mathbf{L}_k \hat{\mathbf{x}}_{k|k} - \mathbf{F}_k \bar{\mathbf{x}} + \mathbf{G}_k \mathbf{u}_\infty$. Moreover the control gains $\mathbf{L}_k$, $\mathbf{F}_k$ and $\mathbf{G}_k$ are independent of the arrival process sequences $\{\gamma_{i,k}\}$, $i = 1, 2$.

Using similar arguments used in [107, 101] it can be inferred that the exact value of the expected error covariance matrix $E_{\gamma_i,\nu}[\mathbf{P}_{k|k}]$ cannot be computed analytically. Nevertheless it can be bounded by computable deterministic quantities as shown in the following lemma whose proof can be recovered along the lines of [101].

Lemma 4 *The expected error covariance matrix $E_{\gamma_i,\nu}[\mathbf{P}_{k|k}]$ satisfies the following bounds:*

$$\underline{\mathbf{P}}_{k|k} \leq E_{\gamma_i,\nu}[\mathbf{P}_{k|k}] \leq \overline{\mathbf{P}}_{k|k}, \quad \forall k \geq 0 \tag{8.30}$$

where matrices $\underline{\mathbf{P}}_{k|k}$, $\overline{\mathbf{P}}_{k|k}$ can be computed as follows:

$$\begin{aligned}
\overline{\mathbf{P}}_{k+1|k} &= \mathbf{A}\overline{\mathbf{P}}_{k|k-1}\mathbf{A}^T + \mathbf{Q} + \bar{\nu}\mathbf{B}\mathbf{R}_N\mathbf{B}^T + \\
&\quad -\bar{\gamma}_1\bar{\gamma}_2\mathbf{A}\overline{\mathbf{P}}_{k|k-1}\mathbf{C}^T(\mathbf{C}\overline{\mathbf{P}}_{k|k-1}\mathbf{C}^T + \bar{\mathbf{R}})^{-1}\mathbf{C}\overline{\mathbf{P}}_{k|k-1}\mathbf{A}^T + \\
&\quad -\bar{\gamma}_1(1-\bar{\gamma}_2)\mathbf{A}\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_1^T(\mathbf{C}_1\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_1^T + \bar{\mathbf{R}}_1)^{-1}\mathbf{C}_1\overline{\mathbf{P}}_{k|k-1}\mathbf{A}^T + \\
&\quad -(1-\bar{\gamma}_1)\bar{\gamma}_2\mathbf{A}\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_2^T(\mathbf{C}_2\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_2^T + \bar{\mathbf{R}}_2)^{-1}\mathbf{C}_2\overline{\mathbf{P}}_{k|k-1}\mathbf{A}^T
\end{aligned} \tag{8.31}$$

$$\begin{aligned}
\overline{\mathbf{P}}_{k|k} &= \overline{\mathbf{P}}_{k|k-1} - \bar{\gamma}_1\bar{\gamma}_2\overline{\mathbf{P}}_{k|k-1}\mathbf{C}^T(\mathbf{C}\overline{\mathbf{P}}_{k|k-1}\mathbf{C}^T + \bar{\mathbf{R}})^{-1}\mathbf{C}\overline{\mathbf{P}}_{k|k-1} + \\
&\quad -\bar{\gamma}_1(1-\bar{\gamma}_2)\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_1^T(\mathbf{C}_1\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_1^T + \bar{\mathbf{R}}_1)^{-1}\mathbf{C}_1\overline{\mathbf{P}}_{k|k-1} + \\
&\quad -(1-\bar{\gamma}_1)\bar{\gamma}_2\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_2^T(\mathbf{C}_2\overline{\mathbf{P}}_{k|k-1}\mathbf{C}_2^T + \bar{\mathbf{R}}_2)^{-1}\mathbf{C}_2\overline{\mathbf{P}}_{k|k-1}
\end{aligned} \tag{8.32}$$

$$\underline{\mathbf{P}}_{k+1|k} = (1-\bar{\gamma}_1)(1-\bar{\gamma}_2)\mathbf{A}\underline{\mathbf{P}}_{k|k-1}\mathbf{A}^T + \mathbf{Q} + \bar{\nu}\mathbf{B}\mathbf{R}_N\mathbf{B}^T \tag{8.33}$$

$$\underline{\mathbf{P}}_{k|k} = (1-\bar{\gamma}_1)(1-\bar{\gamma}_2)\underline{\mathbf{P}}_{k|k-1} \tag{8.34}$$

with initial conditions $\overline{\mathbf{P}}_{0|0} = \underline{\mathbf{P}}_{0|0} = \mathbf{P}_0$. □

From Lemma 4 it follows that also the minimum achievable cost J_N^* cannot be computed analytically, however it can be bounded as follows:

$$J_N^{min} \leq J_N^* \leq J_N^{max} \tag{8.35}$$

$$J_N^{min} = J_N^{(1)}\left(\left\{\underline{\mathbf{P}}_{k|k}\right\}_{k=0}^{N-1}\right) + J_N^{(2)} \tag{8.36}$$

$$J_N^{max} = J_N^{(1)}\left(\left\{\overline{\mathbf{P}}_{k|k}\right\}_{k=0}^{N-1}\right) + J_N^{(2)} \tag{8.37}$$

8.4 Infinite horizon LQG control

The infinite horizon LQG control can be obtained as limit for $N \to \infty$ in all the previous equations. Nevertheless, as matrices $\{\mathbf{P}_{k|k}\}$ depend nonlinearly on the specific arrival sequences $\{\gamma_{i,k}\}$ and $\{\nu_k\}$, both the expected error covariance matrices $\mathrm{E}_{\gamma_i,\nu}[\mathbf{P}_{k|k}]$ and the minimum cost J_N^* cannot be computed analytically and both do not seam to have limit [108]. However we can derive bounds for the cost function and limit behaviours for the optimal control gains.

Let us set for simplicity $\mathbf{W}_k = \mathbf{W}$ and $\mathbf{U}_k = \mathbf{U}$. Moreover let us introduce the following modified algebraic equations

$$\begin{aligned}
g_{\bar{\gamma}_1,\bar{\gamma}_2,\bar{\nu}}(\mathbf{P}) &= \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{Q} + \bar{\nu}\mathbf{B}\mathbf{R}_N\mathbf{B}^T - \bar{\gamma}_1\bar{\gamma}_2\mathbf{A}\mathbf{P}\mathbf{C}^T(\mathbf{C}\mathbf{P}\mathbf{C}^T + \bar{\mathbf{R}})^{-1}\mathbf{C}\mathbf{P}\mathbf{A}^T + \\
&\quad -\bar{\gamma}_1(1-\bar{\gamma}_2)\mathbf{A}\mathbf{P}\mathbf{C}_1^T(\mathbf{C}_1\mathbf{P}\mathbf{C}_1^T + \bar{\mathbf{R}}_1)^{-1}\mathbf{C}_1\mathbf{P}\mathbf{A}^T + \\
&\quad -(1-\bar{\gamma}_1)\bar{\gamma}_2\mathbf{A}\mathbf{P}\mathbf{C}_2^T(\mathbf{C}_2\mathbf{P}\mathbf{C}_2^T + \bar{\mathbf{R}}_2)^{-1}\mathbf{C}_2\mathbf{P}\mathbf{A}^T
\end{aligned} \tag{8.38}$$

$$h_{\bar{\nu}}(\mathbf{S}) = \mathbf{A}^T\mathbf{S}\mathbf{A} + \mathbf{W} - \bar{\nu}\mathbf{A}^T\mathbf{S}\mathbf{B}(\mathbf{B}^T\mathbf{S}\mathbf{B} + \mathbf{U})^{-1}\mathbf{B}^T\mathbf{S}\mathbf{A} \tag{8.39}$$

for the estimation and control problems respectively. The main results in the infinite horizon for the proposed LQG control problem are summarized in the following Lemma whose proof is given in Appendix B.4.

Lemma 5 *Consider the system (8.1)-(8.3) and the optimal estimator under TCP-like protocols given in (8.10) and (8.12)-(8.14). Let* $(\mathbf{A}, \mathbf{B})$, $(\mathbf{A}, \mathbf{Q}^{1/2})$ *be controllable,* $(\mathbf{A}, \mathbf{C})$, $(\mathbf{A}, \mathbf{W}^{1/2})$ *be observable, and* $\mathbf{A}$ *be stable. Moreover assume* $\mathbf{W}_k = \mathbf{W}$ *and* $\mathbf{U}_k = \mathbf{U}$. *Then the following statements hold:*

(a) *The infinite horizon optimal controller gain is constant:*

$$\lim_{k\to\infty} \mathbf{L}_k = \mathbf{L}_\infty = -(\mathbf{B}^T\mathbf{S}_\infty\mathbf{B} + \mathbf{U})^{-1}\mathbf{B}^T\mathbf{S}_\infty\mathbf{A} \tag{8.40}$$

$$\lim_{k\to\infty} \mathbf{F}_k = \mathbf{F}_\infty = -(\mathbf{B}^T\mathbf{S}_\infty\mathbf{B} + \mathbf{U})^{-1}\mathbf{B}^T\mathbf{T}_\infty \tag{8.41}$$

$$\lim_{k\to\infty} \mathbf{G}_k = \mathbf{G}_\infty = -(1-\bar{\nu})(\mathbf{B}^T\mathbf{S}_\infty\mathbf{B} + \mathbf{U})^{-1}\mathbf{B}^T\mathbf{A}^T\mathbf{Z}_\infty^T\mathbf{B} \tag{8.42}$$

(b) *The infinite horizon optimal estimator gains* $\mathbf{K}_k$ *and* $\mathbf{K}_{i,k}$ *are stochastic and time-varying since they depend on the arrival sequences* $\{\gamma_{i,k}\}$ *and* $\{\nu_k\}$.

(c) *The infinite horizon optimal constant control in case of packet loss is*

$$\mathbf{u}_\infty = -\mathbf{M}_{2,\infty}^{-1}\mathbf{B}^T\mathbf{M}_{3,\infty}\mathbf{T}_\infty^T\bar{\mathbf{x}} \triangleq \mathbf{\Lambda}_\infty\bar{\mathbf{x}} \tag{8.43}$$

Let us note that it is a linear function of the target state $\bar{\mathbf{x}}$, *hence* $\mathbf{u}_\infty = 0$ *is optimum only when* $\bar{\mathbf{x}} = 0$.

(d) *Using the steady control law (8.43), the minimum cost can be bounded by two deterministic sequences:*

$$\frac{1}{N}J_N^{min} \leq \frac{1}{N}J_N^* \leq \frac{1}{N}J_N^{max} \tag{8.44}$$

where $\frac{1}{N}J_N^{min}$, $\frac{1}{N}J_N^{max}$ *converge to the following quantities:*

$$j_\infty^{min} \triangleq \lim_{N\to\infty}\frac{1}{N}J_N^{min} = \underline{j}_\infty^{(1)} + j_\infty^{(2)} \tag{8.45}$$

$$j_\infty^{max} \triangleq \lim_{N\to\infty}\frac{1}{N}J_N^{max} = \bar{j}_\infty^{(1)} + j_\infty^{(2)} \tag{8.46}$$

with

$$\begin{aligned}\underline{j}_\infty^{(1)} &\triangleq \lim_{N\to\infty}\frac{1}{N}J_N^{(1)}\left(\left\{\underline{\mathbf{P}}_{k|k}\right\}_{k=0}^{N-1}\right)\\ &= (1-\bar{\gamma}_1)(1-\bar{\gamma}_2)\text{trace}\left(\left(\mathbf{A}^T\mathbf{S}_\infty\mathbf{A} + \mathbf{W} - \mathbf{S}_\infty\right)(\underline{\mathbf{P}}_\infty)\right)\end{aligned} \tag{8.47}$$

$$\begin{aligned}
\bar{j}_{\infty}^{(1)} &\triangleq \lim_{N\to\infty}\frac{1}{N}J_N^{(1)}\left(\{\overline{\mathbf{P}}_{k|k}\}_{k=0}^{N-1}\right)\\
&= \text{trace}\Big(\left(\mathbf{A}^T\mathbf{S}_\infty\mathbf{A}+\mathbf{W}-\mathbf{S}_\infty\right)(\overline{\mathbf{P}}_\infty+\\
&\quad -\bar{\gamma}_1\bar{\gamma}_2\overline{\mathbf{P}}_\infty\mathbf{C}^T(\mathbf{C}\overline{\mathbf{P}}_\infty\mathbf{C}^T+\bar{\mathbf{R}})^{-1}\mathbf{C}\overline{\mathbf{P}}_\infty+\\
&\quad -\bar{\gamma}_1(1-\bar{\gamma}_2)\overline{\mathbf{P}}_\infty\mathbf{C}_1^T(\mathbf{C}_1\overline{\mathbf{P}}_\infty\mathbf{C}_1^T+\bar{\mathbf{R}}_1)^{-1}\mathbf{C}_1\overline{\mathbf{P}}_\infty+\\
&\quad -(1-\bar{\gamma}_1)\bar{\gamma}_2\overline{\mathbf{P}}_\infty\mathbf{C}_2^T(\mathbf{C}_2\overline{\mathbf{P}}_\infty\mathbf{C}_2^T+\bar{\mathbf{R}}_2)^{-1}\mathbf{C}_2\overline{\mathbf{P}}_\infty)\Big) \qquad (8.48)\\
j_{\infty}^{(2)} &\triangleq \lim_{N\to\infty}\frac{1}{N}J_N^{(2)}\\
&= \bar{\nu}\,\text{trace}((\mathbf{B}^T\mathbf{S}_\infty\mathbf{B}+\mathbf{U})\mathbf{R}_N)+tr(\mathbf{S}_\infty\mathbf{Q})+\\
&\quad \bar{\mathbf{x}}^T\left(\mathbf{M}_{1,\infty}-(1-\bar{\nu})\mathbf{T}_\infty\mathbf{M}_{3,\infty}^T\mathbf{B}\mathbf{M}_{2,\infty}^{-1}\mathbf{B}^T\mathbf{M}_{3,\infty}\mathbf{T}_\infty^T\right)\bar{\mathbf{x}} \qquad (8.49)
\end{aligned}$$

and the matrices $\mathbf{S}_\infty$, $\mathbf{T}_\infty$, $\mathbf{Z}_\infty$ $\underline{\mathbf{P}}_\infty$, $\overline{\mathbf{P}}_\infty$ *are solutions of the following equations*

$$\begin{aligned}
\mathbf{S}_\infty &= h_{\bar{\nu}}(\mathbf{S}_\infty) && (8.50)\\
\mathbf{T}_\infty &= \mathbf{W}+\mathbf{T}_\infty\left(\mathbf{A}-\bar{\nu}\mathbf{B}(\mathbf{U}+\mathbf{B}^T\mathbf{S}_\infty\mathbf{B})^{-1}\mathbf{B}^T\mathbf{S}_\infty\mathbf{A}\right) && (8.51)\\
\mathbf{Z}_\infty &= \mathbf{S}_\infty+\mathbf{Z}_\infty\mathbf{A}\left(\mathbf{I}-\bar{\nu}\mathbf{B}(\mathbf{U}+\mathbf{B}^T\mathbf{S}_\infty\mathbf{B})^{-1}\mathbf{B}^T\mathbf{S}_\infty\right) && (8.52)\\
\underline{\mathbf{P}}_\infty &= (1-\bar{\gamma}_1)(1-\bar{\gamma}_2)\mathbf{A}\underline{\mathbf{P}}_\infty\mathbf{A}^T+\mathbf{Q}+\bar{\nu}\mathbf{B}\mathbf{R}_N\mathbf{B}^T && (8.53)\\
\overline{\mathbf{P}}_\infty &= g_{\bar{\gamma}_1,\bar{\gamma}_2,\bar{\nu}}(\overline{\mathbf{P}}_\infty) && (8.54)
\end{aligned}$$

□

It can be shown by numerical simulations that the lower bound (8.45) is quite loose. Differently, the upper bound (8.46) is tight and would represent the effective infinite horizon cost function if we were to use optimal constant Kalman gains in the estimation process (as proposed in Section 7.3), instead of the time varying Kalman filters derived in (8.15)-(8.16). This consideration follows as a simple generalization of the analysis in Section 7.3 when considering $N = 1$ as the receiving buffer length, no packet retransmission and accounting for control signal quantization.

8.4.1 Generalizations for unstable systems in case of negligible quantization error

In the previous sections the problems of state estimation and control are explicitly investigated for a stable system. With this assumption the quantization process at both sensors and controller can be performed with simple time-invariant quantizers without compromising mean square stability. Moreover, in case of small distortion (many quantization bits), the quantization error can be modelled as an additive noise independent of the source signal. Unfortunately this considerations do not always extend to more general unstable systems. Indeed, even assuming perfect packet delivery, if the system is strictly unstable and either $\bar{\mathbf{x}}_0$ or $\mathbf{w}_k$ has infinite support, neither time-invariant memoryless encoders (e.g uniform quantizer) or finite-state predictive quantizers [94], can guarantee a bounded cost function [19]. In fact, in case of perfect packet delivery, only adaptive quantizers with possibly unbounded range can maintain mean square stability for sufficiently high data rates, e.g. using notation of Section 8.5 and 8.6 it must

be $\frac{b_i}{T_i} > \log_2 |\mathbf{A}|$, [109, 19]. Therefore, the framework introduced in the previous sections, accounting for packet loss and limited transmission bandwidth, does not apply to unstable systems for which the problem of signals quantization and error quantization modelling are still unsolved and active areas of research.

Nevertheless, if we can neglect quantization errors, for example when the transmission bandwidth is large enough, i.e. $\mathbf{R}_N = \mathbf{R}_{Z_1} = \mathbf{R}_{Z_2} \simeq \mathbf{0}$, the control laws (8.26) and (8.43) are still optimum for state control around a target state even in case of unstable systems, providing a generalization of the analysis given in [108]. We recall that assuming only packet drops and no quantization errors, an unstable system can be stabilized in the infinite horizon only if $\bar{\gamma}_i$, $\bar{\nu}$ are higher than given thresholds, i.e. $\bar{\gamma}_i \geq \bar{\gamma}_i^{th}$ and $\bar{\nu} \geq \bar{\nu}^{th}$. These thresholds define a stability region for the system whose boundaries depend on system parameters [108, 101]. Interestingly, the convergence of (8.40)-(8.43), (8.51)-(8.52) and (8.45)-(8.46) in the infinite horizon is guaranteed inside the same stability region.

8.5 Optimization of quantization processes in the infinite horizon

In this section we propose an optimization of quantizers at sensors and controller adopting as performance metric the infinite horizon cost metric derived in Section 8.3. For simplicity we refer to a simple scalar system but the technique can be generalized to more complex MIMO systems.

Let us consider a stable, scalar system with $A < 1$, $C_1 = C_2 = 1$, $B = 1$ and $R_1, R_2 > 0$, i.e. the system is observed by two sensors with possible different features. We consider simple uniform quantization and denote with b_i, $i = 1, 2$, and b_3 the number of bits used for signal quantization at sensor i and controller, respectively. In case of small distortion, the quantization errors $z_{i,k}(b_i)$ and $n_k(b_3)$ can be modelled as white additive noises with uniform distribution in $[-\Delta_i/2, \Delta_i/2]$ where the quantization step size Δ_i depends on the source signal and the number of quantization bits [94][3]. If we approximate the system state as Gaussian, $z_{i,k}(b_i)$ and $n_k(b_3)$ have variance $R_{Z_i}(b_i) = 3\sigma_{s_i}^2 2^{-2b_i}$ and $R_N(b_3) = 3\sigma_u^2 2^{-2b_3}$ [94], where $\sigma_{s_i}^2$ and σ_u^2 denote the variance of observation $s_{i,k}$ and control signal u_k^c, respectively. Both $R_{Z_i}(b_i)$ and $R_N(b_3)$ depend on the number of quantization bits and the variance of the source signals $s_{i,k}$ and u_k^c. These signals are interconnected through the state evolution equations (7.1)-(8.4), therefore the optimization of the quantization bits at sensors and controller should be carried out jointly for an effective minimization of the performance metric (8.5) in the infinite horizon.

Taking the expectation of (8.3) and (8.4) and computing the limit for $k \to \infty$ we get the

[3]Under uniform quantization the equivalent observation noise $\bar{v}_{i,k} = v_{i,k} + z_{i,k}$ and process noise $\nu_k Bn_k + w_k$ (see also (8.11)) are not strictly Gaussian and Kalman filtering is not optimum for state estimation [102]. Nevertheless, let us define with $f_{\bar{v}_{i,k}^g}(x)$ the probability density function (PDF) of the Gaussian random variable $\bar{v}_{i,k}^g$ having zero mean and variance $\bar{R}_i = R_i + R_{Z_i}$ and denote with $D_1 = \int |f_{\bar{v}_{i,k}^g}(x) - f_{\bar{v}_{i,k}}(x)| \mathrm{d}x$ the norm one distance between the PDFs of $\bar{v}_{i,k}^g$ and $\bar{v}_{i,k}$. It can be seen that for $R_i \geq R_{Z_i}$ the Gaussian approximation is quite tight as $D_1 \leq 0.05$ for the value of R_i used in Section 8.7. In the cross-layer optimization of Section 8.6 the quantization noise is either negligible with respect to the observation noise ($R_i \gg R_{Z_i}$) or at most comparable ($R_i \approx R_{Z_i}$), therefore the Gaussian approximation is generally tight and Kalman filtering is nearly optimum. The same considerations apply to the process noise as well.

following equations

$$\begin{aligned} m_{s_i} &= \lim_{k\to\infty} \mathrm{E}[s_{i,k}] = C_i \lim_{k\to\infty} \mathrm{E}[x_k] = C_i m_x = m_x && (8.55)\\ m_{u^c} &= \lim_{k\to\infty} \mathrm{E}_{\gamma_i,\nu}[u_k^c] = L_\infty \lim_{k\to\infty} \mathrm{E}_{\gamma_i,\nu}[\mathrm{E}[x_k|\mathcal{I}_k, \bar{x}, u_\infty]] - (F_\infty - G_\infty\Lambda_\infty)\bar{x} \\ &= L_\infty m_x - (F_\infty - G_\infty\Lambda_\infty)\bar{x} && (8.56)\end{aligned}$$

where $L_\infty = -\frac{AS_\infty}{S_\infty+U}$, $F_\infty = -\frac{T_\infty}{S_\infty+U}$, $G_\infty = -(1-\bar{\nu})\frac{Z_\infty A}{S_\infty+U}$ while Λ_∞ is defined in (8.43). From (8.55)-(8.56) and (8.1) we can derive the following system

$$\begin{cases} m_x = Am_x + \bar{\nu}m_{u^c} + (1-\bar{\nu})\Lambda_\infty\bar{x} \\ m_{u^c} = L_\infty m_x - (F_\infty - G_\infty\Lambda_\infty)\bar{x} \end{cases} \tag{8.57}$$

whose solution is given by

$$\begin{aligned} m_x &= \frac{[\bar{\nu}(G_\infty\Lambda_\infty - F_\infty) + (1-\bar{\nu})\Lambda_\infty]\,\bar{x}}{1 - A - \bar{\nu}L_\infty} \triangleq \alpha\bar{x} && (8.58)\\ m_{u^c} &= (L_\infty\alpha - F_\infty + G_\infty\Lambda_\infty)\,\bar{x} && (8.59)\end{aligned}$$

It's interesting to note that m_x depends linearly on the target state $\bar{x}$. Using (8.58) and (8.59) we can get the variance of signals $s_{i,k}$ and u_k^c in the infinite horizon as

$$\begin{aligned} \sigma_{s_i}^2 &= \lim_{k\to\infty} \mathrm{E}[(s_{i,k} - m_{i,s})^2] = \lim_{k\to\infty} \mathrm{E}[((x_k - \bar{x}) + (1-\alpha)\bar{x} + v_{i,k})^2] \\ &= \lim_{k\to\infty} \mathrm{E}[(x_k - \bar{x})^2] + R_i - (1-\alpha)^2\bar{x}^2 && (8.60)\\ \sigma_{u^c}^2 &= \lim_{k\to\infty} \mathrm{E}[(u_k^c - m_{u^c})^2] = \lim_{k\to\infty} \mathrm{E}_{\gamma_i,\nu}[(L_\infty\hat{x}_{k|k} - (F_\infty - G_\infty\Lambda_\infty)\bar{x} - m_{u^c})^2] \\ &= L_\infty^2 \lim_{k\to\infty} \mathrm{E}[(\hat{x}_{k|k} - \alpha\bar{x})^2] = L_\infty^2 \lim_{k\to\infty} \mathrm{E}[((x_k - \bar{x}) - e_{k|k} + (1-\alpha)\bar{x})^2] && (8.61)\end{aligned}$$

From (8.61) we can further get

$$\begin{aligned} \sigma_{u^c}^2 &= L_\infty^2 \lim_{k\to\infty} \left(\mathrm{E}[(x_k - \bar{x})^2] - (1-\alpha)^2\bar{x}^2 + \mathrm{E}[P_{k|k}] - 2\mathrm{E}[(e_{k|k} + \hat{x}_{k|k})e_{k|k}]\right) \\ &= L_\infty^2 \lim_{k\to\infty} \left(\mathrm{E}[(x_k - \bar{x})^2] - \mathrm{E}[P_{k|k}] - (1-\alpha)^2\bar{x}^2\right) && (8.62)\end{aligned}$$

where in the last line of (8.62) we used the orthogonality between estimation error and state estimate, i.e. $\mathrm{E}[\hat{x}_{k|k}e_{k|k}] = 0$.

Variance $\sigma_{s_i}^2$ and $\sigma_{u^c}^2$ can hardly be computed analytically, nevertheless if we use the upper bound $\lim_{k\to\infty}\frac{1}{N}J_N^* \simeq j_\infty^{max}$ we can derive approximations $\tilde{\sigma}_{s_i}^2$ and $\tilde{\sigma}_{u^c}^2$ noticing that in the infinite horizon we have

$$D_x \triangleq \lim_{k\to\infty} \mathrm{E}[(x_k - \bar{x})^2] \simeq \frac{j_\infty^{max} - UM_{u^a}}{W} \tag{8.63}$$

where

$$M_{u^a} \triangleq \lim_{k\to\infty} \mathrm{E}\left[(u_{k^a})^2\right] \simeq \bar{\nu}(\sigma_{u^c}^2 + m_{u^c}^2) + (1-\bar{\nu})u_\infty^2 \tag{8.64}$$

is the average power of the control signal at actuator. Moreover with this specific system setting

we have

$$\begin{aligned} j_{\infty}^{max} &= (S_{\infty}(A^2-1)+W)\tilde{P}_{\infty} + \bar{\nu}(S_{\infty}+U)R_N + S_{\infty}Q + \\ &\quad \bar{x}^2\left(M_{1,\infty} - (1-\bar{\nu})\frac{(T_{\infty}M_{3,\infty})^2}{M_{2,\infty}}\right) \end{aligned} \tag{8.65}$$

with

$$\tilde{P}_{\infty} = \overline{P}_{\infty} - \bar{\gamma}_1\bar{\gamma}_2\frac{\overline{P}_{\infty}^2}{\overline{P}_{\infty}+\bar{R}_{eq}} - \bar{\gamma}_1(1-\bar{\gamma}_2)\frac{\overline{P}_{\infty}^2}{\overline{P}_{\infty}+\bar{R}_1} - (1-\bar{\gamma}_1)\bar{\gamma}_2\frac{\overline{P}_{\infty}^2}{\overline{P}_{\infty}+\bar{R}_2} \tag{8.66}$$

where $\bar{R}_{eq} = \frac{\bar{R}_1\bar{R}_2}{\bar{R}_1+\bar{R}_2}$ and the second term of (8.66) is obtained using the matrix inversion lemma.

From (8.63) we get

$$\sigma_{s_i}^2 \simeq \tilde{\sigma}_{s_i}^2 = D_x + R_i - (1-\alpha)^2\bar{x}^2 \tag{8.67}$$

$$\sigma_{u^c}^2 \simeq \tilde{\sigma}_{u^c}^2 = L_{\infty}^2\left(D_x - \tilde{P}_{\infty} - (1-\alpha)^2\bar{x}^2\right) \tag{8.68}$$

If we use the approximations $R_{Z_i}(b_i) \simeq 3\tilde{\sigma}_{s_i}^2 2^{-2b_i}$ and $R_N(b_3) \simeq 3\tilde{\sigma}_{u^c}^2 2^{-2b_3}$ we can obtain $\tilde{\sigma}_{s_i}^2$ and $\tilde{\sigma}_{u^c}^2$ from (8.67), (8.68), (8.54) and (8.50) after solving for given b_i, $i=1,2,3$, $\bar{\gamma}_i$, $i=1,2$ and $\bar{\nu}$ the following system

$$\begin{cases} \frac{R_{Z_i}}{3\cdot 2^{-2b_i}} = D_x + R_i - (1-\alpha)^2\bar{x}^2, \quad i=1,2 \\ \frac{R_N}{3\cdot 2^{-2b_3}} = \left(\frac{AS_{\infty}}{S_{\infty}+U}\right)^2\left(D_x - \tilde{P}_{\infty} - (1-\alpha)^2\bar{x}^2\right) \\ \overline{P}_{\infty} = A^2\overline{P}_{\infty} + Q + \bar{\nu}R_N - \bar{\gamma}_1\bar{\gamma}_2\frac{A^2\overline{P}_{\infty}^2}{\overline{P}_{\infty}+\bar{R}_{eq}} - \bar{\gamma}_1(1-\bar{\gamma}_2)\frac{A^2\overline{P}_{\infty}^2}{\overline{P}_{\infty}+\bar{R}_1} - (1-\bar{\gamma}_1)\bar{\gamma}_2\frac{A^2\overline{P}_{\infty}^2}{\overline{P}_{\infty}+\bar{R}_2} \\ S_{\infty} = A^2S_{\infty} + W - \bar{\nu}\frac{A^2S_{\infty}^2}{S_{\infty}+U} \end{cases} \tag{8.69}$$

where R_{Z_1}, R_{Z_2}, R_N, $\overline{P}_{\infty}$ and S_{∞} are the unknowns. We notice that the last two equations of (8.69) result from simplifications of (8.54) and (8.50) in case of a scalar system.

We recall that, if we were to use optimal constant Kalman gains instead of time-varying Kalman filtering, with a noteworthy simplification of the estimation process, $\overline{P}_{\infty}$ and j_{∞}^{max} would represent the true infinite horizon error variance and cost function, respectively. In this case both (8.67) and (8.68) would hold as strict equalities and the solution of (8.69) would provide the exact values of the unknowns.

8.6 Cross-layer optimization of quantization processes and resource allocation in the infinite horizon

As in Section 7.4 we adopt a time division multiple access (TDMA) as medium access control (MAC) strategy. Both sensors and controller are allocated a portion T_i ($i=1,2$ for sensor i and $i=3$ for controller) of the available time slot T. We assume a total transmission bandwidth W_B and a block fading model with independent Rayleigh fading realizations for all links, i.e.

each radio link is modelled as $\sqrt{\Gamma_i}h_{i,k}$ with $h_{i,k} \sim \mathcal{CN}(0,1)$ and Γ_i, $i \in \{1,2,3\}$ denoting the average link SNR. Moreover we assume that only the average SNR is available at the two transmitters.

The probability of packet loss depends on channel conditions, the number of quantization bits, the modulation and the coding strategy. The framework introduced in Sections 8.2 and 8.3 is general and can be applied for many transmission strategies.

As examples of application we investigate the cross-layer optimization of quantization processes and resource allocation for two different sensor deployments. The first set up adopts low-cost contemporary sensors with PSK (BPSK or QPSK) modulation and no channel coding. We recall that assuming independent transmitted symbols, the arrival probability for a packet composed by $b_i = T_i W_B$ bits in a flat Rayleigh fading channel is given by (7.22). The same transmission strategy is also used for the link between controller and actuator and $\bar{\nu}$ can be modelled as in (7.22).

Differently, the second set up considers a long-term future sensor deployment where sensors might perform adaptive rate allocation (ad-RA) choosing among a large set of modulation and coding rates. In this case the packet arrival probability for sensor i can be approximated with (7.24). We use the same model for the control packet too.

We can derive a cross-layer optimization of the communication parameters $\{b_i\}$ and $\{T_i\}$ for the minimization of the infinite horizon cost metric (8.46) using the framework introduced in the previous sections. For PSK, b_i depends linearly on T_i as $b_i = W_B T_i$, and we can consider the following constraint optimization problem:

$$\min_{T_i,\, i=1,2,3} j_{\infty}^{max} \tag{8.70a}$$

$$T_i \geq 0,\ \sum_{i=1}^{3} T_i \leq T \tag{8.70b}$$

Differently, for (ad-RA) the optimization problem becomes

$$\min_{b_i, T_i,\, i=1,2,3} j_{\infty}^{max} \tag{8.71a}$$

$$b_i, T_i \geq 0,\ \sum_{i=1}^{3} T_i = T \tag{8.71b}$$

Both (8.70) and (8.71) are generally non-convex and using numerical optimization tools we can only guarantee to achieve local optimum solutions.

8.7 Simulation results

In this section we present two sets of numerical results dealing with: i) cross-layer optimization of quantization processes and resource allocation by solving (7.25)-(7.26) within the framework introduced in the previous sections and ii) representation of state evolution for a given target state trajectory assuming BPSK and a practical wireless channel model. As in Sections 8.5 and 8.6, we consider a simple scalar, stable system with $A = 0.9$ and $C_1 = C_2 = B = 1$.

Moreover $R_1 = R_2 = 10^{-3}$ to model sensors with the same technical features.

In the first set of numerical results the communication parameters b_i and T_i are optimized for the minimization of j_∞^{max} according to the constraint problems (7.25)-(7.26). In Fig. 8.2 we assume $T = 1$ s, $Q = 0.1$, $W = 1$, $\Gamma_1 = \Gamma_3 = 10$ dB, $\Gamma_2 = 5$ dB and $W_B = 20$ Hz. We adopt the second sensor network deployment (ad-RA) and represent j_∞^{max} as a function of $\bar{x}$, for different values of U. Using optimum u_∞ in (8.4) provides a further gain with respect to the suboptimum strategy that keeps the actuator idle in case of packet loss, i.e. $u_\infty = 0$. And this consideration applies even changing the weight U assigned to the power of the control signal. Moreover using u_∞ we achieve with only $W_B = 20$ Hz performance very close to the *ideal scenario* with no packet loss and no quantization.

Under the same system setting, Fig. 8.3 shows the portion of time slot allocated to the different links. Choosing the optimum u_∞ has an additional benefit over $u_\infty = 0$, because resource allocation becomes almost independent of $\bar{x}$ and the network does not need to be reconfigured when $\bar{x}$ changes. Moreover even if Γ_1 = Γ_3 to the controller is allocated more than 50% of the available time slot and the rest is distributed among the two sensors: here $T_1 > T_2$ as a consequence of the higher link reliability, i.e. $\Gamma_1 > \Gamma_2$. Differently, with $u_\infty = 0$ the optimum resource allocation varies with $\bar{x}$. In particular T_3 increases with $\bar{x}$ because increasing the reliability of the link between controller and actuator becomes more and more important for state control. Indeed in case of packet loss the actuator stays idle and the state evolves towards the point of equilibrium $x_k = 0$. This causes an increment of the distance from $\bar{x}$ and this effect becomes more pronounced for increasing $\bar{x}$. The same system setting is considered also in Fig 8.4 where the allocated quantization bits are shown as a function of $\bar{x}$. Still using optimum u_∞ renders the system more robust to variations of $\bar{x}$, moreover the number of quantization bits for both schemes generally reduces with $\bar{x}$ in order to increase link reliability (see (7.23)).

In Fig. 8.5 and 8.6 we keep the same system parameters of Figs. 8.2, 8.3 and 8.4, set $\bar{x} = 5$ and represent respectively j_∞^{max} and T_i as a function of the available bandwidth W_B. Both in case of fix rate (BPSK, QPSK) and ad-RA the cost decreases with W_B because additional bandwidth can only be beneficial, nevertheless good performance is achievable even with a sufficient small bandwidth, e.g. $10 \leq W_B \leq 30$, and there is no significant gain in increasing the bandwidth further, since we are already very close to *ideal scenario* performance. We observe how ad-RA can improve system performance especially for large W_B and approaches the lower bound given by the *ideal scenario* for $W_B \to \infty$. Furthermore BPSK becomes preferable to QPSK if W_B is large enough. Indeed there is no gain in increasing the number of quantization bits above a given threshold because the quantization noise $z_{i,k}$ becomes negligible with respect to the observation noise $v_{i,k}$, therefore for W_B large enough BPSK is preferable because it gives an higher arrival probability than QPSK. For the same reason we notice from Fig. 8.6 that for BPSK and QPSK the total portion of T allocated for transmissions, i.e. $\sum_{i=1}^{3} T_i$, decreases with W_B because above a certain threshold additional quantization bits only cause a degradation of the packet arrival probability without improving the estimation process. Differently, in case of ad-RA we always have $\sum_{i=1}^{3} T_i = 1$ because the additional bandwidth can be used to increase $\bar{\gamma}_i$ and $\bar{\nu}$ with a proper choice of modulation and coding rate. Still $T_3 \gtrapprox 0.5$ and $T_1 > T_2$ as a consequence of $\Gamma_1 > \Gamma_2$. Moreover, $T_2 = 0$ for small W_B, because the two

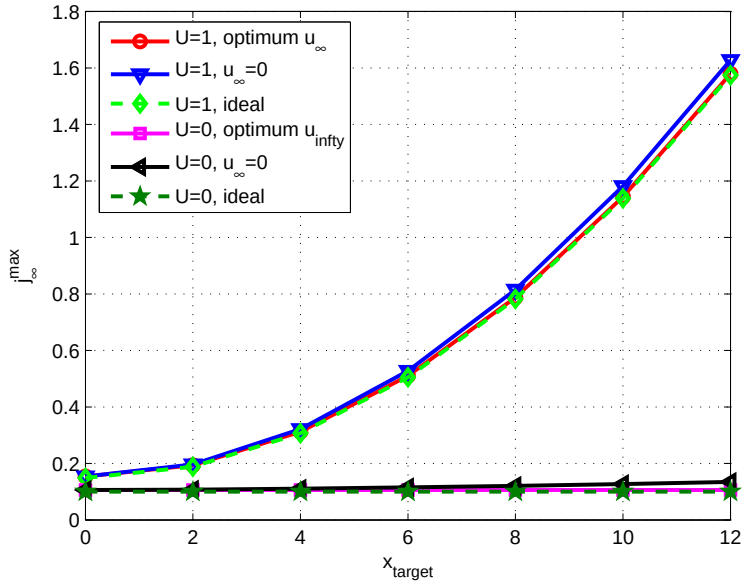

Figure 8.2: j_{∞}^{max} as a function of the target system state $\bar{x}$ for ad-RA. $T = 1$ s, $W_B = 20$ Hz, $A = 0.9$, $Q = 0.1$, $R = 0.001$, $\Gamma_1 = \Gamma_3 = 10$ dB, $\Gamma_2 = 5$ dB.

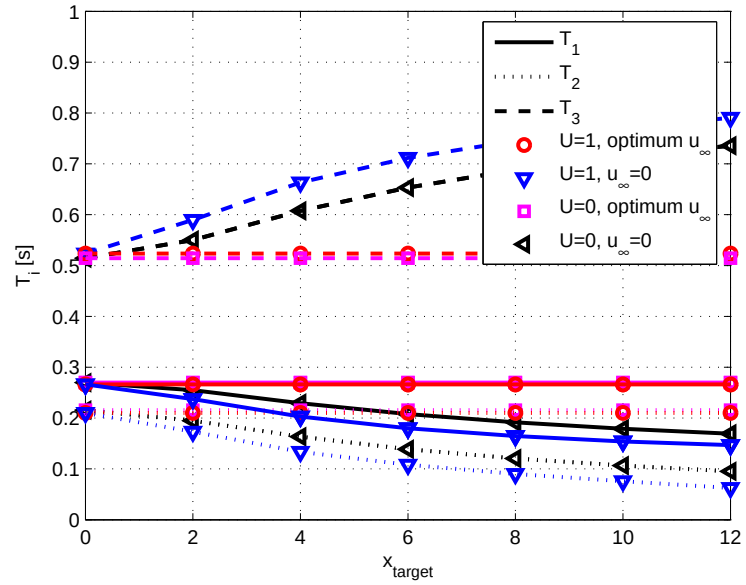

Figure 8.3: Portion of time slot assigned to sensors and controller as a function of the target system state $\bar{x}$ for ad-RA. $T = 1$ s, $W_B = 20$ Hz, $A = 0.9$, $Q = 0.1$, $R = 0.001$, $\Gamma_1 = \Gamma_3 = 10$ dB, $\Gamma_2 = 5$ dB.

sensors provide measurements of the same quality, i.e. $R_1 = R_2$, and with limited bandwidth allocating a portion of the time slot only to the sensor with the highest SNR is preferable.

Figs. 8.7 and 8.8 show respectively j_{∞}^{max} and T_i as a function of Γ_1 for two different values of Γ_2 and for both ad-RA and BPSK. The optimum u_{∞} is adopted for all transmission

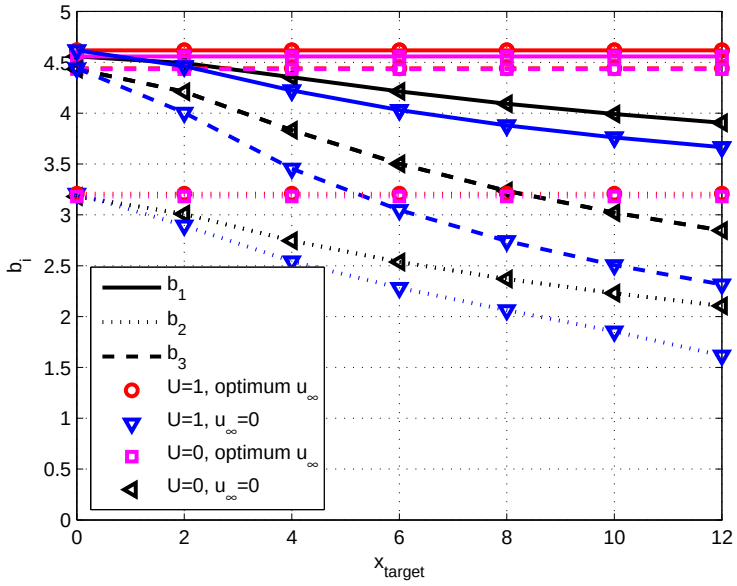

Figure 8.4: Number of quantization bits b_i for sensors and controller as a function of the target system state $\bar{x}$ for ad-RA. $T = 1$ s, $W_B = 20$ Hz, $A = 0.9$, $Q = 0.1$, $R = 0.001$, $\Gamma_1 = \Gamma_3 = 10$ dB, $\Gamma_2 = 5$ dB.

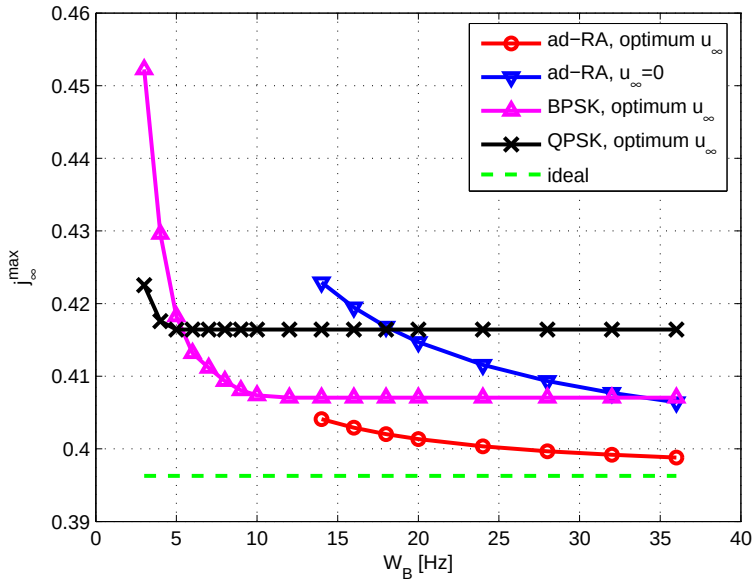

Figure 8.5: j_∞^{max} as a function of the available bandwidth W_B. $T = 1$ s, $A = 0.9$, $Q = 0.1$, $R = 0.001$, $\Gamma_1 = \Gamma_3 = 10$ dB, $\Gamma_2 = 5$ dB, $\bar{x} = 5$.

strategies. When $\Gamma_2 = 10$ dB there is marginal gain in considering an additional sensor even if $\Gamma_2 = 30$ dB, meaning that a single measurement of the state might be sufficient if the link between sensor and controller is reliable. Differently, if $\Gamma_2 = -20$ dB an additional sensor measurement provides a significant gain in terms of j_∞^{max} even with small $\Gamma_1 \simeq 0$ dB.

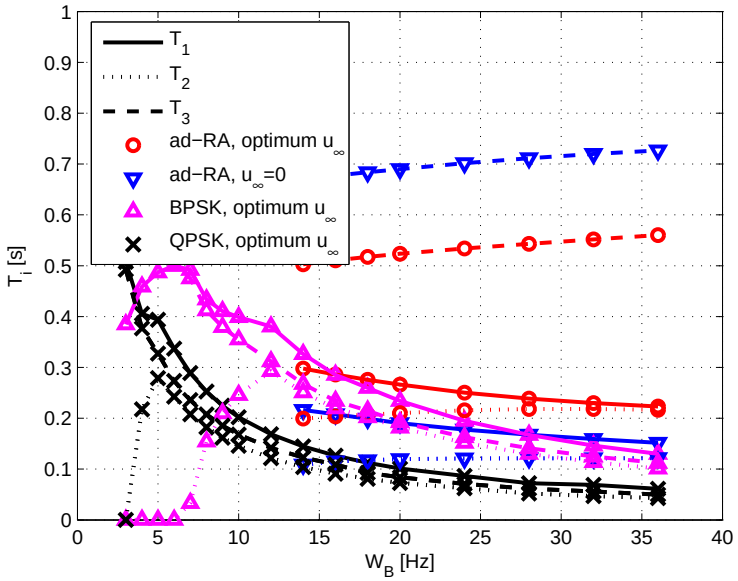

Figure 8.6: Portion of time slot assigned to sensors and controller as a function of the available bandwidth W_B. $T = 1$ s, $A = 0.9$, $Q = 0.1$, $R = 0.001$, $\Gamma_1 = \Gamma_3 = 10$ dB, $\Gamma_2 = 5$ dB, $\bar{x} = 5$.

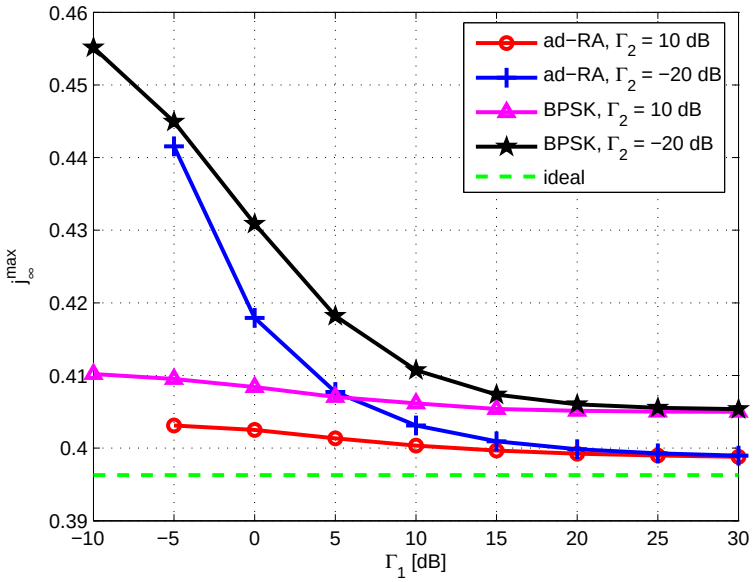

Figure 8.7: j_∞^{max} as a function of Γ_1. $T = 1$ s, $W_B = 20$ Hz, $A = 0.9$, $Q = 0.1$, $R = 0.001$, $\Gamma_2 = 10, -20$ dB, $\Gamma_3 = 10$ dB, $\bar{x} = 5$, optimum u_∞.

From Fig. 8.8 we notice that if the SNRs for the two sensors are significantly different, the available time slot is mostly distributed between the controller and the best sensor. Moreover the optimum allocation among them depends on their SNRs with less time assigned to the link with the highest SNR.

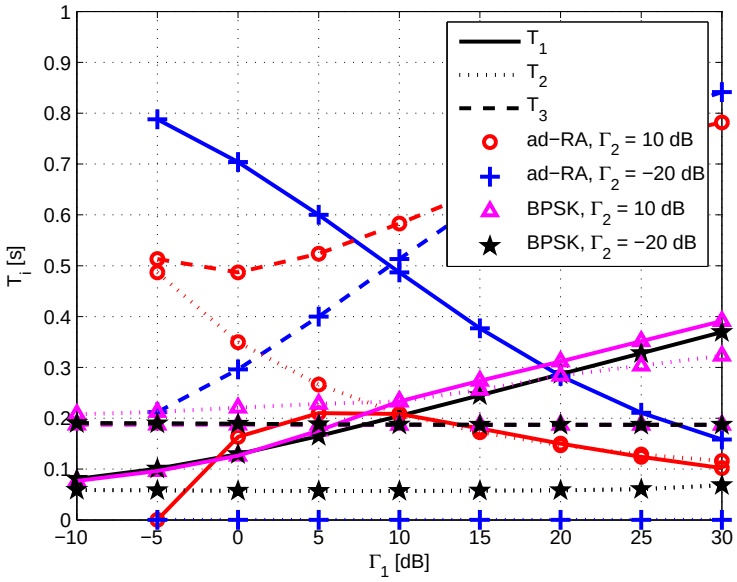

Figure 8.8: Portion of time slot assigned to sensors and controller as a function of Γ_1. $T = 1$ *s,* $W_B = 20$ *Hz,* $A = 0.9$, $Q = 0.1$, $R = 0.001$, $\Gamma_2 = 10, -20$ *dB,* $\Gamma_3 = 10$ *dB,* $\bar{x} = 5$, *optimum* u_∞.

We notice that in Figs. 8.5 and 8.7 the gap in terms of j_∞^{max} between ad-RA and BPSK is usually small, meaning that even a simple BPSK with no channel coding might be effective for state control of a scalar, stable system.

The second set of numerical results considers two sensors using BPSK with a carrier frequency of 2.4 GHz [115]. Each wireless link is modelled as a flat Rayleigh fading channel with classic Doppler spectrum having a Doppler spread of $f_s = 6.7$ Hz [116], which accounts for pedestrian mobility in the surrounding environment. We assume $T = 1$ s, $W = 1$, $U = 0$, $\Gamma_1 = \Gamma_3 = 10$ dB, $\Gamma_2 = 5$ dB, $x_0 = 0$, $W_B = 20$ Hz and perfect channel estimation at demodulators. Note that for $T = 1$ s, since the coherence time of the channel is $T_c \approx 1/f_s = 15$ ms, the assumption of independent packet arrivals in Section 8.1 is generally satisfied. In Fig. 8.9 we use the proposed NCS to control x_k around a target state trajectory. Assuming optimum u_∞, the optimization of b_i in the infinite horizon, according to (7.25) yields $b_1 = b_2 = b_3 = 4$ for $Q = 0.1$ and $b_1 = 5$, $b_2 = 3$, $b_3 = 4$ for $Q = 1$, independently of $\bar{x}$. The independence of optimum b_i from the target state recalls a similar finding for ad-RA in Fig. 8.4. The controller adopts the optimum control law (8.4) with infinite horizon controller gains (8.40)-(8.42) which do not need to be reconfigured as $\bar{x}$ changes. Nevertheless, every time $\bar{x}$ changes, the controller has to transmit u_∞ to the actuator while the estimator feeds back m_{s_i} and $\sigma_{s_i}^2$ to the two sensors for the update of the quantization parameters. We notice from (8.55) and (8.67) that $m_{s_i} = m_x$ and $\sigma_{s_i}^2 - R_i$ are independent of the sensor index, therefore a common information can be transmitted to the two sensors. We observe from Fig. 8.9 that with the proposed framework we are able to have a tight control of the system state around the state trajectory, especially for a slowly time variant system with small Q. Moreover, convergence of x_k in the presence of a switch in the state target is generally accomplished within few time slots, with

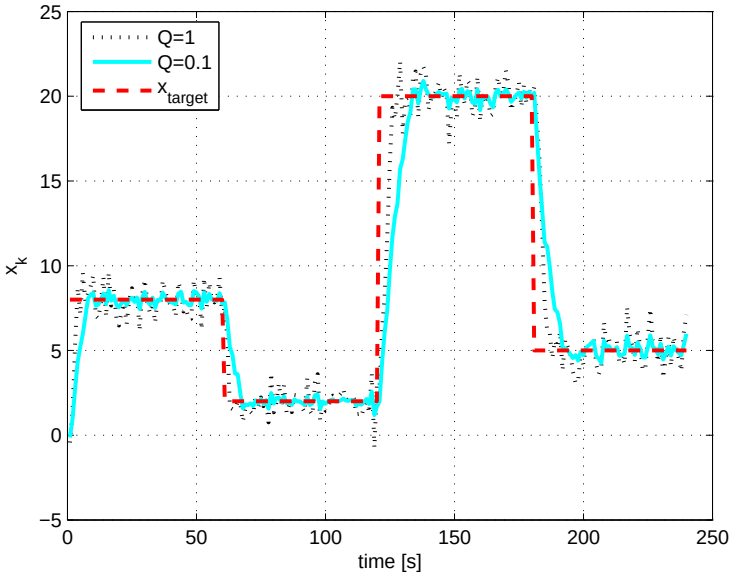

Figure 8.9: State evolution for assigned target states. $T = 1$ *s,* $W_B = 20$ *Hz,* $W = 1$*,* $U = 0$*,* $A = 0.9$*,* $R = 0.001$*,* $\Gamma_1 = \Gamma_3 = 10$ *dB,* $\Gamma_2 = 5$ *dB, BPSK, optimum* b_i *and* u_∞*.*

faster convergence for increasing Q.

8.8 Conclusions

The chapter studies the problem of system control around a target state trajectory in NCSs with multiple sensors, accounting for both packet drops and signal quantization at sensors and controller. Assuming a TCP-like protocol between controller and actuator, we solve the problem of optimum linear quadratic Gaussian (LQG) control around a target state for a stable system and provide a generalization for unstable systems in case of negligible quantization errors, i.e. large transmission bandwidth. Moreover we characterize the limiting behavior of both estimator and controller in the infinite horizon and derive bounds for the mean square distance of the state from the target. By a cross-layer approach we optimize quantization processes and resource allocation in order to minimize a final performance metric. Although the algorithm is proposed for a simple, scalar, stable system the proposed framework is general and could be extended to MIMO systems. Surprisingly even with a small bandwidth transmission and simple BPSK we can reach performance close to the optimum state control achievable with no quantization and no packet loss. This supports the widespread use of low cost sensors in applications dealing with state control around a target trajectory.

This work also attracts attention on several issues that motivate future research activities for a better understanding of NCSs design: i) correlated packet arrivals, ii) time-variant signal quantization in unstable systems, iii) generalization of the cost function to account for sensor energy efficiency, iv) vector quantization modelling in case of MIMO systems and v) generalization of estimation and control techniques adopting “soft” detection/decoding strategies.

Chapter 9

Conclusions

This thesis considers two different multiuser wireless communication systems: the MIMO broadcast channel and a networked control system. Chapters 3-6 deal with the MIMO BC with special interest in FDD systems where the transmitter receives quantized channel information from the receivers. Differently, Chapters 7 and 8 consider NCSs where multiple sensors, controller and actuator exchange low-rate messages to monitor or control a dynamical system. In the following we summarize the main contributions of the different chapters.

In Chapter 2 we assume perfect CSIT and describe a sub-optimum multiuser eigenmode transmission (MET) technique based on ZF BF where the set of active users and the set of eigenmodes per user are selected with a greedy algorithm in order to maximize the weighted throughput. MET outperforms most state-of-the-art linear precoding strategies and achieves a large fraction of DPC capacity in most channel environments.

In Chapter 3 we assume limited uplink FB from single antenna receivers and investigate: i) beamformer design, ii) channel quantization and feedback signalling optimization and iii) user selection. We design a novel MMSE beamformer under incomplete CSIT that outperforms ZF BF when users are randomly selected, but provides marginal gains when user selection follows an opportunistic approach. Moreover we propose various LBG-based FB strategies that exploit spatial and time correlation of the MIMO channel. In particular hierarchical FB and predictive FB provide the largest gains. Finally robust and efficient greedy user selection algorithms are derived for the maximization of the system sum rate.

In Chapter 4 users are equipped with multiple antennas. Under the assumption of at most one stream per selected user we design two transceiver strategies: i) ZF BF with MESC and ii) unitary BF with MMSE combiner. Both channel quantization codebook and FB strategies adapt to the transceiver technique. ZF BF with MESC and hierarchical FB is preferable especially for low mobility users, nevertheless unitary BF with MMSE combiner requires less control signalling and is very competitive for low FB rates.

In Chapter 5 we evaluate the potential gains of MU MIMO over SU MIMO in a multi-cell packet-based cellular network. Network MIMO and higher order sectorization are investigated as possible methods to mitigate inter-cell interference.

In Chapter 6 we consider a MIMO-OFDM BC where the available bandwidth is divided into RBs. We provide joint conditions on the channel coherence bandwidth and the FB rate per

RB that allow for a simpler quantization of the RB channel matrix by a space vector, causing negligible throughput loss. Moreover we show that even for a moderate number of users in the network, concentrating all the available FB bits in characterizing only one RB provides a significant gain in system throughput over a more classical distributed approach.

In Chapter 7 we derive the minimum error covariance estimator and the optimum estimator with constant gains for a NCS where observations from multiple sensors are subject to random delays and packet losses. The effects of measurements quantization are investigated for a stable system. For a scalar stable system, simple BPSK and single-hop communication protocols provide close to optimum estimation performance.

In Chapter 8, assuming a stable system and a TCP-like protocol between controller and actuator, we solve the problem of optimum state control around a target state in case of both packet drops and signal quantization. Generalization for unstable systems is also given for large bandwidth transmissions. The proposed framework is used for a cross-layer design of signal quantization processes and network resource allocation for a scalar system. Again almost optimal control is achievable with small bandwidth transmissions and simple BPSK. This supports the use of low-cost sensors in this kind of applications.

Appendix A

Proof of theorems for MU MIMO downlink systems

A.1 Proof of Theorem 1

We solve the constraint problem (3.19) using Lagrangian multipliers. Define the cost function

$$\begin{aligned} J(\mathbf{G},\beta,\lambda) &= \mathrm{E}\Big[\big|\big|\left(\left(\mathbf{F}+\mathbf{E}\right)\mathbf{G}(\mathcal{S})\mathbf{d}+\mathbf{N}^{-1}\mathbf{n}\right)\beta^{-1}-\mathbf{d}\big|\big|^2\Big]+ \\ &\quad \lambda\left(\mathrm{tr}(\mathbf{G}^H\mathbf{G})-P\right) \\ &= \beta^{-2}\mathrm{tr}(\mathbf{F}\mathbf{G}\mathbf{G}^H\mathbf{F}^H)+\beta^{-2}\mathrm{tr}(\mathrm{E}[\mathbf{E}\mathbf{G}\mathbf{G}^H\mathbf{E}^H])+ \\ &\quad -\beta^{-1}\mathrm{tr}(\mathbf{G}^H\mathbf{F}^H)-\beta^{-1}\mathrm{tr}(\mathbf{F}\mathbf{G})+M+\beta^{-2}\sigma_N^2+ \\ &\quad \lambda\left(\mathrm{tr}(\mathbf{G}^H\mathbf{G})-P\right) \end{aligned} \tag{A.1}$$

From the computation of $\frac{\partial J(\mathbf{G},\beta,\lambda)}{\partial \mathbf{G}}=0$ we get as necessary condition for the constraint minimization problem (3.19), $\mathbf{F}^H\mathbf{F}\mathbf{G}+\mathbf{R}\mathbf{G}-\beta\mathbf{F}^H+\lambda\beta^2\mathbf{G}=0$, which yields

$$\mathbf{G}=\beta\left(\mathbf{F}^H\mathbf{H}+\mathbf{R}+\lambda\beta^2\mathbf{I}\right)^{-1}\mathbf{F}^H\,. \tag{A.2}$$

Introducing $\eta=\lambda\beta^2$ and

$$\bar{\mathbf{G}}(\eta)=\left[\mathbf{F}^H\mathbf{F}+\mathbf{R}+\eta\mathbf{I}\right]^{-1}\mathbf{F}^H \tag{A.3}$$

(A.2) can be equivalently written as

$$\mathbf{G}(\eta)=\beta(\eta)\bar{\mathbf{G}}(\eta) \tag{A.4}$$

where

$$\beta(\eta)=\sqrt{\frac{P}{\mathrm{tr}\left(\bar{\mathbf{G}}(\eta)^H\bar{\mathbf{G}}(\eta)\right)}} \tag{A.5}$$

imposes the average power constraint (3.19b).

For the solution of (3.19) we have to impose $\frac{\partial J(\mathbf{G},\beta,\lambda)}{\partial \beta} = 0$ which yields

$$\text{tr}\left(\mathbf{F}\mathbf{G}\mathbf{G}^H\mathbf{F}^H + \text{E}[\mathbf{E}\mathbf{G}\mathbf{G}^H\mathbf{E}^H] - \beta\Re(\mathbf{F}\mathbf{G}) + \frac{\sigma_N^2}{|\mathcal{S}|}\mathbf{I}_{|\mathcal{S}|}\right) = 0\,, \tag{A.6}$$

where $\Re(\mathbf{A})$ denotes the real part of $\mathbf{A}$. Substituting (A.3) in (A.6) and dividing by β^2 we get, after some computations

$$\text{tr}\left(\mathbf{F}\bar{\mathbf{G}}\bar{\mathbf{G}}^H\mathbf{F}^H + \text{E}[\mathbf{E}\bar{\mathbf{G}}\bar{\mathbf{G}}^H\mathbf{E}^H] - \beta\Re(\mathbf{G}^H\mathbf{F}^H) + \frac{\sigma_N^2}{P}\bar{\mathbf{G}}\bar{\mathbf{G}}^H\right) = 0\,. \tag{A.7}$$

Since $\bar{\mathbf{G}}^H(\mathbf{F}^H\mathbf{F} + \mathbf{R} + \xi\mathbf{I})\bar{\mathbf{G}}$ is a positive semi-definite matrix, its trace is a real number and from (A.3) we have

$$\begin{aligned}\text{tr}\left(\bar{\mathbf{G}}^H(\mathbf{F}^H\mathbf{F} + \mathbf{R} + \eta\mathbf{I})\bar{\mathbf{G}}\right) &= \text{tr}\left((\mathbf{F}^H\mathbf{F} + \mathbf{R} + \eta\mathbf{I})\bar{\mathbf{G}}\bar{\mathbf{G}}^H\right)\\ &= \text{tr}\left(\bar{\mathbf{G}}^H\mathbf{F}^H\right)\\ &= \text{tr}\left(\Re\left(\bar{\mathbf{G}}^H\mathbf{F}^H\right)\right)\end{aligned} \tag{A.8}$$

Applying (A.8) in (A.7) we get, $\text{tr}\left(\left(-\eta\mathbf{I} + \frac{\sigma_N^2}{P}\mathbf{I}\right)\bar{\mathbf{G}}\bar{\mathbf{G}}^H\right) = 0$ and the resulting value for η becomes

$$\eta = \frac{\sigma_N^2}{P}\,. \tag{A.9}$$

Substituting (A.9) in (A.3), (A.5) and (A.4) we finally get (3.20), (3.21) and (3.22), respectively.

A.2 Proof of Lemma 1

Firstly we derive $\text{E}\left[\tilde{\boldsymbol{\epsilon}}_k^H\tilde{\boldsymbol{\epsilon}}_k\right]$ under the hypothesis: a) $\bar{\mathbf{h}}_k\tilde{\boldsymbol{\epsilon}}_k^H = 0$, b) $\tilde{\boldsymbol{\epsilon}}_k\tilde{\boldsymbol{\epsilon}}_k^H = 1$, c) $\bar{\mathbf{h}}_k\bar{\mathbf{h}}_k^H = 1$, d) all direction of $\tilde{\boldsymbol{\epsilon}}_k$ in the space orthogonal to $\bar{\mathbf{h}}_k$ are equally probable.

From vector $\bar{\mathbf{h}}_k$, by the orthonormalization procedure of Gram-Schmidt, we obtain a $(M-1)\times M$ orthonormal matrix $\mathbf{A}_k$, such that $\mathbf{A}_k\bar{\mathbf{h}}_k^H = \mathbf{0}$ and $\tilde{\boldsymbol{\epsilon}}_k = \mathbf{x}_k\mathbf{A}_k$, with $\mathbf{x}_k$ a $1\times M-1$ unit-norm vector. We also have

$$\text{E}[\tilde{\boldsymbol{\epsilon}}_k^H\tilde{\boldsymbol{\epsilon}}_k] = \mathbf{A}_k^H\text{E}[\mathbf{x}_k^H\mathbf{x}_k]\mathbf{A}_k\,. \tag{A.10}$$

Then, we can write $[\mathbf{x}_k]_q = |\,[\mathbf{x}_k]_q\,|e^{j\varphi_{k,q}}$. We assume that $\varphi_{k,q}$ are i.i.d. uniform random variables in $(0, 2\pi]$, while $[|\boldsymbol{x}_k|]_i$ are i.i.d. zero mean variables, so that

$$\text{E}[x_p x_q^*] = \begin{cases} 0 & p \neq q \\ \text{E}[|[\boldsymbol{x}_k]_p|^2] & p = q\,. \end{cases} \tag{A.11}$$

We now write $\boldsymbol{x}_k$ in hyperspherical coordinates as

$$|[\mathbf{x}_k]_i| = \cos(\phi_i)\prod_{p=1}^{i-1}\sin(\phi_p)\,, \quad i = 1, 2, \ldots, M-2\,, \tag{A.12}$$

$$|[\mathbf{x}_k]_{M-1}| \;=\; \prod_{p=1}^{M-2} \sin(\phi_p)\,, \tag{A.13}$$

where ϕ_i, $i = 1, 2, \ldots, M-2$ are independent uniform random variables in the range $(0, 2\pi]$. Hence we obtain $\mathrm{E}[|[\mathbf{x}_k]_p|^2] = \frac{1}{2^p}$, $p < M-1$ and $\mathrm{E}[|[\mathbf{x}_k]_{M-1}|^2] = \frac{1}{2^{M-2}}$. From (A.10) and (A.11) we obtain (3.25).

Lastly, from the definition of $\mathbf{R}$, i.e. $\mathbf{R} \triangleq \mathrm{E}[\mathbf{E}^H\mathbf{E}]$, and the assumption that all vectors $\boldsymbol{\epsilon}_k$ are independent, (3.25) leads to (3.23).

A.3 Suboptimum performance metric for RBCM-Q

Applying Jensen's inequality and exploiting the concavity of the function $\log_2(1+x)$, an upper bound of (6.5) for a generic codematrix $\mathbf{C}_i$ is

$$R(\mathbf{H}, \mathbf{C}_i) \leq L \log_2 \left(1 + \frac{1}{L} \sum_{\ell=0}^{L-1} \frac{\rho ||\mathbf{h}(\ell)||^2 |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2}{1 + \rho ||\mathbf{h}(\ell)||^2 (1 - |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2)} \right) . \tag{A.14}$$

Since $\log_2(1+x)$ is a strictly monotonically increasing function, maximizing the right hand side of (A.14) is equivalent to maximizing

$$R_{bound}(\mathbf{H}, \mathbf{C}_i) = \sum_{\ell=0}^{L-1} \frac{\rho ||\mathbf{h}(\ell)||^2 |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2}{1 + \rho ||\mathbf{h}(\ell)||^2 (1 - |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2)} . \tag{A.15}$$

Note that (A.15) depends on the transmitted power P, through ρ, hence different codebooks should be designed for different SNRs. With the aim of designing practical codebooks, we focus only on two limit situations: low-SNR and high-SNR regimes.

In the low-SNR regime, when multiuser interference is negligible with respect to the channel noise, the expectation of (A.15) yields

$$\begin{aligned} \text{Low-SNR} \qquad \mathrm{E}\left[R_{bound}(\mathbf{H}, \mathbf{C}_i)\right] &= \mathrm{E}\left[\sum_{\ell=0}^{L-1} \rho ||\mathbf{h}(\ell)||^2 |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2\right] \\ &= \rho \sum_{\ell=0}^{L-1} \mathrm{E}\left[||\mathbf{h}(\ell)||^2\right] \mathrm{E}\left[|\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2\right] \end{aligned} \tag{A.16}$$

where in the last equality the independence between the norm and the direction of the MIMO channel has been used. Moreover, as $\mathrm{E}\left[||\mathbf{h}(\ell)||^2\right]$ is the same for all subcarriers, the maximization of (A.16) follows from maximizing the expectation of (6.13).

In the high-SNR regime, multiuser interference becomes predominant and applying Jensen's inequality to the expectation of (A.15) yields

$$\mathrm{E}\left[R_{bound}(\mathbf{H}, \mathbf{C}_i)\right] \;\geq\; L\,\mathrm{E}\left[\frac{\sum_{\ell=0}^{L-1} \frac{1}{L} |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2}{1 - \sum_{\ell=0}^{L-1} \frac{1}{L} |\tilde{\mathbf{h}}(\ell)\ \mathbf{c}_i(\ell)^H|^2}\right] \tag{A.17}$$

A further application of Jensen's inequality on (A.17) leads to the following

$$\text{High-SNR} \qquad \mathrm{E}\left[R_{bound}(\mathbf{H},\mathbf{C}_i)\right] \;\geq\; L\frac{\frac{1}{L}\mathrm{E}\left[\sum_{\ell=0}^{L-1}|\tilde{\mathbf{h}}(\ell)\;\mathbf{c}_i(\ell)^H|^2\right]}{1-\frac{1}{L}\mathrm{E}\left[\sum_{\ell=0}^{L-1}|\tilde{\mathbf{h}}(\ell)\;\mathbf{c}_i(\ell)^H|^2\right]}\,, \quad \text{(A.18)}$$

showing that (6.13) is a reasonable metric also in the high-SNR regime, as (A.18) is maximized by maximizing the expectation of (6.13).

Appendix B

Proof of theorems for networked control systems

B.1 Optimal state estimation in case of packet drops and delays: proof of Theorem 3

(**a**) Since the information available at the estimator at time t is given by the time-varying linear stochastic system given by (7.1)-(7.5), then the optimal estimator is given by its corresponding time-varying Kalman filter [102]

$$\begin{aligned}
\hat{\mathbf{x}}^t_{k|k} &= \mathbf{A}\hat{\mathbf{x}}^t_{k-1|k-1} + \tilde{\mathbf{K}}^t_k\left(\tilde{\mathbf{y}}^t_k - \mathbf{C}^t_k\mathbf{A}\hat{\mathbf{x}}^t_{k-1|k-1}\right) && \text{(B.1)}\\
\tilde{\mathbf{K}}^t_k &= \mathbf{P}^t_{k|k-1}\left(\mathbf{C}^t_k\right)^T\left(\mathbf{C}^t_k\mathbf{P}^t_{k|k-1}\left(\mathbf{C}^t_k\right)^T + \mathbf{R}^t_k\right)^{\dagger} && \text{(B.2)}\\
\mathbf{P}^t_{k+1|k} &= \mathbf{A}\mathbf{P}^t_{k|k-1}\mathbf{A}^T + \mathbf{Q} + \\
&\quad -\mathbf{A}\mathbf{P}^t_{k|k-1}\left(\mathbf{C}^t_k\right)^T\left(\mathbf{C}^t_k\mathbf{P}^t_{k|k-1}\left(\mathbf{C}^t_k\right)^T + \mathbf{R}^t_k\right)^{\dagger}\mathbf{C}^t_k\mathbf{P}^t_{k|k-1}\mathbf{A}^T && \text{(B.3)}\\
\hat{\mathbf{x}}^t_{0|0} &= \bar{\mathbf{x}}_0\,, \quad \mathbf{P}^t_{1,0} = \mathbf{P}_0 && \text{(B.4)}
\end{aligned}$$

where $\mathbf{C}^t_k = \left[\gamma^t_{1,k}\mathbf{C}^T_1, \gamma^t_{2,k}\mathbf{C}^T_2\right]^T$ and $\mathbf{R}^t_k = \text{diag}\left(\gamma^t_{1,k}\mathbf{R}_1, \gamma^t_{2,k}\mathbf{R}_2\right)$. Using the properties of the pseudo-inverse it's easy to verify that

$$\begin{aligned}
\left(\mathbf{C}^t_k\right)^T\left(\mathbf{C}^t_k\mathbf{P}^t_{k|k-1}\left(\mathbf{C}^t_k\right)^T + \mathbf{R}^t_k\right)^{\dagger}\mathbf{C}^t_k = \\
\gamma^t_{1,k}\gamma^t_{2,k}\mathbf{C}^T\left(\mathbf{C}\mathbf{P}^t_{k|k-1}\mathbf{C}^T + \mathbf{R}\right)^{-1}\mathbf{C} + \\
\gamma^t_{1,k}(1-\gamma^t_{2,k})\mathbf{C}^T_1\left(\mathbf{C}_1\mathbf{P}^t_{k|k-1}\mathbf{C}^T_1 + \mathbf{R}_1\right)^{-1}\mathbf{C}_1 + \\
(1-\gamma^t_{1,k})\gamma^t_{2,k}\mathbf{C}^T_2\left(\mathbf{C}_2\mathbf{P}^t_{k|k-1}\mathbf{C}^T_2 + \mathbf{R}_2\right)^{-1}\mathbf{C}_2 \qquad \text{(B.5)}
\end{aligned}$$

which is equivalent to consider only correctly received observations in the state estimation process. From (B.1)-(B.5) we easily get (7.8)-(7.11).

(**b**) Let us consider $t > N$. If $\gamma^t_{i,k} = \gamma^{t-1}_{i,k}$, $i = 1, 2$, $\forall k \geq 1$, $\forall t \geq k + N$, then also

$\mathbf{P}^t_{k+1|k} = \mathbf{P}^{t-1}_{k+1|k}$ and $\hat{\mathbf{x}}^t_{k|k} = \hat{\mathbf{x}}^{t-1}_{k|k}$ hold under the same conditions on the time and sensor indices. In particular for $k = t - N$ we have $\mathbf{P}^t_{t-N+1|t-N} = \mathbf{P}^{t-1}_{t-N+1|t-N}$ and $\hat{\mathbf{x}}^t_{t-N|t-N} = \hat{\mathbf{x}}^{t-1}_{t-N|t-N}$. Therefore it is not necessary to compute $\mathbf{P}^t_{t+1|t}$ and $\hat{\mathbf{x}}^t_{t|t}$ at any time step t starting from $k = 1$, but it is sufficient to use the values $\mathbf{P}^{t-1}_{t-N+1|t-N}$ and $\hat{\mathbf{x}}^{t-1}_{t-N|t-N}$ pre-computed at the previous time step $t-1$, as in (7.12)-(7.13), and then iterate (7.8)-(7.11) for the last N observations.

B.2 Optimal estimator with constant gains: proof of Theorem 4

The proof of Theorem 4 can be recovered along the lines of [18, Theorem 3] and is only outlined in the following.

Let us define the following operator:

$$\begin{aligned}\mathcal{L}_{\lambda_1,\lambda_2}(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{P}) &= (1-\lambda_1)(1-\lambda_2)\left(\mathbf{APA}^T+\mathbf{Q}\right)+\lambda_1\lambda_2\left(\mathbf{AFPF}^T\mathbf{A}^T+\mathbf{V}\right)+\\ &\quad \lambda_1(1-\lambda_2)\left(\mathbf{AF}_1\mathbf{PF}_1^T\mathbf{A}^T+\mathbf{V}_1\right)+\\ &\quad (1-\lambda_1)\lambda_2\left(\mathbf{AF}_2\mathbf{PF}_2^T\mathbf{A}^T+\mathbf{V}_2\right) \qquad \text{(B.6)}\end{aligned}$$

where $\mathbf{F} = \mathbf{I}-\mathbf{KC}$, $\mathbf{F}_i = \mathbf{I}-\mathbf{K}_i\mathbf{C}_i$, $\mathbf{V} = \mathbf{Q}+\mathbf{AKRK}^T\mathbf{A}^T$ and $\mathbf{V}_i = \mathbf{Q}+\mathbf{AK}_i\mathbf{R}_i\mathbf{K}_i^T\mathbf{A}^T$. From (7.12)-(7.14), exploiting the independence between $\mathbf{w}_k$, $\mathbf{v}_{i,k}$, $\gamma^t_{i,k}$, $\tilde{\mathbf{e}}^t_{k+1|k}$ and the fact that $\mathbf{v}_{i,k}$, $\mathbf{w}_k$ are zero mean random vectors we get

$$\overline{\mathbf{P}}^t_{t-N+2|t-N+1} = \mathcal{L}_{\lambda_{1,N-1},\lambda_{2,N-1}}\left(\mathbf{K}_{N-1},\mathbf{K}_{1,N-1},\mathbf{K}_{2,N-1},\overline{\mathbf{P}}^{t-1}_{t-N+1|t-N}\right) \qquad \text{(B.7)}$$

$$\overline{\mathbf{P}}^t_{t-k+1|t-k} = \mathcal{L}_{\lambda_{1,k},\lambda_{2,k}}\left(\mathbf{K}_k,\mathbf{K}_{1,k},\mathbf{K}_{2,k},\overline{\mathbf{P}}^t_{t-k|t-k-1}\right),\ k = N-2,\ldots,0 \qquad \text{(B.8)}$$

Notice that (B.7)-(B.8) define a set of linear deterministic equations for fixed $\lambda_{i,k}$, $\mathbf{K}_k$ and $\mathbf{K}_{i,k}$. In particular if we define $\mathbf{S}_t = \overline{\mathbf{P}}^t_{t-N+1|t-N}$ then (B.7) can be rewritten as

$$\mathbf{S}_{t+1} = \mathcal{L}_{\lambda_{1,N-1},\lambda_{2,N-1}}\left(\mathbf{K}_{N-1},\mathbf{K}_{1,N-1},\mathbf{K}_{2,N-1},\mathbf{S}_t\right). \qquad \text{(B.9)}$$

Since all matrices $\overline{\mathbf{P}}^t_{t-k+1|t-k}$, $k = 0,\ldots,N-1$ can be obtained from $\mathbf{S}_t$, it follows that the stability of the constant-gains estimator can be completely inferred from the properties of the operator $\mathcal{L}_{\lambda_1,\lambda_2}(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{P})$. In the following lemma we study the properties of this operator. The proof of Lemma 6 can be recovered along the lines of [18, Theorem 2] using results from [101].

Lemma 6 *Assume that* $\mathbf{P} \geq 0$, $(\mathbf{A},\mathbf{C})$ *is observable,* $(\mathbf{A},\mathbf{Q}^{1/2})$ *is controllable and* $0 \leq \lambda_i \leq 1$. *Then the following statements are true:*

(a) $\mathcal{L}_{\lambda_1,\lambda_2}(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{P}) = \Phi_{\lambda_1,\lambda_2}(\mathbf{P})+\Xi_{\lambda_1,\lambda_2}(\mathbf{K}-\mathbf{K}_P,\mathbf{K}_1-\mathbf{K}_{1,P},\mathbf{K}_2-\mathbf{K}_{2,P},\mathbf{P})$ *where*

$$\begin{aligned}\Xi_{\lambda_1,\lambda_2}(\mathbf{K}-\mathbf{K}_P,\mathbf{K}_1-\mathbf{K}_{1,P},\mathbf{K}_2-\mathbf{K}_{2,P},\mathbf{P}) = \\ \lambda_1\lambda_2\mathbf{A}\left(\mathbf{K}-\mathbf{K}_P\right)\left(\mathbf{C}\mathbf{P}\mathbf{C}^T+\mathbf{R}\right)\left(\mathbf{K}-\mathbf{K}_P\right)^T\mathbf{A}^T + \\ \lambda_1(1-\lambda_2)\mathbf{A}\left(\mathbf{K}_1-\mathbf{K}_{1,P}\right)\left(\mathbf{C}_1\mathbf{P}\mathbf{C}_1^T+\mathbf{R}_{11}\right)\left(\mathbf{K}_1-\mathbf{K}_{1,P}\right)^T\mathbf{A}^T + \\ (1-\lambda_1)\lambda_2\mathbf{A}\left(\mathbf{K}_2-\mathbf{K}_{2,P}\right)\left(\mathbf{C}_2\mathbf{P}\mathbf{C}_2^T+\mathbf{R}_{22}\right)\left(\mathbf{K}_2-\mathbf{K}_{2,P}\right)^T\mathbf{A}^T\end{aligned} \quad \text{(B.10)}$$

(b) $\mathcal{L}_{\lambda_1,\lambda_2}\left(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{P}\right) \geq \Phi_{\lambda_1,\lambda_2}(\mathbf{P}) = \mathcal{L}_{\lambda_1,\lambda_2}\left(\mathbf{K}_P,\mathbf{K}_{1,P},\mathbf{K}_{2,P},\mathbf{P}\right), \forall \mathbf{K},\mathbf{K}_i$

(c) $0 \leq \mathbf{P}_1 \leq \mathbf{P}_2 \implies \Phi_{\lambda_1,\lambda_2}(\mathbf{P}_1) \leq \Phi_{\lambda_1,\lambda_2}(\mathbf{P}_2)$

(d) $0 \leq \lambda_1 \leq 1$ *fixed and* $0 \leq \lambda_2^{(1)} \leq \lambda_2^{(2)} \leq 1 \implies \Phi_{\lambda_1,\lambda_2^{(1)}}(\mathbf{P}) \geq \Phi_{\lambda_1,\lambda_2^{(2)}}(\mathbf{P})$. *Similarly for* $0 \leq \lambda_2 \leq 1$ *fixed and* $0 \leq \lambda_1^{(1)} \leq \lambda_1^{(2)} \leq 1 \implies \Phi_{\lambda_1^{(1)},\lambda_2}(\mathbf{P}) \geq \Phi_{\lambda_1^{(2)},\lambda_2}(\mathbf{P})$.

(e) *If there exists* $\mathbf{P}^*$ *such that* $\mathbf{P}^* = \mathcal{L}_{\lambda_1,\lambda_2}\left(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{P}^*\right)$, *then* $\mathbf{P}^* \geq 0$ *and it is unique. Consequently this is true also for* $\mathbf{K} = \mathbf{K}_P$, $\mathbf{K}_i = \mathbf{K}_{i,P}$, $i = 1,2$, *where* $\mathbf{P}^* = \Phi_{\lambda_1,\lambda_2}(\mathbf{P}^*)$.

(f) *If* $0 \leq \lambda_1 \leq 1$ *is fixed* $0 \leq \lambda_2^{(1)} \leq \lambda_2^{(2)} \leq 1$ *and there exists* $\mathbf{P}_1^*$, $\mathbf{P}_2^*$ *such that* $\mathbf{P}_1^* = \Phi_{\lambda_1,\lambda_2^{(1)}}(\mathbf{P}_1^*)$ *and* $\mathbf{P}_2^* = \Phi_{\lambda_1,\lambda_2^{(2)}}(\mathbf{P}_2^*)$, *then* $\mathbf{P}_1^* \geq \mathbf{P}_2^*$. *A similar property holds switching the role of* λ_1 *and* λ_2.

(g) *Let* $\mathbf{S}_{t+1} = \mathcal{L}_{\lambda_1,\lambda_2}\left(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{S}_t\right)$ *and* $\mathbf{S}_0 \geq 0$. *If* $\mathbf{S}^* = \mathcal{L}_{\lambda_1,\lambda_2}\left(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{S}^*\right)$ *has a solution, then* $\lim_{t\to\infty}\mathbf{S}_t = \mathbf{S}^*$, *otherwise the sequence* $\mathbf{S}_t$ *is unbounded.*

(h) *If there exists* $\mathbf{S}^*$, $\mathbf{K}$, $\mathbf{K}_i$ *such that* $\mathbf{S}^* = \mathcal{L}_{\lambda_1,\lambda_2}\left(\mathbf{K},\mathbf{K}_1,\mathbf{K}_2,\mathbf{S}^*\right)$, *then also matrix* $\mathbf{P}^* = \Phi_{\lambda_1,\lambda_2}(\mathbf{P}^*)$ *exists and* $\mathbf{P}^* \leq \mathbf{S}^*$.

(i) *If* $\mathbf{A}$ *is strictly stable, then* $\mathbf{P}^* = \Phi_{\lambda_1,\lambda_2}(\mathbf{P}^*)$ *has always a solution. Otherwise for fixed* $0 \leq \lambda_1 \leq 1$ *there exists* λ_2^c *such that* $\mathbf{P}^* = \Phi_{\lambda_1,\lambda_2}(\mathbf{P}^*)$ *has solution if and only if* $\lambda_2 > \lambda_2^c$. *Moreover for a given* λ_1 *we have an upper bound and a lower bound for* λ_2^c, *i.e.* $\underline{\lambda}_2^c \leq \lambda_2^c \leq \overline{\lambda}_2^c$ *[101]. Similar considerations hold for fixed* $0 \leq \lambda_2 \leq 1$ *and varying* λ_1. □

From Lemma 6, we can now prove (**a**) and (**b**) in Theorem 4. First we prove by contradiction that, for a given N and with $\mathbf{A}$ unstable, there is an instability region for which no stable estimator with constant gains exists. Hence for given $0 \leq \lambda_{1,N-1} \leq 1$, let us assume there exists a stable estimator when $\lambda_{2,N-1} \leq \lambda_{2,N-1}^c$, i.e. there exist $\{\mathbf{K}_k\}_{k=0}^{N-1}$, $\{\mathbf{K}_{i,k}\}_{k=0}^{N-1}$, $i = 1,2$ such that $\overline{\mathbf{P}}_{t|t}^t$ is bounded for all t. Since $\overline{\mathbf{P}}_{t+1|t} = \mathbf{A}\overline{\mathbf{P}}_{t|t}^t\mathbf{A}^T + \mathbf{Q}$ also $\overline{\mathbf{P}}_{t+1|t}$ must be bounded for all t. From (B.7)-(B.8) it follows that $\overline{\mathbf{P}}_{t+1|t}$ is bounded if and only if $\overline{\mathbf{P}}_{t-k+1|t-k}^t$, $k = 0,\ldots,N-1$ are bounded for all t. Since the bounded sequence $\mathbf{S}_t = \overline{\mathbf{P}}_{t-N+1|t-N}^t$ needs to satisfy (B.9), from Lemma 6(**g**) it follows that $\mathbf{S}^* = \mathcal{L}_{\lambda_{1,N-1},\lambda_{2,N-1}}\left(\mathbf{K}_{N-1},\mathbf{K}_{1,N-1},\mathbf{K}_{2,N-1},\mathbf{S}^*\right)$ has a solution. Moreover from Lemma 6(**h**) it follows that also $\mathbf{P}^* = \Phi_{\lambda_{1,N-1},\lambda_{2,N-1}}(\mathbf{P}^*)$ has a solution. However according to Lemma 6(**i**)

$\mathbf{P}^* = \Phi_{\lambda_{1,N-1},\lambda_{2,N-1}}(\mathbf{P}^*)$ cannot have a solution, which contradicts the hypothesis that a stable estimator exists. The same arguments can be used for fixed $0 \leq \lambda_{2,N-1} \leq 1$ and varying $\lambda_{1,N-1}$.

Consider now the case in which the couple $(\lambda_{1,N-1}, \lambda_{2,N-1})$ belongs to the stability region. From Lemma 6(**h**), it follows that (7.18)-(7.21) are well defined and have a solution. Let us consider any other set of gains $\{\tilde{\mathbf{K}}_k\}_{k=0}^{N-1}$, $\{\tilde{\mathbf{K}}_{i,k}\}_{k=0}^{N-1}$, $i = 1, 2$ for which the following equations hold:

$$\mathbf{T}_{N-1}^N = \mathcal{L}_{\lambda_{1,N-1},\lambda_{2,N-1}}(\tilde{\mathbf{K}}_{N-1}, \tilde{\mathbf{K}}_{1,N-1}, \tilde{\mathbf{K}}_{2,N-1}, \mathbf{T}_{N-1}^N) \tag{B.11}$$

$$T_k^N = \mathcal{L}_{\lambda_{1,k},\lambda_{2,k}}(\tilde{\mathbf{K}}_k, \tilde{\mathbf{K}}_{1,k}, \tilde{\mathbf{K}}_{2,k}, \mathbf{T}_{k+1}^N),\ k = N-2, \ldots, 0 \tag{B.12}$$

From Lemma 6(**g**) it follows that $\lim_{t\to\infty} \overline{P}_{t-k+1|t-k}^t = \mathbf{V}_k^N$ for the optimal gains $\{\mathbf{K}_k\}_{k=0}^{N-1}$, $\{\mathbf{K}_{i,k}\}_{k=0}^{N-1}$ and $\lim_{t\to\infty} \overline{P}_{t-k+1|t-k}^t = \mathbf{T}_k^N$ when using generic gains $\{\tilde{\mathbf{K}}_k\}_{k=0}^{N-1}$, $\{\tilde{\mathbf{K}}_{i,k}\}_{k=0}^{N-1}$, $i = 1, 2$. From Lemma 6(**h**) it follows that $\mathbf{V}_{N-1}^N \leq \mathbf{T}_{N-1}^N$. From Lemma 6(**c**) we have

$$\begin{aligned}
\mathbf{V}_{N-2}^N &= \Phi_{\lambda_{1,N-2},\lambda_{2,N-2}}(\mathbf{V}_{N-1}^N) && \text{(B.13)}\\
&\leq \mathcal{L}_{\lambda_{1,N-2},\lambda_{2,N-2}}(\tilde{\mathbf{K}}_{N-2}, \tilde{\mathbf{K}}_{1,N-2}, \tilde{\mathbf{K}}_{2,N-2}, \mathbf{V}_{N-1}^N) && \text{(B.14)}\\
&\leq \mathcal{L}_{\lambda_{1,N-2},\lambda_{2,N-2}}(\tilde{\mathbf{K}}_{N-2}, \tilde{\mathbf{K}}_{1,N-2}, \tilde{\mathbf{K}}_{2,N-2}, \mathbf{T}_{N-1}^N) && \text{(B.15)}\\
&= \mathbf{T}_{N-2}^N && \text{(B.16)}
\end{aligned}$$

where (B.14) and (B.15) derive from Lemma 6(**b**) and (**c**), respectively. Inductively it's easy to show that $\mathbf{V}_k^N \leq \mathbf{T}_k^N$ for all $k = 0, \ldots, N-1$.

Now we want to show that $\mathbf{V}_0^{N+1} \leq \mathbf{V}_0^N$. From Lemma 6(**f**) and the property $\lambda_{i,N} \geq \lambda_{i,N-1}$ we get

$$\begin{aligned}
\mathbf{V}_N^{N+1} &= \Phi_{\lambda_{1,N},\lambda_{2,N}}(\mathbf{V}_N^{N+1})\\
&\leq \tilde{\mathbf{S}} = \Phi_{\lambda_{1,N},\lambda_{2,N-1}}(\tilde{\mathbf{S}})\\
&\leq \mathbf{V}_{N-1}^N = \Phi_{\lambda_{1,N-1},\lambda_{2,N-1}}(\mathbf{V}_{N-1}^N) && \text{(B.17)}
\end{aligned}$$

Therefore $\mathbf{V}_{N-1}^{N+1} = \Phi_{\lambda_{1,N-1},\lambda_{2,N-1}}(\mathbf{V}_N^{N+1}) \leq \Phi_{\lambda_{1,N-1},\lambda_{2,N-1}}(\mathbf{V}_{N-1}^N) = \mathbf{V}_{N-1}^N$ and inductively $\mathbf{V}_k^{N+1} \leq \mathbf{V}_k^N$ for all $k = N-1, \ldots, 0$, which proves the statement.

Finally, if τ_{max} is finite, then $\lambda_{i,k} = \lambda_{i,\tau_{max}}$ for all $k \geq \tau_{max}$. Assume $N > \tau_{max}$, then

$$\begin{aligned}
\mathbf{V}_{N-1}^N &= \Phi_{\lambda_{1,N-1},\lambda_{2,N-1}}(\mathbf{V}_{N-1}^N) = \Phi_{\lambda_{1,N-2},\lambda_{2,N-2}}(\mathbf{V}_{N-1}^N)\\
&= \mathbf{V}_{N-2}^N = \Phi_{\lambda_{1,N-2},\lambda_{2,N-2}}(\mathbf{V}_{N-2}^N) = \Phi_{\lambda_{1,N-3},\lambda_{2,N-3}}(\mathbf{V}_{N-2}^N) = \ldots\\
&= \mathbf{V}_{\tau_{max}}^N = \Phi_{\lambda_{1,\tau_{max}},\lambda_{2,\tau_{max}}}(\mathbf{V}_{\tau_{max}}^N)\,. && \text{(B.18)}
\end{aligned}$$

Since we also have $\mathbf{V}_{\tau_{max}}^{\tau_{max}} = \Phi_{\lambda_{1,\tau_{max}},\lambda_{2,\tau_{max}}}(\mathbf{V}_{\tau_{max}}^{\tau_{max}})$, then from Lemma 6(**e**) we have $\mathbf{V}_{\tau_{max}}^{\tau_{max}} = \mathbf{V}_{\tau_{max}}^N$. From (7.21) we then obtain $\mathbf{V}_k^{\tau_{max}} = \mathbf{V}_k^N$ for $k = \tau_{max} \ldots, 0$, which concludes the proof.

B.3 Optimal state control under TCP-like protocols: proof of Lemma 3

The proof is by induction. Let us define for ease of notation $\mathcal{F}_k = \{\mathcal{I}_k, \bar{\mathbf{x}}, \mathbf{u}_\infty\}$. The claim is clearly true for $k = N$ with the choice of parameters $\mathbf{S}_N = \mathbf{W}_N$, $\mathbf{T}_N = \mathbf{W}_N$, $\mathbf{Z}_N = \mathbf{0}$ and $c_N = \bar{\mathbf{x}}^T \mathbf{W}_N \bar{\mathbf{x}}$. Suppose now that the claim is true for $k+1$, the value function for time k is given by:

$$\begin{aligned} V_k(\mathbf{x}_k) &= \min_{\mathbf{u}_k^c} \mathrm{E}\big[(\mathbf{x}_k - \bar{x})^T \mathbf{W}_k (\mathbf{x}_k - \bar{\mathbf{x}}) + \nu_k (\bar{\mathbf{u}}_k^c)^T \mathbf{U}_k \bar{\mathbf{u}}_k^c + \\ &\quad (1-\nu_k)\mathbf{u}_\infty^T \mathbf{U}_k \mathbf{u}_\infty + V_{k+1}(\mathbf{x}_{k+1}) | \mathcal{F}_k \big] \\ &= \min_{\mathbf{u}_k^c} \mathrm{E}[(\mathbf{x}_k - \bar{\mathbf{x}})^T \mathbf{W}_k (\mathbf{x}_k - \bar{\mathbf{x}}) + \nu_k (\bar{\mathbf{u}}_k^c)^T \mathbf{U}_k \bar{\mathbf{u}}_k^c + \\ &\quad + (1-\nu_k)\mathbf{u}_\infty^T \mathbf{U}_k \mathbf{u}_\infty + \mathbf{x}_{k+1}^T \mathbf{S}_{k+1} \mathbf{x}_{k+1} - 2\bar{\mathbf{x}}^T \mathbf{T}_{k+1} \mathbf{x}_{k+1} + \\ &\quad 2(1-\bar{\nu})\mathbf{u}_\infty^T \mathbf{B}^T \mathbf{Z}_{k+1} \mathbf{A} \mathbf{x}_{k+1} + c_{k+1} | \mathcal{F}_k] \end{aligned} \tag{B.19}$$

where we used the property $\mathrm{E}[\mathrm{E}[g(\mathbf{x}_{k+1})|\mathcal{F}_{k+1}]|\mathcal{F}_k] = \mathrm{E}[g(\mathbf{x}_{k+1})|\mathcal{F}_k]$, $\forall g(\cdot)$ [108], to get the last equality. After some computations we can equivalently express (B.19) as:

$$\begin{aligned} V_k(\mathbf{x}_k) &= \mathrm{E}\big[\mathbf{x}_k^T (\mathbf{W}_k + \mathbf{A}^T \mathbf{S}_{k+1} \mathbf{A}) \mathbf{x}_k | \mathcal{F}_k\big] + \bar{\nu}\,\mathrm{tr}\left((\mathbf{B}^T \mathbf{S}_{k+1} \mathbf{B} + \mathbf{U}_k)\mathbf{R}_N\right) + \\ &\quad \mathrm{tr}(\mathbf{S}_{k+1}\mathbf{Q}) - 2\bar{\mathbf{x}}^T (\mathbf{W}_k + \mathbf{T}_{k+1}\mathbf{A})\hat{\mathbf{x}}_{k|k} + \\ &\quad 2(1-\bar{\nu})\mathbf{u}_\infty^T \mathbf{B}^T (\mathbf{S}_{k+1} + \mathbf{Z}_{k+1}\mathbf{A})\mathbf{A}\hat{\mathbf{x}}_{k|k} + \\ &\quad (1-\bar{\nu})\mathbf{u}_\infty^T \left(\mathbf{B}^T \mathbf{S}_{k+1}\mathbf{B} + \mathbf{U}_k + (1-\bar{\nu})\mathbf{B}^T(\mathbf{Z}_{k+1}\mathbf{A} + \mathbf{A}^T \mathbf{Z}_{k+1}^T)\mathbf{B}\right)\mathbf{u}_\infty + \\ &\quad \bar{\mathbf{x}}^T \mathbf{W}_k \bar{\mathbf{x}} - 2(1-\bar{\nu})\bar{\mathbf{x}}^T \mathbf{T}_{k+1}\mathbf{B}\mathbf{u}_\infty + \mathrm{E}[c_{k+1}|\mathcal{F}_k] + \\ &\quad \bar{\nu} \min_{\mathbf{u}_k^c} \Big((\mathbf{u}_k^c)^T (\mathbf{B}^T \mathbf{S}_{k+1}\mathbf{B} + \mathbf{U}_k)\mathbf{u}_k^c + \\ &\quad 2(\mathbf{u}_k^c)^T \mathbf{B}^T \left(\mathbf{S}_{k+1}\mathbf{A}\hat{\mathbf{x}}_{k|k} - \mathbf{T}_{k+1}^T \bar{\mathbf{x}} + (1-\bar{\nu})\mathbf{A}^T \mathbf{Z}_{k+1}^T \mathbf{B}\mathbf{u}_\infty\right)\Big) \end{aligned} \tag{B.20}$$

We notice that $V_k(\mathbf{x}_k)$ is a quadratic function of $\mathbf{u}_k^c$, therefore the optimum control signal can be simply obtained by solving the equation $\frac{\partial V_k}{\partial \mathbf{u}_k^c} = 0$ which gives (8.26). If we substitute the optimum $\mathbf{u}_k^c$ back into (B.20) after some computations we get:

$$\begin{aligned} V_k(\mathbf{x}_k) &= \mathrm{E}\Big[\mathbf{x}_k^T (\mathbf{W}_k + \mathbf{A}^T \mathbf{S}_{k+1}\mathbf{A} + \\ &\quad -\bar{\nu}\mathbf{A}^T \mathbf{S}_{k+1}\mathbf{B}(\mathbf{U}_k + \mathbf{B}^T \mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T \mathbf{S}_{k+1}\mathbf{A})\mathbf{x}_k | \mathcal{F}_k\Big] + \\ &\quad -2\bar{\mathbf{x}}^T (\mathbf{W}_k + \mathbf{T}_{k+1}\mathbf{A} - \bar{\nu}\mathbf{T}_{k+1}\mathbf{B}(\mathbf{U}_k + \mathbf{B}^T \mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T \mathbf{S}_{k+1}\mathbf{A})\hat{\mathbf{x}}_{k|k} + \\ &\quad 2(1-\bar{\nu})\mathbf{u}_\infty^T \mathbf{B}^T \Big(\mathbf{S}_{k+1} + \mathbf{Z}_{k+1}\mathbf{A} + \\ &\quad -\bar{\nu}\mathbf{Z}_{k+1}\mathbf{A}\mathbf{B}(\mathbf{U}_k + \mathbf{B}^T \mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T \mathbf{S}_{k+1}\Big)\mathbf{A}\hat{\mathbf{x}}_{k|k} + \\ &\quad \bar{\nu}\,\mathrm{tr}(\mathbf{A}^T \mathbf{S}_{k+1}\mathbf{B}(\mathbf{U}_k + \mathbf{B}^T \mathbf{S}_{k+1}\mathbf{B})^{-1}\mathbf{B}^T \mathbf{S}_{k+1}\mathbf{A}\mathbf{P}_{k|k}) + \mathrm{tr}(\mathbf{S}_{k+1}\mathbf{Q}) + \\ &\quad \bar{\nu}\,\mathrm{tr}(\mathbf{B}^T(\mathbf{S}_{k+1}\mathbf{B} + \mathbf{U}_k)\mathbf{R}_N) + E[c_{k+1}|\mathcal{F}_k] + \bar{\mathbf{x}}^T \mathbf{M}_{1,k}\bar{\mathbf{x}} + \\ &\quad (1-\bar{\nu})\mathbf{u}_\infty^T \mathbf{M}_{2,k}\mathbf{u}_\infty + 2(1-\bar{\nu})\mathbf{u}_\infty^T \mathbf{B}^T \mathbf{M}_{3,k}\mathbf{T}_{k+1}^T \bar{\mathbf{x}} \end{aligned} \tag{B.21}$$

where we used the property $\mathrm{E}[\mathbf{x}_k^T\mathbf{S}\mathbf{x}_k|\mathcal{F}_k] = \hat{\mathbf{x}}_{k|k}^T\mathbf{S}\hat{\mathbf{x}}_{k|k} + \mathrm{tr}(\mathbf{S}\mathbf{P}_{k|k})$, $\forall\mathbf{S} \geq 0$. Recalling (8.7), the claim given by (8.17) is satisfied for time k and for all $\mathbf{x}_k$ if and only if (8.20)-(8.23) are satisfied.

B.4 Optimal state control under TCP-like protocols in the infinite horizon: proof of Lemma 5

The proof of Lemma 5 can be recovered along the lines of [108, Theorem 5.6] and is only outlined in the following.

(a) Since the system is stable it can be shown that $\lim_{k\to\infty}\mathbf{S}_k = \mathbf{S}_\infty$ for all initial condition $\mathbf{S}_0 \geq \mathbf{0}$ [108]. Moreover the stability of the system also guarantee $\lim_{k\to\infty}\mathbf{T}_k = \mathbf{T}_\infty$ and $\lim_{k\to\infty}\mathbf{Z}_k = \mathbf{Z}_\infty$. Therefore (8.40)-(8.42) follows from (8.26).

(b) This is because the optimal state estimator described in Section 8.2 depends on the arrival sequences $\{\gamma_{i,k}\}$ and $\{\nu_k\}$.

(c) As it can be seen from (8.27), only $J_N^{(2)}$ depends on $\mathbf{u}_\infty$. Moreover, $j_\infty^{(2)} \triangleq \lim_{k\to\infty}\frac{1}{N}J_N^{(2)}$ exists as a consequence of system stability, and it is a quadratic function of $\mathbf{u}_\infty$. Therefore the optimum $\mathbf{u}_\infty$ can be simply obtained by solving the equation $\frac{\partial j_\infty^{(2)}}{\partial\mathbf{u}_\infty} = 0$, and this yields (8.43).

(d) Since the system is stable, using similar arguments of [101] we can show that all the following limits exist: $\lim_{k\to\infty}\overline{\mathbf{P}}_{k+1|k}$, $\lim_{k\to\infty}\overline{\mathbf{P}}_{k|k}$, $\lim_{k\to\infty}\underline{\mathbf{P}}_{k+1|k}$ and $\lim_{k\to\infty}\underline{\mathbf{P}}_{k|k}$. Therefore the claim follows from (8.35)-(8.37), (8.31)-(8.34) and (8.43).

Bibliography

[1] D. Tse and P. Viswanath, *Foundamentals of Wireless Communication*, Cambridge University Press, 2004.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.

[3] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit: the general case," in *Int. Symp. Info. Theory (ISIT)*, Nice, France, June 2007.

[4] Syed A. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Info. Theory*, vol. 53, no. 7, pp. 2637–2642, July 2007.

[5] V. R. Cadambe and Syed A. Jafar, "Interference alignment and the degrees of freedom for the k user interference channel," *IEEE Trans. Info. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[6] W. Yu, *Competition and Cooperation in Multi-user Communication Environments*, PhD thesis, Stanford University, 2002.

[7] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for gaussian vector multiple access channels," in *Int. Symp. Info Theory (ISIT)*, Washington, DC, 2001.

[8] W. Rhee, W. Yu, and J. M. Cioffi, "The optimality of beamforming in uplink multiuser wireless systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 86–96, Jan. 2004.

[9] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[10] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691 – 1706, July 2003.

[11] W. Yu and J. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875 – 1892, Sept. 2004.

[12] P. Viswanath and D. Tse, "Sum capacity of the vector gaussian channel and uplink-downlink duality," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1912 – 1921, Aug. 2003.

[13] N. Jindal, W. Rhee, S. Vishwanath, S.A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna gaussian broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1570–1580, April 2005.

[14] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[15] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. Telecomm. Related Technol.*, vol. 10, pp. 585–595, Nov.-Dec. 1999.

[16] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Magazine*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[17] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.

[18] L. Schenato, "Optimal estimation in networked control systems subject to random delay and packet drop," *IEEE Trans. Automat. Control*, vol. 53, no. 5, pp. 1311–1317, June 2008.

[19] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: an overview," *Proc. IEEE Special issue on "The Emerging Technology of Networked Control Systems"*, vol. 95, no. 1, pp. 108–137, Jan. 2007.

[20] N. Benvenuto and G. Cherubini, *Algorithms for Communications Systems and their Applications*, Wiley, 2002.

[21] G. Dimić and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Tran. Commun.*, vol. 53, no. 10, pp. 3857 – 3868, Oct. 2005.

[22] Samsung, "Downlink MIMO for EUTRA," 3GPP TGS RAN WG1,R1-060335, Feb. 2006.

[23] M. Kountouris and D. Gesbert, "Memory-based opportunistic multi-user beamforming," in *Int. Symp. Info. Theory (ISIT)*, Adelaide, Australia, Sept. 2005.

[24] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439 – 441, May 1983.

[25] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "MIMO precoding for decentralized receivers," *Intern. Symp. Info. Theory (ISIT)*, July 2002.

[26] C. Windpassinger, *Detection and Precoding for Multiple Input Multiple Output Channels*, PhD thesis, University of Erlangen-Nuremberg, 2004.

[27] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication - part II: perturbation," *IEEE. Trans. Commun.*, vol. 53, pp. 537 – 544, Mar. 2005.

[28] F. Boccardi, F. Tosato, and G. Caire, "Precoding schemes for the MIMO-GBC," *IEEE Intern. Zurich Seminar on Commun.*, Feb. 2006.

[29] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," *IEEE Trans. Info. Theory*, vol. 51, no. 11, pp. 3820 – 3833, Nov. 2005.

[30] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528 – 541, Mar. 2006.

[31] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication - part I: channel inversion and regularization," *IEEE. Trans. Commun.*, vol. 53, pp. 195 – 202, Jan. 2005.

[32] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Comm.*, vol. 21, no. 5, pp. 802–811, June 2003.

[33] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[34] Lai-U Choi and Ross D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20 – 24, Jan. 2004.

[35] Y. Wu, J. Zhang, H. Zheng, X. Xu, and S. Zhou, "Receive antenna selection in the downlink of multiuser MIMO systems," in *IEEE Vehic. Techn. Conf. (VTC) Spring*, Dallas, TX, Sept. 2005.

[36] Z. Pan, K.-K. Wong and T.-S. Ng, "Generalized multiuser orthogonal space-division multiplexing," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1969–1973, Nov. 2004.

[37] Z. Shen, R. Chen, J. G. Andrews, Jr. R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser mimo systems with block diagonalization," *IEEE Trans. on Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sept. 2006.

[38] F. Boccardi and H. Huang, "A near-optimum technique using linear precoding for the MIMO broadcast channel," in *IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Honolulu, Hawaii, Apr. 2007.

[39] F. Boccardi, H. Huang, and M. Trivellato, "A near-optimum precoding technique for downlink multiuser MIMO transmissions," *to appear in Bell Labs Tech. Journal*, Oct. 2008.

[40] F. Boccardi and H. Huang, "Zero-forcing precoding for the MIMO broadcast channel under per-antenna power constraints," in *IEEE Intern. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Cannes, France, June 2006.

[41] D. J. Love, R. W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, pp. 54–59, Oct. 2004.

[42] N. Jindal T. Yoo and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sept. 2007.

[43] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1468–1477, Sept. 2007.

[44] K. Huang, J. G. Andrews, and Jr R. W. Heath, "Performance of orthogonal beamforming for SDMA with limited feedabck," *to appear in IEEE Trans. Vehic. Tech.*, May 2008.

[45] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Info. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[46] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Info. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[47] D. J. Love and R. W. Heath Jr., "Grassmannian beamforming on correlated MIMO channels," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dallas, TX, Dec. 2004.

[48] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel feedback for MIMO broadcast channels," in *IEEE Intern. Conf. Commun. (ICC)*, Beijing, China, May 2008.

[49] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO downlink made practical: Achievable rates with simple channel state estimation and feedback schemes," *submitted to IEEE Trans. Info. Theory*, Nov. 2007.

[50] M. Trivellato, F. Boccardi, and F. Tosato, "User selection schemes for MIMO broadcast channels with limited feedback," in *IEEE Vehic. Techn. Conf. (VTC) Spring*, Dublin, Ireland, Apr. 2007.

[51] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Joint low-rate feedback and channel quantization for the MIMO broadcast channel," in *IEEE Africon'07*, Windhoek, Namibia, Sept. 2007.

[52] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Predictive channel quantization and beamformer design for MIMO-BC with limited feedback," in *IEEE Global Commun. Conf. (GLOBECOM)*, Washington DC, Nov. 2007.

[53] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Low-rate predictive feedback for the OFDM MIMO broadcast channel," in *Tyrrenhian Intern. Workshop on Digital Commun. (TIWDC)*, Napoli, Italy, Sept. 2007.

[54] Philips, "Comparison between MU-MIMO codebook-based channel reporting techniques for LTE downlink," 3GPP TSG RAN WG1 Meeting 46bis, R1-062483, Oct. 2006.

[55] Philips, "System-level simulation results for channel vector quantisation feedback for mu-mimo," 3GPP TSG RAN WG1 Meeting 47, R1-063028, Nov. 2006.

[56] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Commun. Mag.*, vol. 38, pp. 70–77, July 2000.

[57] C. R. Murthy and B. D. Rao, "Quantization methods for equal gain transmission with finite rate feedback," *IEEE Trans. Sign. Process.*, vol. 55, no. 1, pp. 233–245, Jan. 2007.

[58] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84–95, Jan. 1980.

[59] L. Liu and H. Jafarkhani, "Novel transmit beamforming schemes for time-selective fading multiantenna systems," *IEEE Trans. Sign. Process.*, vol. 54, pp. 4767 – 4781, Dec. 2006.

[60] J. Salo, G. Del Galdo, J. Salmi, P. Kyösti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP Spatial Channel Model," Tech. Rep., 3GPP TR 25.996, Jan. 2005, Available: http://www.tkk.fi/Units/Radio/scm/.

[61] S. Zhou, B. Li, and P. Willett, "Recursive and trellis-based feedback reduction for MIMO-OFDM with transmit beamforming," in *IEEE Global. Commun. Conf. (GLOBECOM)*, St. Louis, MO, Nov. 2005.

[62] N. Jindal, "Antenna combining for the MIMO downlink channel," *to appear in IEEE Trans. Wireless Commun.*, Apr. 2007.

[63] F. Boccardi, H. Huang, and M. Trivellato, "Multiuser eigenmmode transmission for MIMO broadcast channels with limited feedback," in *IEEE Intern. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

[64] M. Trivellato, H. Huang, and F. Boccardi, "Antenna combining and codebook design for the MIMO broadcast channel with limited feedback," in *Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2007.

[65] M. Trivellato, F. Boccardi, and H. Huang, "On transceiver design and channel quantization for downlink multiuser MIMO systems with limited feedback," *IEEE J. Sel. Areas Commun. (JSAC)*, vol. 6, no. 8, pp. 1494–1504, Oct. 2008.

[66] M. Trivellato, F. Boccardi, and H. Huang, "Zero-forcing vs unitary beamforming in multiuser MIMO systems with limited feedback," in *IEEE Inter. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Cannes, France, Sept. 2008.

[67] F. Boccardi, *Precoding schemes for MIMO downlink transmissions*, PhD Thesis, University of Padova, 2007.

[68] A. Bayesteh and A. K. Khandani, "On the user selection for MIMO broadcast channels," in *IEEE Intern. Symp. on Info. Theory (ISIT)*, Adelaide, Australia, Sept. 2005.

[69] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge Univ. Press, 2003.

[70] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*, Now publisher, 2004.

[71] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666–672, May 1998.

[72] F. Boccardi, H. Huang, and A. Alexiou, "Hierarchical quantization and its application to multiuser eigenmode transmissions for MIMO broadcast channels with limited feedback," in *IEEE Intern. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Athens, Greece, Sept. 2007.

[73] R. A. Valenzuela M. K. Karakayali, G. J. Foschini, "Network co ordination for spectrally efficent communication in cellular systems," *IEEE Wireless Communications Magazine*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[74] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," in *Info. Theory and Applications Workshop*, San Diego, CA, Jan. 2007.

[75] S. Venkatesan, "Coordinating base stations for greater uplink spectral efficiency in a cellular network," in *IEEE Intern. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Athens, Greece, Sept. 2007.

[76] P. Marsch and G. Fettweis, "A framework for optimizing the downlink performance of distributed antenna systems under a constrained backhaul," in *Proc. Eurpoean Wireless Conf. (EW '07)*, Apr. 2007.

[77] H. Huang and M. Trivellato, "Performance of multiuser MIMO and network coordination in downlink cellular networks," in *Workshop on Resource Alloc. in Wireless Networks (RAWNET)*, Berlin, Germany, Mar. 2008.

[78] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *accepted for publication in IEEE Trans. Wireless Commun. Letter*, Jan. 2009.

[79] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[80] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, 2005.

[81] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, Wiley, 1981.

[82] P. J. Smith, M. Shafi, and L. M. Garth, "Performance analysis for adaptive MIMO SVD transmission in a cellular system," in *Australian Commun. Theory Workshop*, Perth, Australia, Feb. 2006.

[83] H. Huang and R.A. Valenzuela, "Foundamental simulated performance of downlink fixed wireless cellular networks with multiple antennas," in *IEEE Intern Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sept. 2005, vol. 1, pp. 161–165.

[84] H. Ekstrom, A. Furuskar, J. Karlsson, M. Meyer, and S. Parkvall, "Technical solutions for the 3G long-term evolution," *IEEE Commun. Mag.*, pp. 38–45, Mar 2006.

[85] J. Choi and J. R. W. Heath, "Interpolation based transmit beamforming for MIMO-OFDM with limited feedback," *IEEE Trans. Signal Process.*, vol. 53, pp. 4125–4135, Aug. 2005.

[86] F. She, W. Chen, H. Luo, and X. Yang, "Minimum MSE-based MIMO-OFDM precoded spatil multiplexing systems with limited feedback," in *IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, Nov. 2007.

[87] M. Sternad, T. Svensson, and G. Klang, "The WINNER B3G system MAC concept," in *IEEE Vehic. Tech. Conf. (VTC) fall*, Montreal, Canada, Sept. 2006.

[88] S. Zhou, Z. Wang, and G. B. Giannakis, "Quantifying the power loss when transmit beamforming relies on finite-rate feedback," *IEEE Tran. Wireless Commun.*, vol. 4, no. 4, pp. 1948–1957, July 2005.

[89] H. S. Mehr and G. Caire, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," *submitted to IEEE Trans. Commun*, Apr. 2008.

[90] M. Trivellato, S. Tomasin, and N. Benvenuto, "Channel quantization and feedback strategies for multiuser MIMO-OFDM downlink systems," in *IEEE Global Commun. Conf. (GLOBECOM)*, New Orleans, LA, Nov. 2008.

[91] M. Trivellato, S. Tomasin, and N. Benvenuto, "On channel quantization and feedback strategies for multiuser MIMO-OFDM downlink systems," *accepted for publication in IEEE Trans. Commun.*, Nov. 2008.

[92] 3GPP, "3rd generation partnership project; technical specification group radio access network; physical layer procedures (FDD)," Tech. Rep., 3GPP TS 25.214, 2002-03, Available: http://www.mumor.org/public/background/25214-500.pdf.

[93] V. Raghavan, R. W. Heath, and A. M. Sayeed, "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1298 – 1310, Sept. 2007.

[94] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publisher, 1992.

[95] "Sensor networks and applications," *Proceedings of the IEEE. Special issue*, vol. 91, Aug. 2003.

[96] M. Kinter-Meyer and R. Conant, "Opportunities of wireless sensors and controls for building operation," *Energy Engineering Journal*, vol. 102, no. 5, pp. 27–48, 2005.

[97] M. Nekovee, "Ad hoc sensor networks on the road: the promises and challenges of vehicular ad hoc networks," in *Workshop on Ubiquitous Computing and e-Research*, Edinburgh, U.K., May. 2005.

[98] A. Willing, K. Matheus, and A. Wolisz, "Wireless technology in industrial networks," *Proceeding of the IEEE*, vol. 93, no. 6, pp. 1130–1151, June 2005.

[99] S. Oh, L. Schenato, P. Chen, and S. Sastry, "Tracking and coordinations of multiple agents using sensor networks: systems design, algorithms and experiments," *Proceeding of the IEEE*, vol. 95, no. 1, pp. 234–254, Jan. 2007.

[100] X. Liu and A. Goldsmith, "Wireless network design for distributed control," in *IEEE Conf. on Decision and Control (CDC)*, Paradise Island, Bahamas, Dec. 2004.

[101] X. Liu and A. Goldsmith, "Kalman filtering with partial observation losses," in *IEEE Conf. on Decision and Control (CDC)*, Paradise Island, Bahamas, Dec. 2004, pp. 1413–1418.

[102] D. Bertsekas, *Digital Programming and Optimal Control*, Athena Scientific, Belmont, Massachusetts, 1995.

[103] M. Trivellato and N. Benvenuto, "On estimation in networked control systems with random delays and partial observation losses," in *Intern Symp. Info. Theory and its Applications (ISITA)*, Auckland, New Zealand, Dec. 2008.

[104] M. Huang and S. Dey, "Stability of kalman filtering with markovian packet losses," *Automatica*, vol. 43, pp. 598–607, 2007.

[105] IEEE Std 802.11a 1999(R2003), "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. High-speed physical layer in the 5 GHz band," June 2003.

[106] P. Chen and S. Sastry, "Latency and connectivity analysis tools for wireless mesh networks," in *in Proc. First Intern. Conf. on Robot Commun. and Coord. (ROBO-COMM)*, Athens, Greece, Oct. 2007.

[107] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1453–1464, Sept. 2004.

[108] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proc. IEEE*, vol. 95, pp. 163–187, Jan. 2007.

[109] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, July 2004.

[110] J. H. Braslavsky, R. H. Middleton, and J. S. Freudenberg, "Feedback stabilization over signal-to-noise ratio constrained channels," *IEEE Trans. Automat. Control*, vol. 52, no. 8, pp. 1391–1403, Aug. 2004.

[111] A.C.F. Chan, D.H.K. Tsang, and S. Gupta, "TCP (transmission control protocol) over wireless links," in *IEEE Vehic. Techn. Conf., (VTC)*, Phoenix, AZ, May. 1997.

[112] M. Trivellato and N. Benvenuto, "State control in networked control systems under packet drops and limited transmission bandwidth," *submitted to IEEE Trans. Commun.*, June 2008.

[113] M. Trivellato and N. Benvenuto, "Cross-layer design of networked control systems," in *to appear in IEEE Intern. Conf. Commun. (ICC)*, Dresden, Germany, June 2009.

[114] L. Schenato, "Optimal sensor fusion for distributed sensors subject to random delay and packet loss," in *IEEE Conf. Decision and Control (CDC)*, New Orleans, LA, Dec. 2007.

[115] IEEE Std 802.15.4a 2007, "Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (WPANs)," pp. 1–203, Aug. 2007.

[116] G. Zanca, F. Zorzi, A. Zanella, and M. Zorzi, "Experimental comparison of RSSI-based localization algorithms for indoor wireless sensors networks," in *Proc. of ACM RealWSN*, Glasgow, Scotland, 2008.

Publications

► Journal Publications (submitted)

[JS1] M. Trivellato and N. Benvenuto, "State Control in Networked Control Systems under Packet Drops and Limited Transmission Bandwidth," submitted to *IEEE Trans. Commun.*, June 2008.

► Journal Publications (accepted/published)

[J1] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," accepted for publication in *IEEE Trans. Wireless Commun. Letter*.

[J2] M. Trivellato, S. Tomasin, and N. Benvenuto, "On Channel Quantization and Feedback Strategies for Multiuser MIMO-OFDM Downlink Systems," accepted for publication in *IEEE Trans. Commun.*.

[J3] M. Trivellato, F. Boccardi, and H. Huang, "On Transceiver Design and Channel Quantization for Downlink Multiuser MIMO Systems with Limited Feedback," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 8, pp. 1494-1504, Oct. 2008.

[J4] F. Boccardi, H. Huang, and M. Trivellato, "A near-optimum precoding technique for downlink multiuser MIMO transmissions," accepted for publication in *Bell Labs Tech. Journal*.

► Conference Publications (accepted/published)

[C1] M. Trivellato and N. Benvenuto, "Cross-Layer Design of Networked Control Systems," to appear in *Intern. Conf. on Commun. (ICC)*, Dresden, Germany, June 2009.

[C2] M. Trivellato and N. Benvenuto, "On Estimation in Networked Control Systems with Random Delays and Partial Observation Losses," in Proc. *Intern. Symp. Info. Theory and its Applications (ISITA)*, Auckland, New Zealand, Dec. 2008.

[C3] M. Trivellato, S. Tomasin, and N. Benvenuto, "Channel Quantization and Feedback Optimization in Multiuser MIMO-OFDM Downlink Systems," in Proc. *IEEE Global Commun. Conf. (GLOBECOM)*, New Orleans, LA, Nov. 2008

[C4] M. Trivellato, F. Boccardi, and H. Huang, "Zero-Forcing vs Unitary Beamforming in Multiuser MIMO Systems with Limited Feedback," in Proc. *IEEE Intern. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Cannes, France, Sept. 2008

[C5] H. Huang and M. Trivellato, "Performance of multiuser MIMO and network coordination in downlink cellular networks," in Proc. *Workshop on Resource Alloc. in Wireless Networks (RAWNET)*, Berlin, Germany, Mar. 2008.

[C6] M. Trivellato, H. Huang, and F. Boccardi, "Antenna Combining and Codebook Design for the MIMO Broadcast Channel with Limited Feedback," (invited paper), in Proc. *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, Nov. 2007.

[C7] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Low-rate Predictive Feedback for the OFDM MIMO Broadcast Channel," in Proc. *Tyrrhenian Intern. Workshop on Digital Commun. (TIWDC)*, Napoli, Italy, Sept. 2007.

[C8] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Predictive Channel Quantization and Beamformer Design for MIMO-BC with Limited Feedback," in Proc. *IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, Nov. 2007.

[C9] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Joint Low-Rate Feedback and Channel Quantization for the MIMO Broadcast Channel," in Proc. *IEEE Africon'07*, Windhoek, Namibia, Sept. 2007.

[C10] F. Boccardi, H. Huang, and M. Trivellato, "Multiuser Eigenmode Transmission for MIMO Broadcast Channels with Limited Feedback," in Proc. *IEEE Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

[C11] M. Trivellato, F. Boccardi, and F. Tosato, "User Selection Schemes for MIMO Broadcast Channels with Limited Feedback," in Proc. *Vehic. Techn. Conf. (VTC)*, Dublin, Ireland, April 2007.

[C12] A. Assalini, M. Trivellato, and S. Pupolin, "Performance Analysis of OFDM-OQAM Systems," in Proc. *Wireless Personal Multimedia Commun. (WPMC)*, Aalborg, Denmark, Sept. 2005.