**Télécom Paris (ENST)**
**Institut Eurécom**

**THESE**

Présentée pour Obtenir le Grade de Docteur
de l'Ecole Nationale Supérieure
des Télécommunications

Spécialité: Communication et Electronique

**Mari Kobayashi**

# Utilisation des Antennes Multiples dans un Lien Descendant des Systèmes Sans Fils

| | |
|---|---|
| Président | J.C.Belfiore, ENST (Paris,France) |
| Rapporteurs | R.Kohno, Yokohama National University (Yokohama, Japan) |
| | E.Leonardi, Politecnico di Torino (Turin, Italy) |
| Examinateurs | J. Boutros, ENST (Paris,France) |
| | D.Gesbert, Institut Eurécom (Sophia Antipolis, France) |
| Thesis supervisor | G. Caire, Institut Eurécom (Sophia Antipolis, France) |

24 Juin 2005

**Télécom Paris (ENST)**
**Institut Eurécom**

# THESIS

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
from Ecole Nationale Supérieure
des Télécommunications

Specializing: Communication and Electronics

**Mari Kobayashi**

# On the Use of Multiple Antennas for Downlink of Wireless Systems

| | |
|---|---|
| President | J.C.Belfiore, ENST (Paris,France) |
| Rapporteurs | R.Kohno, Yokohama National University (Yokohama, Japan) |
| | E.Leonardi, Politecnico di Torino (Torino, Italy) |
| Examiners | J. Boutros, ENST (Paris,France) |
| | D.Gesbert, Institut Eurécom (Sophia Antipolis, France) |
| Thesis supervisor | G. Caire, Institut Eurécom (Sophia Antipolis, France) |

June 24$^{th}$ 2005

Je tiens à exprimer mes plus sincères remerciements à mon directeur de thèse Professeur Giuseppe Caire pour la rigeur avec laquelle il a encadré ma thèse. Il a su me transmettre son exigence et son enthousiasme pour la recherche.

Mes remerciements vont également aux membres de mon jury, Professeur Jean-Claude Belfiore, Professeur Ryuji Kohno, Professeur Emilio Leonardi, Docteur Joseph Boutros, Docteur David Gesbert, pour leur attention et leurs commentaires constructifs envers mon travail.

Je suis également extrêmement reconnaissante envers David Gesbert pour la collaboration fructueuse que nous avons eu.

France Telecom R&D a rendu possible ce travail grâce au financement qu'il a apporté à Eurecom. Je tiens à ce propos à remercier Docteur Raphaël Visoz pour son suivi de ma thèse.

Je remercie mes collègues d'Eurecom avec qui j'ai eu des discussions constructives au cours de mon travail, notamment Maxime, Albert, Kader, Marios, Merouane et Dirk Slock. J'ai partagé des moments enrichissants avec de nombreux autres, en particulier Souad, Navid, Stefania, Saad, Ruben, Hilde.

Je remercie mes parents, ma source d'énergie fondamentale, pour avoir su respecter mon choix de faire une thèse.

Enfin, je veux remercier Julien, mon mari, pour son soutien continu pendant ces trois ans de thèse. Je souhaite la réussite de sa thèse dans un an.

le 4 Juillet 2005
Mari Kobayashi
Sophia Antipolis

# Abstract

The role of multiple antennas in wireless communications has evolved greatly in the recent years. For point-to-point communication, Multi-Input Multi-Output (MIMO) technology has emerged as an attractive solution to increase the capacity by providing multiplexing gain and/or improve the link quality by diversity gain without channel state information at transmitter (CSIT). Chapter 2 focuses on the use of multiple antennas to increase the reliability of wireless channels via space-time coding (STC): coding schemes whose codewords are transmitted across the time and space dimension introduced by multiple transmit antennas. For single-carrier transmission over frequency selective multi-input single-output (MISO) fading channels, we propose a STC scheme that concatenates trellis-coded modulation (TCM) with time-reversal orthogonal space-time block coding (TR-STBC). The decoder is based on reduced-state joint equalization and decoding that operates directly on the TCM trellis without trellis state expansion. Hence, the decoder complexity is independent of the channel memory and of the constellation size. We show that, in the limit of large block length, the proposed scheme with reduced-state joint equalization and decoding can achieve the full diversity offered by the MISO multipath channel. Numerical examples show that our proposed scheme offers similar/superior performance at significantly lower complexity with respect to the best schemes previously proposed.

Chapter 3 and Chapter 4 investigate the use of multiple transmit antennas in a point-to-multipoint downlink system (MIMO broadcast channel) with CSIT. When the users in a downlink system experience an i.i.d. fading channel and the transmitter has perfect and instantaneous CSI, serving the user enjoying the best channel conditions yields another dimension of diversity called "multiuser diversity". If the transmitter is equipped with multiple antennas, opportunistic beamforming can be exploited by inducing an artificial fading that varies faster than its coherence time so as to prevent the users

from staying in bad fades. This is an efficient use of multiple antennas over a slow fading channel where the fading is considered as a constructive source of multiuser diversity. In Chapter 3, we compare STC (transmit diversity) and random "opportunistic" beamforming by considering a simple rate feedback. We take into account the following fundamental and realistic aspects missed in the existing works : random packet arrivals, correlated block-fading channels and non-perfect CSIT. We derive an adaptive policy that stabilizes the transmit queues whenever the arrival rates are in the system stability region. The realistic assumption of non-perfect CSIT yields a non-trivial tradeoff between multiuser diversity achieved by opportunistic beamforming and transmit diversity achieved by STC. In particular, the ability of accurately predicting CSIT emerges as a key factor of multiuser diversity based scheduling. For a given feedback delay, the relative merit of opportunistic beamforming versus STC strongly depends on the channel Doppler bandwidth. For slow fading channels opportunistic beamforming can achieve the best average delay by making the channel vary i.i.d. from one slot to another, while for fast fading channels STC performs better because it increases the link reliability ("channel-hardening effect").

In Chapter 4, we consider a more informative feedback that enables a smarter use of the multiple antennas at the transmitter. In particular, we consider a distortion-wise optimal "analog feedback" in which each user terminal sends back its estimated channel vector without quantizing and coding over a feedback link. We restrict our signaling strategy to linear beamforming because it is simple and can be generalized to non-perfect CSIT case. Then, we propose a novel method for user selection and linear beamforming that maximizes the weighted sum rate over a beamforming matrix. Under perfect CSIT, the proposed scheme achieves near dirty-paper-coding performance by exploiting multiplexing gain, power gain, and multiuser diversity. Over a time-varying channel with a given feedback delay, our scheme achieves the smallest average delay than the previously considered STC and opportunistic beamforming for any channel Doppler bandwidth. For slow varying channels, our proposed scheme can exploit multiplexing gain and power gain to balance the buffer queues. For faster channels, while opportunistic $M$-beamforming becomes strongly interference-limited and then unstable, the proposed scheme essentially allocates the whole power to the user with the largest weighted rate at each slot. Compared to STC that also operates in TDMA mode, our proposed scheme achieves better average delay because of the larger outage rate.

# Contents

# List of Figures

# List of Tables

# Acronyms

Here are the main acronyms used in this document. The meaning of an acronym is usually indicated once, when it first occurs in the text.

| | |
|---|---|
| ARQ | Automatic Repeat Request |
| AWGN | Additive White Gaussian Noise |
| BC | Broadcast Channel |
| BER | Bit Error Rate |
| BICM | Bit Interleaved Coded Modulation |
| CDMA | Code Division Multiple Access |
| CSI | Channel State Information |
| CSIT | Channel State Information at Transmitter |
| i.i.d. | independent and identically distributed |
| ISI | Inter Symbol Interference |
| FER | Frame Error Rate |
| GA | Gaussian Approximation |
| GM | Gauss-Markov |
| GOD | Generalized Orthogonal Design |
| MFB | Matched Filter Bound |
| MISO | Multiple-Input Single-Output |
| MIMO | Multiple-Input Multiple-Output |
| ML | Maximum Likelihood |
| MMSE | Minimum Mean-Square Error |
| MMSE-DFE | Minimum Mean Square Error Decision Feedback Equalizer |
| PSK | Phase Shift Keying |
| PSP | Per Survivor Processing |
| QAM | Quadrature Amplitude Modulation |
| SDMA | Space Division Multiple Access |
| SINR | Signal-to-Interference-plus-Noise Ratio |

| | |
|---|---|
| SISO | Single-Input Single-Output |
| SNR | Signal-to-Noise Ratio |
| STC | Space-Time Code |
| TCM | Trellis Coded Modulation |
| TR-STBC | Time-Reversal Space-Time Block Code |
| TDMA | Time Division Multiple Access |
| VA | Viterbi Algorithm |
| WER | Word Error Rate |

# Notations

Here is a list of the main notations and symbols used in this document. We have tried to keep consistent notations throughout the document, but some symbols have different definitions depending on when they occur in the text.

### General Notations

| | |
|---|---|
| $\mathbb{C}$ | Set of complex numbers |
| $\mathbb{R}$ | Set of real numbers |
| $x$ | Scalar variable |
| $\mathbf{x}$ | Vector variable |
| $\mathbf{X}$ | Matrix variable |
| $(\cdot)^*$ | Complex conjugate operator |
| $(\cdot)^T$ | Transpose operator |
| $(\cdot)^H$ | Hermitian transpose operator |
| $\mathrm{tr}(\cdot)$ | Trace of a matrix |
| $M$ | Number of transmit antennas |
| $N_r$ | Number of receive antennas |

### Chapter2 : Transmission without CSIT

| | |
|---|---|
| $E_b$ | Energy per information bit |
| $E_b/N_0$ | Signal to noise ratio per information bit |
| $N_0$ | One sided noise power spectral density of the AWGN |
| SNR | Signal to Noise ratio per receive antenna |
| $k$ | Number of input symbols in a STBC |
| $D$ | Block interleaver depth |
| $T$ | Time dimension of a STBC |
| $N$ | Block size (in symbol) |

| | |
|---|---|
| $L$ | Channel impulse response length (in symbol) |
| $\mathcal{X}$ | Constellation size |
| $P$ | Number of separable paths |
| $R_{\text{tcm}}$ | TCM code rate (in bit/symbol) |
| $R_{\text{stbc}}$ | STBC rate (in symbol/channel use) |
| $\eta$ | Spectral efficiency (in bit/channel use) |

**Chapter 3: MIMO broadcast channel with SNR Feedback**

| | |
|---|---|
| $d$ | Delay in a feedback link (in slot) |
| $K$ | Number of user terminals |
| $B$ | Number of beams (streams) |
| $T$ | Number of channel users per slot |
| $\sigma_e^2$ | Prediction MMSE |
| $v$ | Mobile speed (in km/h) |
| $\gamma$ | Transmit Signal to Noise Ratio |
| $L$ | Order of a Gauss-Markov process |
| $\mathsf{SINR}_{k,j}$ | Signal to Interference plus Noise ratio of user $k$ in beam $j$ |

**Chapter 4: MIMO broadcast channel with Analog Feedback**

| | |
|---|---|
| $N_0^{\text{up}}$ | One sided noise power spectral density of the uplink AWGN |
| $N_0^{\text{down}}$ | One sided noise power spectral density of the downlink AWGN |
| $\mathsf{SINR}_k$ | Signal to Interference plus Noise ratio of user $k$ |

# Chapter 1

# Introduction

## 1.1 Overview

In third, fourth generation (3G, 4G) wireless data-access networks, some applications supported by such networks will require a high data rate downlink for delay-tolerant packet communications. To provide such downlink, there are two systems currently commercialized : High Speed Downlink Packet Access (HSDPA)[1] for WCDMA proposed by 3GPP and High Data Rate (HDR) system [2] also known as cdma2000 1xEvolution Data Optimized (1xEV-DO) [3] defined in the 3GPP2 cdma 2000 IS-856 [4]. HSDPA Release 5 offers a peak data rate around 10Mbit/s over a signal bandwidth of 5 MHz while 1xEV-DO Rev.0 offers a maximum date rate of 2.4 Mbits/s over a signal bandwidth of 1.25 MHz[5]. In order to further increase the system capacity and enhance the quality of services, Multi-Input Multi-Output (MIMO) technology is envisioned in Release 6 of HSDPA (peak data rate up to 30Mbits/s) as well as in 1xEV-DO Rev.A (peak data rate up to 3.1 Mbits/s).

**MIMO transmission without CSIT** In fact, MIMO technique has been considered as a promising solution that can offer both *diversity gain* and *spatial multiplexing gain* for a point-to-point communication systems [6, 7, 8, 9].

The diversity gain is a straightforward consequence of independent physical channels created by multiple transmit and/or receive antennas under rich scattering environment. The diversity gain makes the channel variation small and thus improves the reliability of wireless channel. Under $M \times N_r$ frequency-flat MIMO fading channel where $M$ and $N_r$ denotes the number of transmit and receive antennas respectively, block error probability decreases as $O(\mathsf{SNR}^{-MN_r})$ for high signal-to-noise ratio (SNR), $MN_r$ times faster than in a Single-Input Single-Output (SISO) system. In addition, MIMO channels offer an extra spatial dimension for communication and this additional degree of freedom can be exploited by transmitting up to $\min\{M, N_r\}$ independent data streams in parallel [8]. The spatial multiplexing gain allows for increasing capacity, thus achieving much higher spectral efficiency.

In typical wireless cellular systems user terminals are miniaturized handsets and cannot host more than a single antenna. On the other hand, basestations can be easily equipped with multiple antennas. This motivated the researchers in the early 1990' to reproduce receive diversity gain by using multiple antennas at the transmitter to achieve transmit diversity [9]. *Space-time coding* (STC) [10, 11, 9], coding schemes whose codewords are transmitted across the time and space dimensions that take potential advantage of MIMO/Multi-Input Single-Output (MISO) diversity gain and/or MIMO spatial multiplexing gain, has received considerable attention recently [9, 7, 6]. The attractive feature of STC is that it achieves transmit diversity without *Channel State Information* at transmitter (CSIT). In the absence of reliable CSIT, the decoding error probability for good codes in high SNR is dominated by so-called (*information outage* event) [12]: the event that the transmitted rate falls below the instantaneous mutual information of the fading channel. In such condition, the improvement of error probability by STC with respect to SISO systems is very large.

Chapter 2 of the thesis focuses on the use of multiple transmit antennas to increase the reliability of a point-to-point downlink without CSIT. For pedestrian users in a urban environment, the propagation channel is typically slowly-fading and frequency-selective. For single-carrier transmission as used in current 3G standards [13, 14], frequency-selectivity generates intersymbol interference (ISI). Then, the STC that achieves maximum diversity gain over multipath fading MIMO channel is required. The design of STC for single-carrier transmission over frequency-selective MIMO channels exists in literature [15, 16, 17, 18, 19]. Maximum-likelihood (ML) decoding in MIMO frequency-selective channels is generally too complex for practical channel

memory and modulation constellation size. Hence, research has focused on suboptimal low complexity schemes such as trellis coding and bit-interleaved coded modulation (BICM) with turbo equalization in [16, 19] or the trellis STC with turbo equalization in [20]. Unfortunately, all those schemes suffer from the exponential complexity in the channel memory and in the constellation size, thus can not be applicable in practice. Therefore, we develop a scheme that

- achieves the maximum diversity gain over multipath fading channels.
- can be decoded with lower complexity than the existing schemes.
- offers different spectral efficiencies without increasing the decoding complexity.

**MIMO broadcast channel with CSIT** For a point-to-multipoint downlink system where the base-station communicates with multiple users, accurate and timely CSIT plays a very important role for total system throughput. The downlink of a single cell system is modeled as a fading Gaussian broadcast channel, whose capacity region has been completely characterized under different assumptions in several papers (e.g., [21, 22]). In particular, it is known that, under fading ergodicity, when the base station is equipped with a single antenna and has perfect CSI, the average throughput (long-term average sum rate) is maximized by serving the user with the largest fading coefficient at each time instant (e.g., [23]). Compared to point-to-point communication, the multiuser setting provides more opportunity to exploit because of an additional degree of freedoms, users [8]. This additional choice of which user to transmit to allows the system to benefit from another dimension of diversity called "*multiuser diversity*" [24, 25, 23, 26]. Motivated by the result of [23], HSDPA[1] and the 1xEV-DO [2, 3] consider multiuser diversity based scheduling such as the max C/I scheduling and proportional fair scheduling[23]. The max-C/I scheduling assigns the user with the largest signal-to-interference ratio (SINR) at each slot. The proportional fair scheduling transmits to a user when its channel gain is near its peak within some delay constraint [8]. In these scheduling schemes, the users are assumed to measure the received SNR and request the corresponding quantized rate over the feedback link [2].

When the base station is equipped with $M > 1$ antennas, the single-cell downlink falls in the class of *vector Gaussian broadcast channels*, whose

capacity region with perfect CSIT has been fully characterized in [27] and references therein. In particular, for a system with $M$ transmit antennas and $K \geq M$ users, a multiplexing gain of $M$ can be achieved with full knowledge of $K$ users' channels at the transmitter. As a result, the average throughput scales as $M \log \mathsf{SNR}$ for high SNR and $M$ users can be served simultaneously on each slot. In [28, 29, 30], it has been shown that the sum-rate capacity of a MIMO broadcast channel can be achieved by dirty paper coding (DPC)[31]. Since DPC requires the transmitter to have perfect CSI of all users, it is difficult to implement in practice especially when the number of users is large. Hence, various practical precoding strategies [32] have been proposed for a MIMO broadcast channel including linear beamforming [33, 34, 35, 36, 37], non-linear precoding [38, 39]. In [40, 32, 37], the linear beamforming scheme has been shown to asymptotically approach the DPC performance, i.e. the sum rate scales as $M \log \mathsf{SNR}$ when $\mathsf{SNR} \to \infty$ for fixed $K >> M$, and scales as $M \log \log K$ when $K \to \infty$ and for fixed $\mathsf{SNR}$. These are precisely the scaling laws achieved by the DPC. Moreover, the recent results [35, 36] show that the sum-rate scaling $M \log \log K$ is achieved even for a practical number of users. A much simpler strategy based on TDMA and multiuser diversity scheduling consists of the so called *opportunistic beamforming* proposed in [23]. In opportunistic beamforming, the multiple antennas are used to generate a random beam inducing an artificial fading that varies slowly enough to be measured and fed back by the users but rapidly enough to make the scheduling algorithm share the channel fairly among the users [1]. A spatial-multiplexing version of opportunistic beamforming is proposed and analyzed in [26], where $M$ mutually orthogonal random beams are simultaneously used to serve the best $M$ users at each time. It is shown that for $K \gg M$ and assuming perfect SNR instantaneous feedback, the same multiplexing gain of $M$ as for the DPC is achievable.

Multiuser diversity based scheduling can also be applied on top of STC. It is natural to investigate the relative merit of transmit diversity achieved by STC and multiuser diversity achieved by opportunistic beamforming because both STC and opportunistic-based scheduling are currently considered for standardization in evolutionary 3G systems such as [1, 2, 3] as mentioned before. Partial answers to the question have been provided, for example, in [41, 42, 43]. These works, as well as many others, indicate essentially that

[1] Here, the fairness means that every user in the system receives equal average throughput over some time scale.

transmit diversity always *decreases* the throughput of a downlink system under opportunistic scheduling. In fact, multiuser diversity benefits from highly variable user channel conditions. This effect is amplified by random beamforming that induces larger and faster fluctuation of the fading channel. On the contrary, transmit diversity improves the reliability of each user's link individually and tends to equalize the channel of each user as the number of antennas increases (the "channel hardening" effect [42]). Therefore, the total system throughput is decreased due to small fluctuation between users. The assumptions underlying these results are: 1) channel errors never occur, i.e., once a user is scheduled and is allocated a given (channel dependent) rate, the message will be successfully received with probability 1; 2) transmission queues have infinite buffers and all users are backlogged. In Chapter 3, we take a deeper look into the interaction between transmit diversity achieved by STC and multiuser diversity achieved by opportunistic beamforming. Here we restrict ourselves to a simple SNR feedback as currently considered in the systems such as HDR[2] or HSDPA[1]. We take into account the fundamental aspects that were neglected in works such as [41, 42, 23, 26]: random packet arrivals with finite transmission buffers and time-varying fading channels with a delay in the feedback link. Under these assumptions, the questions that we pose are:

- What is a meaningful notion of fairness under random packet arrivals ?
- Is STC always harmful ?
- Can opportunistic $M$-beamforming exploit multiplexing gain ?
- How does a time-varying channel impact the performance of STC and opportunistic-beamforming under non-perfect CSIT ?

In Chapter 4, we consider a more informative type of feedback that enables a smarter use of the multiple antennas at the transmitter in MIMO broadcast channels. In particular, we consider a distortion-wise optimal "analog feedback" scheme [44, 45]. Namely, each user terminal sends back its estimated channel vector without quantizing and coding over the feedback link. We restrict our investigation to linear beamforming strategies because it is simple and it can be suitably generalized to the non-perfect CSIT case. Moreover, linear beamforming has been shown to provide a fraction of optimal DPC scheme performance [40, 32, 37, 35, 36] as mentioned before. There

exist several methods for linear beamforming and user selection over MIMO broadcast channel in the literature [36, 35, 46, 37]. The optimality criteria of the works [36, 35, 46, 37] is sum-rate maximization since the infinite backlog for transmission is assumed in those works. In fact, all these works [36, 35, 46, 37] are based on the underlying assumptions given above : 1) channel errors never occur; 2) transmission queues have infinite buffers and all users are backlogged. By considering again the random packet arrivals with finite transmission buffers and time-varying fading channels with a delay in the feedback link, we will examine the following questions:

- What is the optimal scheduling under linear beamforming strategy ?
- Can linear beamforming outperform STC and opportunistic beamforming ?
- How does a time-varying channel impact the performance of linear beamforming under non-perfect CSIT ?

## 1.2 Contributions

The contribution of Chapter 2 is the design of a low-complexity STBC for a single-carrier MISO wireless system [47, 48]. The proposed STBC concatenates two well-known techniques, trellis-coded modulation (TCM) and time-reversal orthogonal space-time block coding (TR-STBC). The decoder is based on reduced-state joint equalization and decoding that operates directly on the TCM trellis without trellis state expansion. Hence, the decoder complexity is independent of the channel memory and of the constellation size. This makes our scheme applicable in practice even for large signal constellations and channel impulse response length as specified in 3G standards, while the schemes proposed in [16, 17, 19, 20] are not. We show that in the limit of large block size our proposed scheme achieves the full diversity offered by the channel. The numerical examples show that our scheme achieves similar/superior performance with respect to the best previously proposed schemes at significantly lower complexity. A significant advantage of the proposed TCM-TR-STBC scheme is that TCM easily implements adaptive modulation by adding uncoded bits (i.e., parallel transitions in the TCM trellis) and by expanding correspondingly the signal constellation [49]. This fact has particular relevance in the implementation of high-speed downlink

schemes based on dynamic scheduling, where adaptive modulation is required [1]. Therefore, it represents an attractive solution to implement transmit diversity in high-speed downlink system such as HSDPA or HDR.

The main contribution of Chapter 3 and Chapter 4 is the investigation of the practical uses of multiple transmit antennas in MIMO broadcast channels [50, 51, 52, 53, 54]. We especially take into account the following fundamental and realistic aspects missed in the existing works : random packet arrivals, correlated block-fading channels and non-perfect CSIT. Under random packet arrival, the notion of fairness is replaced by the notion of *stability* [55, 56]. In other words, achieving any point in the system stability region may be considered as the single most important goal of a downlink resource-allocation policy.

The first contribution of Chapter 3 is the formulation of the stability region of a SDMA/TDMA downlink system under non-perfect CSIT and the derivation of the max-stability policy. The formulation allows us to make a fair comparison between transmit diversity achieved by STC and multiuser diversity achieved by opportunistic beamforming under the max-stability policy. The realistic assumption of non-perfect CSIT yields a non-trivial tradeoff between transmit diversity and multiuser diversity. This tradeoff is missed by the analysis of [41, 42, 43, 23, 26, 57, 5, 58, 59] that neglects the fact that decoding errors may occur with non-zero probability. In our more realistic setting, the ability of accurately predicting the channel SNR clearly emerges as one of the main limiting factors for opportunistic beamforming.

The second contribution of Chapter 3 is the proposition of the improved opportunistic beamforming scheme where the users are synchronized and have a priori knowledge of the random beamforming matrices. This decouples the problem of channel prediction from the speed of variation of the beamformer. Then, we demonstrate that varying random beams every slot minimizes the average delay of the improved opportunistic beamforming.

The third contribution of Chapter 3 is the evaluation of the relative merit of STC and opportunistic beamforming over the time-varying channel under non-perfect CSIT due to a feedback delay. Our comparison suggests the following choice : opportunistic beamforming is better for a slow fading channel because it can decrease the average delay by letting the channel vary i.i.d. from one slot to another, while STC is interesting for a fast fading channel because it achieves the larger outage rate thanks to channel-hardening effect.

The contribution of Chapter 4 [54] is the development of linear-beamforming based scheduling to a queued downlink system under the non-perfect CSIT

assumption. The first contribution of Chapter 4 is the derivation of max-stability stationary policies under linear beamforming strategies that stabilize the system whenever the arrival rate is inside the system stability region. Unfortunately such policies are difficult to implement in practice.

The second contribution of Chapter 4 is the proposition of a novel heuristic method for user selection and linear beamforming that maximizes the weighted sum rate over a beamforming matrix. The proposed scheme can be seen as a practical version of the max-stability stationary policies. Interestingly, numerical simulations show that the proposed scheme stabilizes the system until the arrival rate achieves the boundary of the stability region.

The third contribution of Chapter 4 is the evaluation of the relative merit of our linear beamforming scheme, STC and opportunistic $M$-beamforming over the time-varying channel. At a given feedback delay, it is found that our proposed scheme achieves the best average delay over any range of mobile speed. For slow fading channels, our proposed scheme can exploit multiplexing gain and power gain to balance the transmission queues, while for fast fading channels, it essentially allocates the whole power to the user with the largest weighted rate at each slot by operating in TDMA mode. We can conclude that our scheme offers an attractive solution for a next generation of high-speed downlink system exploiting MIMO technology such as HSDPA Release 6 or 1xEV-DO Rev.A.

# Chapter 2

# MIMO Transmission without CSIT

*In this chapter, we present the use of multiple antennas to improve the reliability of a point-to-point wireless channel via space-time coding (STC). STC has been considered as an attractive solution that achieves the transmit diversity with no Channel State Information at a transmitter (CSIT). For a single-carrier transmission over frequency selective multi-input single-output (MISO) fading channels, we propose a STC that concatenates trellis-coded modulation (TCM) and time-reversal orthogonal space-time block coding (TR-STBC). The decoder is based on reduced-state joint equalization and decoding that operates on the TCM trellis without trellis state expansion. Hence, the decoder complexity is independent of the channel memory and of the constellation size. We show that, in the limit of large block length, the proposed scheme with reduced-state joint equalization and decoding can achieve the full diversity offered by the MISO multipath channel. Numerical examples show that our proposed scheme offers similar/superior performance with respect to the best previously proposed schemes at significantly lower complexity. Thus, it represents an attractive solution to implement transmit diversity in high-speed TDM-based downlink of 3G systems, such as EDGE and UMTS.*

# 2.1 Introduction

## 2.1.1 Motivation

In classical wireless cellular systems user terminals are miniaturized handsets and typically cannot host more than a single antenna. On the other hand, base-stations can be easily equipped with multiple antennas. Then we are in the presence of a multi-input single-output (MISO) channel. For pedestrian users in a urban environment, the propagation channel is typically slow and frequency-selective fading. For single-carrier transmission, as used in current 3G standards [13, 14], frequency-selectivity generates inter-symbol interference (ISI). In systems that do not make use of spread-spectrum waveforms, such as GPRS and EDGE [14] or certain modes of wideband CDMA [13] using very small spreading factors, ISI must be handled by linear/decision-feedback equalization or maximum-likelihood sequence detection [60].

Due to the slowly-varying nature of the fading channel, a codeword spans a limited number of fading degrees of freedom. In the absence of reliable channel state information at the transmitter, the word-error probability (WER) is dominated by the so-called information outage event: namely, the event that the transmitted coding rate is above the mutual information of the channel realization spanned by the transmitted codeword [61]. In such conditions, the WER can be greatly improved by using space-time codes (STCs), i.e., coding schemes whose codewords are transmitted across a time and space dimension introduced by the multiple transmit antennas [11].

In a frequency-selective MISO Rayleigh fading channel with $M$ transmit antennas and $P$ independent (separable) multipath components, it is immediate to show that the best WER behavior achievable by STC for high SNR is $O(\mathsf{SNR}^{-d_{\max}})$, where $d_{\max} \stackrel{\Delta}{=} MP$ is the maximum achievable *diversity order* of the channel, equal to the number of fading degrees of freedom. We hasten to say that in this Chapter we focus on MISO channels and on STC design for achieving maximum diversity. Obviously, the STC scheme proposed in this Chapter can be trivially applied to the case of multiple receiver antennas (MIMO channel). However, for $N_r > 1$ antennas at the receiver our scheme (as well as the competitor schemes mentioned in the following) would not be able to exploit the spatial multiplexing capability of the channel, i.e., the ability of creating up to $\min\{M, N_r\}$ parallel channels between transmitter and receiver, thus achieving much higher spectral efficiency. The optimal tradeoff between the achievable spatial *multiplexing gain* and *diversity gain*

in frequency-flat MIMO channels was investigated in [62] and the analysis has been recently extended to the frequency-selective case in [63].

The design of STC for single-carrier transmission over frequency-selective MIMO channels exists in literature [15, 16, 17, 18, 19]. Maximum-likelihood (ML) decoding in MIMO frequency-selective channels is generally too complex for practical channel memory and modulation constellation size. Hence, research has focused on suboptimal low complexity schemes such as trellis coding and bit-interleaved coded modulation (BICM) with turbo equalization in [16, 19] or the trellis STC with turbo equalization in [20]. Unfortunately, all those schemes suffer from the exponential complexity in the channel memory and in the constellation size, thus can not be applicable in practice. Therefore, Therefore, we develop a scheme that

- achieves the maximum diversity gain over multipath fading channels.
- can be decoded with lower complexity than the existing schemes.
- offers different spectral efficiencies without increasing the decoding complexity.

### 2.1.2 Contribution

We consider the concatenation of Time reversal orthogonal space-time block code (TR-STBC) with an outer trellis coded modulation (TCM) [49]. TR-STBC [15, 9] converts the MISO channel into a standard single-input single-output (SISO) channel with ISI, to which conventional equalization/sequence detection techniques, or turbo-equalization, can be applied. At the receiver, we apply reduced-state sequence detection based on joint minimum mean-square-error (MMSE) decision-feedback equalization (DFE) and decoding (notice that sequence detection for TR-STBC without outer coding has been considered in [64, 65]). The decisions for the MMSE-DFE are found on the surviving paths of the Viterbi decoder acting on the trellis of the TCM code. Since the joint equalization and decoding scheme works on the trellis of the original TCM code, without trellis state expansion due to the ISI channel, the receiver complexity is independent of the channel length and of the constellation size. This makes our scheme applicable in practice even for large signal constellations and channel impulse response length as specified in 3G generation standards, while the schemes proposed in [16, 17, 19] are not.

We show that, in the limit of large block length, the TCM-TR-STBC scheme with reduced-state joint equalization and decoding can achieve the full diversity offered by the MISO multipath channel. Remarkably, simulations show that the proposed scheme achieves full diversity for short (practical) block length and simple TCM codes.

A significant advantage of the proposed TCM-TR-STBC scheme is that TCM easily implements adaptive modulation by adding uncoded bits (i.e., parallel transitions in the TCM trellis) and by expanding correspondingly the signal constellation [49]. Since the constellation size has no impact on the decoder complexity, variable-rate (adaptive) modulation can be easily implemented. This fact has particular relevance in the implementation of high-speed downlink schemes based on dynamic scheduling, where adaptive modulation is required [1]. Remarkably, simulations show that the TCM-TR-STBC scheme achieves WER performance at least as good as (if not better than) previously proposed schemes[16, 17, 19], that are more complex and less flexible in terms of variable-rate coding implementation. Therefore, it represents an attractive solution to implement transmit diversity in high-speed TDM-based downlink of 3G systems, such as EDGE and UMTS.

### 2.1.3 Related works

The design of STC for single-carrier transmission over frequency-selective MISO channels has been investigated in a number of recent contributions [15, 16, 17, 18]. We may group these approaches into two classes. The first approach is based on mitigating ISI by some MISO equalization technique, and then designing a space-time code/decoder for the resulting flat fading channel. For example, the use of a linear MMSE equalizer combined with Alamouti's space-time block code [10] has been investigated in [18]. However, this scheme does not achieve in general the maximum diversity order offered by the channel. The second approach is based on designing the STC by taking into account the ISI channel and then performing joint equalization and decoding. For example, trellis coding and BICM with turbo equalization have been proposed in [16, 19]. Turbo-equalization schemes need a soft-in soft-out MAP decoder for the ISI channel, whose complexity is exponential in the channel impulse response length and in the constellation size. For example, MAP symbol-by-symbol detection implemented by the BCJR algorithm [66] runs on a trellis with $|\mathcal{X}|^{L-1}$ states, where $|\mathcal{X}|$ denotes the size of the transmitted signal constellation $\mathcal{X} \subset \mathbb{C}$, and $L$ denotes the channel

impulse response length (expressed in symbol intervals).

## 2.2 System model.

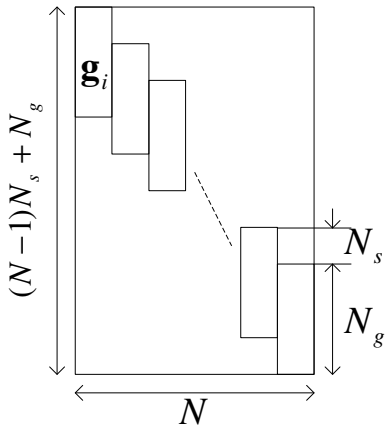


Figure 2.1: Structure of the matrix $\mathcal{H}(\mathbf{g}_i)$

The channel from the $i$-th transmit antenna to the receive antenna is formed by a pulse-shaping transmit filter (e.g., a root-raised cosine pulse [60]), a multipath fading channel with $P$ separable paths, an ideal low-pass filter with bandwidth $[-N_s/(2T_s), N_s/(2T_s)]$ where $N_s \geq 2$ is an integer and $T_s$ is the symbol interval, and a sampler taking $N_s$ samples per symbol. We assume that the fading channels are random but constant in time for a large number of symbol intervals (quasi-static assumption). We also assume that the overall channel impulse response spans at most $L$ symbol intervals, corresponding to $N_g \triangleq LN_s$ receiver samples. Let $(s_i[0], \ldots, s_i[N-1], 0, \ldots, 0)$ denote the sequence of symbols transmitted over antenna $i$, where we add a tail of $L-1$ zeros in order to avoid inter-block interference. The discrete-time complex baseband equivalent MISO channel model can be written in vector form as

$$\mathbf{r} = \sum_{i=1}^{M} \mathcal{H}(\mathbf{g}_i)\mathbf{s}_i + \mathbf{w} \tag{2.1}$$

where $\mathbf{r} \in \mathbb{C}^{N_s(N-1)+N_g}$, $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, N_0\mathbf{I})$ is the complex circularly-symmetric Additive White Gaussian Noise (AWGN), $\mathbf{s}_i = (s_i[0], \ldots, s_i[N-1])^{\mathsf{T}}$, and

$\mathcal{H}(\mathbf{g}_i) \in \mathbb{C}^{(N_s(N-1)+N_g)\times N}$ is the convolution matrix obtained from the overall sampled channel impulse response $\mathbf{g}_i \in \mathbb{C}^{N_g}$ as follows: the $n$-th column of $\mathcal{H}(\mathbf{g}_i)$ is given by

$$(\underbrace{0,\ldots,0}_{nNs}, g_i[0],\ldots,g_i[N_g-1], \underbrace{0,\ldots,0}_{N_s(N-n-1)})^{\mathsf{T}} \tag{2.2}$$

for $n = 0,\ldots,N-1$. The structure of the matrix $\mathcal{H}(\mathbf{g}_i)$ is illustrated in Fig. 2.1.

## 2.3 Preliminaries

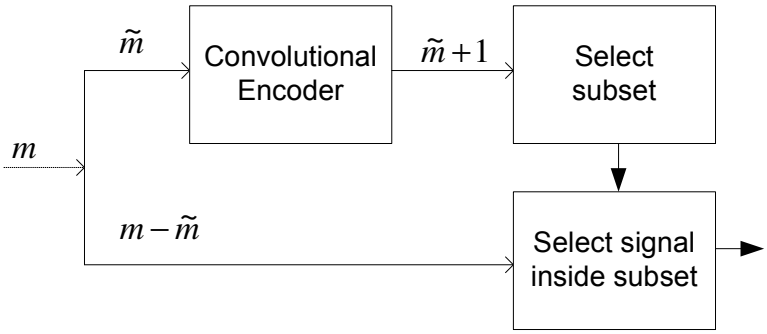


Figure 2.2: Ungerboeck TCM encoder structure

### 2.3.1 Trellis Coded Modulation (TCM)

Trellis coded modulation (TCM) is the standard combined coding and modulation scheme for bandwidth limited channels [67]. The basic idea consists of transmitting $m$ information bits per symbol using a modulator with a set of $\mathcal{X}$ constellations where we have $\mathcal{X} > 2^m$. In [68], Ungerboeck showed that a constellation of size $\mathcal{X} = 2^{m+1}$ concatenated with a binary convolutional code of rate $\frac{m}{m+1}$ suffices to achieve a considerable fraction of the available coding gain. The resulting rate of the TCM is given by $R_{\text{tcm}} = m$ in bit/symbol. Figure 2.2 shows the block diagram of the Ungerboeck TCM encoder where a modulated symbol is obtained by two steps. In the first step, $\tilde{m} \leq m$ information bits are expanded by a rate $\frac{\tilde{m}}{\tilde{m}+1}$ binary convolutional encoder into

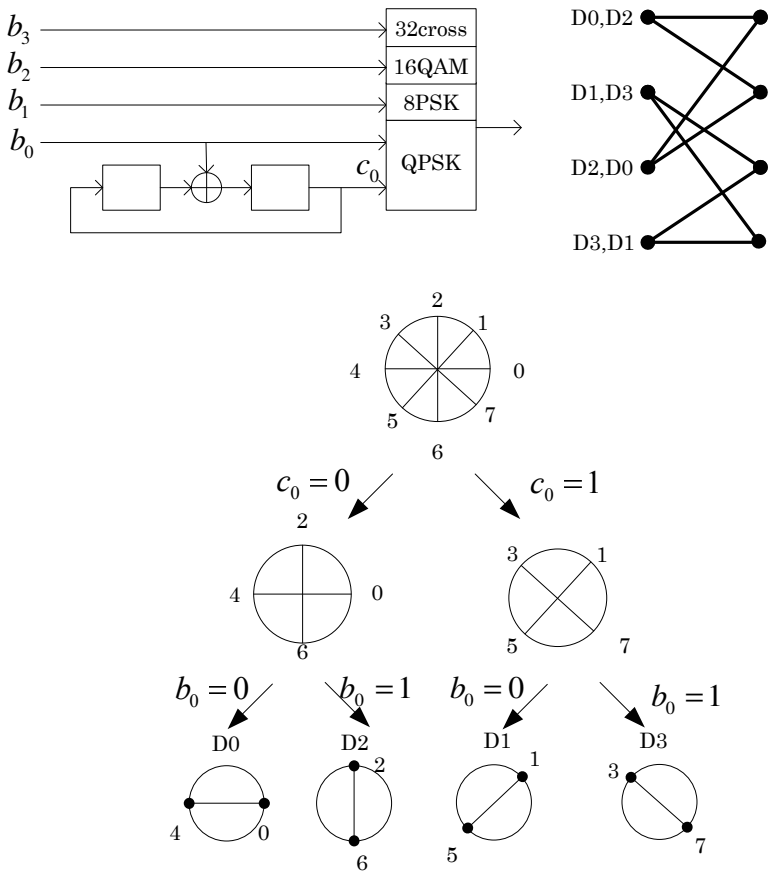


Figure 2.3: 4-state Ungerboeck TCM encoder

$\widetilde{m}+1$ coded bits. These coded bits are used to select one of $2^{\widetilde{m}+1}$ subsets of a redundant $\mathcal{X}$-ary signal set. Then in the second step, the remaining $m-\widetilde{m}$ uncoded bits determine which of the signal in this subset is to be transmitted [49]. The uncoded $m-\widetilde{m}$ bits produce parallel transitions in the TCM trellis: an edge in the trellis diagram is associated with $2^{m-\widetilde{m}}$ signals. In other words, the same encoder of rate $\frac{\widetilde{m}}{\widetilde{m}+1}$ can be used for any constellation of size $\mathcal{X} = 2^{m+1} \geq 2^{\widetilde{m}+1}$.

Figure 2.3 shows the Ungerboeck's 4-state TCM encoder as well as its trellis structure to generate QPSK, 8PSK, 16QAM, and 32cross symbols for $m = 1, 2, 3$, and 4 respectively ($\widetilde{m} = 1$). In addition, Figure 2.3 shows the two steps to generate 8PSK symbol : the two coded bits $b_0, c_0$ are used to

choose one of the four subsets denoted by $D_0, D_1, D_2, D_3$, and then the uncoded bit $b_1$ is used to choose the symbol within a subset. To increase the constellation, it is sufficient to add uncoded bits, i.e. add parallel transitions in trellis while keeping the same 4-state convolutional encoder. This structure is very useful to implement adaptive modulation without increasing the decoder complexity. A further description of TCM can be found in [69, 49].

### 2.3.2 Generalized Orthogonal Designs (GOD)

The theory of generalized orthogonal design (GOD) is related to mathematics [70]. The application to wireless communication was introduced by the famous Alamouti STBC [10] for $M = 2$. The generalization of the Alamouti STBC for $M > 2$ was done by [70] and further developed in [71, 72]. Here we provide the formal definition of GOD for flat-fading channel that will be extended to the frequency-selective fading in the following.

A $[T, M, k]$-GOD is defined by a mapping $\mathbf{S} : \mathbb{C}^k \rightarrow \mathbb{C}^{T \times M}$ such that, for all $\mathbf{x} \in \mathbb{C}^k$, the corresponding matrix $\mathbf{S}(\mathbf{x})$ satisfies $\mathbf{S}(\mathbf{x})^{\mathsf{H}}\mathbf{S}(\mathbf{x}) = |\mathbf{x}|^2\mathbf{I}$. Moreover, the elements of $\mathbf{S}(\mathbf{x})$ are linear combinations of elements of $\mathbf{x}$ and of $\mathbf{x}^*$.

Let consider a simple example of the Alamouti's STBC for $[T = 2, M = 2, k = 2]$. Let $\mathbf{h} = [h_1, h_2]^T$ denote 2-input 1-output frequency flat channel. The received signal is given by

$$\mathbf{r} = \mathbf{S}(\mathbf{x})\mathbf{h} + \mathbf{w} \tag{2.3}$$

where $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(0, N_0\mathbf{I})$ is AWGN and the Alamouti's STBC is given explicitly by

$$\mathbf{S}(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix}$$

Since $\mathbf{S}(\mathbf{x})$ has the property that each row contains either elements of $\mathbf{x}$ or elements of $\mathbf{x}^*$, then we have

$$\widetilde{\mathbf{S}(\mathbf{x})\mathbf{h}} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \mathbf{x} \triangleq \mathcal{H}\mathbf{x}$$

where $\widetilde{\ }$ indicates the complex conjugation of the elements containing elements of $\mathbf{x}^*$. Noticing that the corresponding channel matrix $\mathcal{H}$ is orthogonal matrix, the matched filter output is given by

$$\begin{aligned} \mathcal{H}^H\mathbf{r} &= \mathcal{H}^H\mathcal{H}\mathbf{x} + \mathbf{z} \\ &= |\mathbf{h}|^2\mathbf{x} + \mathbf{z} \end{aligned}$$

where $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(0, N_0 \mathcal{H}^H \mathcal{H})$ is colored noise. The important remark is that the resulting matched filter output is decoupled and the channel is coherently combined. This allows the receiver to perform maximum-likelihood detection on the decoupled symbols while achieving the diversity gain.

### 2.3.3 Time Reversal-Space Time Block Code (TR-STBC)

TR-STBC [15, 9] is a clever extension of orthogonal space-time block codes based on GOD [71, 72, 70] to the frequency-selective channel. As we shall briefly review in the following, the TR-STBC turns a frequency-selective MISO [1] into a standard SISO channel with ISI, by simple linear processing given by matched filtering and combining.

Let $\mathbf{S}$ be a $[T, M, k]$-GOD with the following additional property: A1) the row index $\{1, \ldots, T\}$ set can be partitioned into two subsets, $\mathcal{T}_1$ and $\mathcal{T}_2$, such that all elements of the $t$-th rows with $t \in \mathcal{T}_1$ are given by $a_{t,i} x_{\pi(t,i)}$, for $i = 1, \ldots, M$, and all elements of the $t$-th rows with $t \in \mathcal{T}_2$ are given by $a_{t,i} x^*_{\pi(t,i)}$, for $i = 1, \ldots, M$, where $a_{t,i}$ are given complex coefficients and $\pi : \{1, \ldots, T\} \times \{1, \ldots, M\} \to \{1, \ldots, k\}$ is a given indexing function.

Given a $[T, M, k]$-GOD $\mathbf{S}$ satisfying property A1) and two integers $N \geq 1$ and $L \geq 1$, we define the associated TR-STBC $\mathbf{T}$ with parameters $[T, M, k, N, L]$ as the mapping $\mathbb{C}^{N \times k} \to \mathbb{C}^{T(N+L-1) \times M}$ that maps the $k$ vectors $\{\mathbf{x}_j \in \mathbb{C}^N : j = 1, \ldots, k\}$ into the matrix $\mathbf{T}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ defined as follows. For all $t \in \mathcal{T}_1$, replace the $i$-th element of $\mathbf{S}$ by the vector $a_{t,i} \mathbf{x}_{\pi(t,i)}$ followed by $L-1$ zeros. For all $t \in \mathcal{T}_2$, replace the $i$-th element of $\mathbf{S}$ by the vector $a_{t,i} (\circ \mathbf{x}_{\pi(t,i)})$ followed by $L-1$ zeros, where the complex conjugate time-reversal operator $\circ$ is defined by

$$\circ (v[0], \ldots, v[N-1])^{\mathsf{T}} = (v^*[N-1], \ldots, v^*[0])^{\mathsf{T}}$$

The time-reversal operator satisfies the following elementary properties: B1) Let $\mathcal{H}(\mathbf{g})$ be a convolution matrix as defined by (2.2) and $\mathbf{s} \in \mathbb{C}^N$. Then, $\circ(\mathcal{H}(\mathbf{g}) \circ \mathbf{s}) = \mathcal{H}(\circ \mathbf{g})\mathbf{s}$. B2) Let $\mathbf{g}$ and $\mathbf{h}$ be two impulse responses of length $N_g$, then $\mathcal{H}(\mathbf{g})^{\mathsf{H}} \mathcal{H}(\mathbf{h}) = \mathcal{H}(\circ \mathbf{h})^{\mathsf{H}} \mathcal{H}(\circ \mathbf{g})$.

In order to transmit the $k$ blocks of $N$ symbols each over the MISO channel defined by (2.1) by using a TR-STBC scheme with parameters $[T, M, k, N, L]$,

[1] Extension to MIMO is straightforward, but as anticipated in 2.1, it is less relevant due to the significant spectral efficiency loss of orthogonal STBCs in MIMO channels.

the columns of $\mathbf{T}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ are transmitted in parallel, over the $M$ antennas, in $T(N+L-1)$ symbol intervals. Due to the insertion of the tails of $L-1$ zeros, the received signal can be partitioned into $T$ blocks of $N_s(N-1)+N_g$ samples each, without inter-block interference. If $t \in \mathcal{T}_1$, the $t$-th block takes on the form

$$\mathbf{r}_t = \sum_{i=1}^{M} a_{t,i} \mathcal{H}(\mathbf{g}_i) \mathbf{x}_{\pi(t,i)} + \mathbf{w}_t \tag{2.4}$$

If $t \in \mathcal{T}_2$, by using property B1) and the fact that when $\mathbf{w}_t \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, N_0\mathbf{I})$ then $\circ\mathbf{w}_t$ and $\mathbf{w}_t$ are identically distributed, the $t$-th block takes on the form

$$\circ\mathbf{r}_t = \sum_{i=1}^{M} a_{t,i}^{*} \mathcal{H}(\circ\mathbf{g}_i) \mathbf{x}_{\pi(t,i)} + \mathbf{w}_t \tag{2.5}$$

We form the observation vector $\widetilde{\mathbf{r}}$ by stacking blocks $\{\mathbf{r}_t : t \in \mathcal{T}_1\}$ and $\{\circ\mathbf{r}_t : t \in \mathcal{T}_2\}$. The resulting vector can be written as

$$\widetilde{\mathbf{r}} = \mathcal{Q}(\mathbf{g}_1, \ldots, \mathbf{g}_M) \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_k \end{bmatrix} + \mathbf{w} \tag{2.6}$$

where $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, N_0\mathbf{I})$. The matrix $\mathcal{Q}(\mathbf{g}_1, \ldots, \mathbf{g}_M)$ has dimensions $T(N_s(N-1)+N_g) \times Nk$ and it is formed by $Tk$ blocks of size $(N_s(N-1)+N_g) \times N$. The $(t,j)$-th block is given by $a_{t,i}\mathcal{H}(\mathbf{g}_i)$ for $t \in \mathcal{T}_1$ and $j = \pi(t,i)$, or by $a_{t,i}^{*}\mathcal{H}(\circ\mathbf{g}_i)$ for $t \in \mathcal{T}_2$ and $j = \pi(t,i)$. From the orthogonality property of the underlying GOD $\mathbf{S}$ and from property B2) it is straightforward to show that

$$\mathcal{Q}(\mathbf{g}_1, \ldots, \mathbf{g}_M)^{\mathsf{H}} \mathcal{Q}(\mathbf{g}_1, \ldots, \mathbf{g}_M) = \begin{bmatrix} \mathbf{\Gamma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Gamma} \end{bmatrix} \tag{2.7}$$

where we define the combined total channel response

$$\begin{aligned}\boldsymbol{\Gamma} &\triangleq \sum_{i=1}^{M} \mathcal{H}(\mathbf{g}_i)^{\mathsf{H}}\mathcal{H}(\mathbf{g}_i) \qquad (2.8)\\ &= \begin{bmatrix} \gamma_0 & \gamma_{-1} & \cdots & \gamma_{-L} & 0 & \cdots & 0 \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots & \gamma_{-L} & \ddots & \vdots \\ \vdots & & \ddots & & & \ddots & 0 \\ \gamma_L & \cdots & \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots & \gamma_{-L} \\ 0 & \ddots & & & \ddots & & \vdots \\ \vdots & \ddots & \gamma_L & \cdots & \gamma_1 & \gamma_0 & \gamma_{-1} \\ 0 & \cdots & 0 & \gamma_L & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}\end{aligned}$$

where $\boldsymbol{\Gamma}$ is a $N \times N$ Hermitian symmetric non-negative definite Toeplitz matrix. Therefore, by passing the received signal $\widetilde{\mathbf{r}}$ through the bank of matched filters for the channel impulse responses $\mathbf{g}_i$ and combining the matched filter outputs (sampled at the symbol rate), the $k$ blocks of transmitted symbols are completely decoupled. The equivalent channel for any of these blocks (we drop the block index from now on for simplicity) is given by

$$\mathbf{y} = \boldsymbol{\Gamma}\mathbf{x} + \mathbf{z} \qquad (2.9)$$

where $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, N_0\boldsymbol{\Gamma})$. The TR-STBC scheme has turned the MISO frequency-selective channel into a standard SISO channel with ISI, and the channel model (2.9) represents the so-called *Sampled Matched-Filter* output of the equivalent SISO channel, in block form. Notice that the noise $\mathbf{z}$ is correlated.

### 2.3.4 MMSE-DFE for ISI channel

Since the TR-STBC converts MISO ISI channel into SISO ISI channel, we review the classical equalizer techniques for ISI channel that can be applied to the equivalent channel given in (2.9). Following the block formulation of [73], we present the block Minimum Mean Square Error-Decision Feedback Equalizer (MMSE-DFE). Consider the same model given in (2.9)

$$\overline{\mathbf{y}} = \boldsymbol{\Gamma}\overline{\mathbf{x}} + \overline{\mathbf{z}} \qquad (2.10)$$

where the ordering of $\overline{\mathbf{y}}, \overline{\mathbf{x}}, \overline{\mathbf{z}}$ is reversed such that $\overline{\mathbf{y}} = (y_{N-1}, \ldots, y_0)^T$. The MMSE-DFE is composed by a feedforward filter $\mathbf{F}$ ($N \times N$ matrix), a strictly

causal feedback filter $\mathbf{B}-\mathbf{I}$, and a threshold detector. $\mathbf{B}$ is upper triangular with unit diagonal elements. Assuming that the past decisions in the block are correct, the decision vector and the error vector are given by

$$\begin{aligned} \mathbf{x}' &= \mathbf{F}\overline{\mathbf{y}} - (\mathbf{B}-\mathbf{I})\overline{\mathbf{x}} && (2.11)\\ \mathbf{e} &= \mathbf{x}' - \overline{\mathbf{x}} = \mathbf{F}\overline{\mathbf{y}} - \mathbf{B}\overline{\mathbf{x}} && (2.12) \end{aligned}$$

We wish to find optimal $\mathbf{F}$ and $\mathbf{B}$ so that MMSE $\mathbb{E}[|\mathbf{e}|^2]$ should be minimized. From the orthogonality principle $\mathbb{E}[e_n y_l^*] = 0\ \forall n, l$, we have

$$\mathbb{E}[\mathbf{e}\overline{\mathbf{y}}^H] = \mathbf{F}\mathbb{E}[\overline{\mathbf{y}\mathbf{y}}^H] - \mathbf{B}\mathbb{E}[\overline{\mathbf{x}\mathbf{y}}^H] = \mathbf{0} \quad (2.13)$$

By substituting $\mathbb{E}[\overline{\mathbf{y}\mathbf{y}}^H] = \mathbf{\Gamma}\mathbf{\Gamma}^H + N_0\mathbf{\Gamma}$ and $\mathbb{E}[\overline{\mathbf{x}\mathbf{y}}^H] = \mathbf{\Gamma}$, we obtain

$$\mathbf{F} = \mathbf{B}(N_0\mathbf{I} + \mathbf{\Gamma})^{-1} \quad (2.14)$$

Now, we would like to compute $\mathbf{B}$. Plugging (2.14) into (2.12), we have the covariance matrix of $\mathbf{e}$ as follows.

$$\begin{aligned} \mathbf{e} &= \mathbf{B}\mathbf{u} && (2.15)\\ \mathbf{u} &= (N_0\mathbf{I} + \mathbf{\Gamma})^{-1}\overline{\mathbf{y}} - \overline{\mathbf{x}} = (N_0\mathbf{I} + \mathbf{\Gamma})^{-1}(\overline{\mathbf{z}} - N_0\overline{\mathbf{x}})\\ \mathbb{E}[\mathbf{u}\mathbf{u}^H] &= N_0(N_0\mathbf{I} + \mathbf{\Gamma})\\ \mathbb{E}[\mathbf{e}\mathbf{e}^H] &= N_0\mathbf{B}(N_0\mathbf{I} + \mathbf{\Gamma})^{-1}\mathbf{B}^H \end{aligned}$$

Using the Cholesky factorization, we can further express the covariance matrix of $\mathbf{e}$.

$$\begin{aligned} N_0\mathbf{I} + \mathbf{\Gamma} &= \mathbf{C}^H\mathbf{\Delta}\mathbf{C} && (2.16)\\ (N_0\mathbf{I} + \mathbf{\Gamma})^{-1} &= \mathbf{A}^H\mathbf{\Delta}^{-1}\mathbf{A} && (2.17) \end{aligned}$$

where $\mathbf{A}, \mathbf{C}^H$ are lower triangular with unit diagonal elements and $\mathbf{\Delta} = \mathrm{diag}(\sigma[N-1], \ldots, \sigma[0])$ is a diagonal matrix with positive real diagonal elements. Then we have,

$$\mathbb{E}[\mathbf{e}\mathbf{e}^H] = N_0\mathbf{B}\mathbf{A}^H\mathbf{\Delta}^{-1}\mathbf{A}\mathbf{B}^H = N_0\mathbf{\Delta}^{-1} \quad (2.18)$$

In order to minimize $\mathbb{E}[|\mathbf{e}|^2]$, $\mathbf{B}$ should decorrelate $\mathbf{u}$, in other words should be a whitening filter. So that $\mathbf{e}$ has uncorrelated components, $\mathbf{B}\mathbf{A}^H$ must be a diagonal matrix with unit elements because both $\mathbf{B}$ and $\mathbf{A}^H$ are monic. Since the Cholesky factorization (2.16) is unique, $\mathbf{B}$ should match $\mathbf{C}$ and we

have $\mathbf{B}\mathbf{A}^H = \mathbf{C}\mathbf{A}^H = \mathbf{I}$. As a result, the MMSE-DFE forward filter is given by

$$\mathbf{F} = \mathbf{\Delta}^{-1}\mathbf{B}^{-H} \tag{2.19}$$

In order to compute the MMSE-DFE forward filter with linear complexity in the channel length $L$ and in the block size $N$, Schur algorithm can be applied by considering the banded Toeplitz structure of $\mathbf{\Gamma}$ where each row of contains at most $2L-1$ non-zero elements [73]. The output of this filter can be obtained efficiently by applying back-substitution to $\mathbf{y}$, yielding linear complexity in $L$ and $N$.

Finally, we can calculate the SINR of the decision vector $\mathbf{x}'$. Substituting (2.19) into (2.11),

$$\begin{aligned}
\mathbf{x}' &= \mathbf{\Delta}^{-1}\mathbf{B}^{-H}\overline{\mathbf{y}} - (\mathbf{B}-\mathbf{I})\overline{\mathbf{x}} && (2.20)\\
&= \mathbf{\Delta}^{-1}\mathbf{B}^{-H}(\mathbf{\Gamma}\overline{\mathbf{x}} + \overline{\mathbf{z}}) - (\mathbf{B}-\mathbf{I})\overline{\mathbf{x}} && (2.21)\\
&= -N_0\mathbf{\Delta}^{-1}\mathbf{B}^{-H}\overline{\mathbf{x}} + \mathbf{\Delta}^{-1}\mathbf{B}^{-H}\overline{\mathbf{z}} + \overline{\mathbf{x}} && (2.22)\\
&= (\mathbf{I} - N_0\mathbf{\Delta}^{-1})\overline{\mathbf{x}} + \boldsymbol{\nu} && (2.23)
\end{aligned}$$

where $(\mathbf{I} - N_0\mathbf{\Delta}^{-1})$ is diagonal and $\boldsymbol{\nu} = N_0\mathbf{\Delta}^{-1}\left(\mathbf{I} - \mathbf{B}^{-H}\right)\overline{\mathbf{x}} + \mathbf{\Delta}^{-1}\mathbf{B}^{-H}\overline{\mathbf{z}}$ denotes the residual interference plus noise term. Then, the $x'[n]$ in a natural ordering has the following SINR for $n = 0, \ldots, N-1$. Denoting

$$\mathsf{SINR}_n = \frac{(1-\epsilon_n)^2}{\mathrm{var}[\nu_n]} = \frac{1}{\epsilon_n} - 1 \tag{2.24}$$

where $\epsilon_n = N_0/\sigma[N-1-n]$ denotes the MMSE for the symbol $x[n]$ and we find the expression for $\mathrm{var}[\nu_n]$ as follows.

$$\epsilon_n = \mathbb{E}[|x'_n - \overline{x}_n|^2] = \mathbb{E}[|(1-\epsilon_n)\overline{x}_n + \nu_n - \tilde{x}_n|^2] = \epsilon_n^2 + \mathrm{var}[\nu_n]$$

## 2.4 The concatenated TCM-TR-STBC scheme

We wish to concatenate an outer code defined over a complex signal constellation $\mathcal{X} \subset \mathbb{C}$ with an inner TR-STBC scheme. For outer coding we choose standard TCM [69, 68, 49, 67] for the following reasons: 1) it is very easy to implement variable-rate coding by adding uncoded bits, expanding the signal set correspondingly, and increasing the number of parallel transitions in the same basic encoder trellis; 2) they can be easily decoded by the

Viterbi Algorithm (VA) which is particularly suited to the low-complexity joint equalization and decoding scheme proposed in the next section; 3) after the TR-STBC combining, we are in the presence of an ISI channel whose impulse response is given by the coherent combination of the $M$ channel impulse responses of the underlying MISO channel. Due to the inherent diversity combining, the effect of fading is reduced and it makes sense to choose the outer coding scheme in a family optimized for classical AWGN-ISI channels [67].

In the proposed TCM-TR-STBC scheme, the blocks of symbols $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ of the TR-STBC transmit matrix $\mathbf{T}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ are obtained by interleaving the output sequence produced by a TCM encoder. As we shall see in the next section, a block interleaver with suitable depth $D$ is necessary in order to enable the low-complexity joint equalization and decoding scheme to work efficiently. We consider a row-column interleaver formed by an array of size $N \times D$, where the symbols produced by the TCM encoder are written by rows, and the columns form the blocks $\mathbf{x}_j$ mapped into the TR-STBC transmit matrix.

Figure 2.4(a) shows the block diagram of the proposed concatenated scheme for $M = 3$, based on the rate-3/4 STBC with block length $T = 4$ defined by

$$\mathbf{S}(x_1, x_2, x_3) = \begin{bmatrix} x_1 & x_2 & x_3 \\ -x_2^* & x_1^* & 0 \\ -x_3^* & 0 & x_1^* \\ 0 & -x_3 & x_2 \end{bmatrix}$$

The sequence generated by the TCM encoder is arranged in the interleaving array by rows. The resulting $D$ vectors of length $N$ are mapped onto the $T(N + L - 1) \times M$ TR-STBC transmit matrix (this is shown transposed in Figure 2.4(a) where the shadowed areas correspond to zeros). The spectral efficiency of the resulting concatenated scheme is given by $\eta = \frac{N}{N+L-1} R_{\text{stbc}} R_{\text{tcm}}$, where $R_{\text{stbc}}$ is the rate [symbol/channel use] of the underlying STBC, and $R_{\text{tcm}}$ is the rate [bit/symbol] of the outer TCM code. The factor $\frac{N}{N+L-1}$ is the rate loss due to the insertion of the zero-padding, and can be made small by letting $N \gg L$.

## 2.5 Reduced-state joint equalization and decoding

ML decoding of the overall concatenated scheme is too complex, since it would require running a VA on an expanded trellis, where the number of states depends on the channel length and on the constellation size. To overcome this problem, we propose a reduced-state joint equalization and decoding approach based on the *per-survivor processing* (PSP) principle [74], similar to the scheme proposed in [75] for trellis STCs over the frequency-flat MIMO channel. The block diagram of the receiver is shown in Figure 2.4(b). An MMSE-DFE deals with the causal part of ISI by using the reliable decisions found on the survivors of the VA operating on the trellis of the underlying TCM code. The non-causal part of the ISI is mitigated by the forward filter of the MMSE-DFE. The efficient method for computing the MMSE-DFE forward filter with linear complexity in the channel length $L$ and in the TR-STBC block size $N$ is derived in 2.3.4.

Let $\{z[j]\}$ be the sequence of symbol-rate samples obtained after forward filtering and block deinterleaving (see Figure 2.4(b) ). Due to the structure of the interleaver, the decisions in the decision-feedback section of the equalizer

can be found on the survivors of the VA acting on the original TCM trellis (i.e., without state expansion). The resulting VA is fully defined by its branch metric. Consider the $q$-th parallel transition at the $j$-th trellis step, extending from state $s$ and merging to state $s'$. The corresponding branch metric is given by

$$m_{s,s',q}[j] = \left| z[j] - (1-\epsilon_n)\, x(s,s',q) - \sum_{\ell=1}^{L'} b_{n,\ell}\, \widehat{x}_{j-\ell D}(s) \right|^2 \tag{2.25}$$

where $L' = \min\{L, n\}$, $n = \lfloor j/D \rfloor$, $\epsilon_n = N_0/\sigma[N-1-n]$ denotes the MMSE at the output of the backward filter, $x(s,s',q)$ is the constellation symbol labeling the $q$-th parallel transition of the trellis branch $s \to s'$, $\widehat{x}_{n-\ell D}(s)$ are the tentative decisions found on the surviving path terminating in state $s$, and $(b_{n,1}, \ldots, b_{n,L})$ are the coefficients of the MMSE-DFE feedback filter where $b_{n,\ell}$ is the $(N-1-n, N-1-n+\ell)$-th element of the matrix $\mathbf{B}$.

Thanks to the interleaving depth $D$, the tentative decisions are found at least $D$ trellis steps before the symbol of interest (step $j$ in the trellis). If $D$ is larger than the Viterbi decoding delay (typically 5 or 6 times the code constraint length), the corresponding decisions are reliably obtained from the Viterbi decoder output [75]. As a matter of fact, simulations show that the scheme is extremely robust and, even if $D$ is much smaller than the typical Viterbi decoding delay, the WER performance of the proposed scheme is almost identical to that of a genie-aided scheme that makes use of ideal feedback decisions. The minimal $D$ for which ideal-feedback performance is attained depends on the specific code and should be optimized by extensive simulation.

The joint MMSE-DFE equalization and decoding algorithm is summarized as follows. For simplicity, we assume that all codewords starts at state 0 and are terminated on state 0 after $K = ND$ trellis steps where $K$ denotes the frame length. $S$ denotes the number of states in the TCM trellis.

1. Initialize the VA path metrics:
   $M_0(0) = 0, M_0(s) = +\infty$ for $s = 1, \ldots S-1$.

2. Add compare and select recursion: for $j = 1, \ldots, K-1$,

   - for $s' = 0, \ldots, S-1$, update the path metrics according to

$$M_j(s') = \min_{s \in \Pi(s'), \forall q} \{M_{j-1}(s) + m_j(s,s',q)\} \tag{2.26}$$

where $\Pi(s')$ the set of states $s$ for which a transition $s \to s'$ exists (parent states of state $s'$) and $m_j(s, s', q)$ is given by (2.25).

- update the survivors terminating in each state;

3. Trace back recursion: retrieve the sequence of information bits corresponding to the survivor terminating in state 0.

## 2.6 WER analysis

In this section, we provide two approximations to the WER of the proposed TCM-TR-STBC scheme. Both approximations are based on the assumption of a genie that helps the equalization and decoding scheme.

**Matched Filter Bound (MFB).** Assuming that a genie eliminates the whole ISI and that deinterleaving suffices to decorrelate the Gaussian noise, (2.9) is turned into the ISI-free AWGN channel

$$y^{(MFB)}[j] = \sqrt{\gamma_0} x[j] + w[j] \tag{2.27}$$

where $w[j] \sim \mathcal{N}_{\mathbb{C}}(0, N_0)$ is AWGN, $\mathbb{E}[|x[j]|^2] = \mathcal{E}$ and $\gamma_0 = \sum_{i=1}^{M} |\mathbf{g}_i|^2$. The corresponding SNR is given $\gamma_0 \mathcal{E}/N_0$. The coefficient $\gamma_0$ can be expressed by using the eigen decomposition of the covariance matrix of $\mathbf{g}_i$, given by $\mathbf{R}_g \triangleq \mathbb{E}[\mathbf{g}_i \mathbf{g}_i^{\mathsf{T}}]$, that we assume independent of $i$ for simplicity. We let $\mathbf{R}_g = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathsf{H}}$ where $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_P\}$ contains the non-zero eigenvalues on the diagonal and $\mathbf{U} \in \mathbb{C}^{N_g \times P}$ has orthonormal columns. The number $P$ of positive eigenvalues of $\mathbf{R}_g$ represents the number of fading effective degrees of freedom of the multipath channel, i.e., the number of *separable paths*.[2] In the rest of this Chapter we assume Rayleigh fading, uncorrelated scattering, and that the channel impulse responses of different antennas are statistically independent. We use the Karhunen-Loeve decomposition

$$\mathbf{g}_i = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{h}_i$$

[2]Notice that we have not made any constraining assumption about the channel delay-intensity profile [60]. Therefore, this definition applies to both diffuse and discrete multipath models.

where $\mathbf{h}_i = (h_i[1], \ldots, h_i[P])^{\mathsf{T}}$ are complex circularly symmetric Gaussian vectors with i.i.d. components $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. It follows that

$$\begin{aligned} \gamma_0 &= \sum_{p=1}^{P} \lambda_p \left( \sum_{i=1}^{M} |h_i[p]|^2 \right) \\ &= \sum_{p=1}^{P} \lambda_p \alpha[p] \end{aligned} \tag{2.28}$$

where the $\alpha[p]$'s are i.i.d. central Chi-squared random variables with $2M$ degrees of freedom.

The WER conditioned with respect to $\gamma_0$ under the MFB assumption is upper bounded by

$$P_w^{(MFB)}(e|\gamma_0) \leq K \sum_d A_d Q\left( \sqrt{\frac{\mathcal{E} d^2 \gamma_0}{2N_0}} \right) \tag{2.29}$$

where $K = DN$ denotes the frame length in trellis steps and $A_d$ is the average number of *simple* error events at *normalized* squared Euclidean distance $d^2$.[3] This function can be evaluated numerically by using the Euclidean distance enumerator $\{A_d\}$ of the TCM code. In practice, the (possibly truncated) distance enumerator can be computed by several algorithms depending on geometrical uniformity of the TCM code under examination [69, 76, 77]. In order to obtain the average WER over the realization of the channel $\gamma_0$, we cannot average the conditional union bound (2.29) term by term because the union bound averaged over the fading statistics may be very loose or even not converge if an infinite number of terms is taken into account in the union bound summation (see [78]). Then, we follow the approach of [78] and obtain

$$P_w^{(MFB)}(e) \leq \mathbb{E}_{\gamma_0} \left[ \min \left\{ 1, K \sum_d A_d Q\left( \sqrt{\frac{\mathcal{E} d^2 \gamma_0}{2N_0}} \right) \right\} \right] \tag{2.30}$$

where the expectation is with respect to the statistics of $\gamma_0$, that can be easily obtained by numerical integration. Since we have used a union upper bound in the MFB lower bound, (2.30) is neither a lower nor an upper bound. Rather, it provides a useful approximation for the actual WER $P_w(e)$.

[3] Having put in evidence the average symbol energy $\mathcal{E}$, we define the normalized Euclidean distance $d$ between two code sequences $\mathbf{x}$ and $\mathbf{x}'$ by $d^2 = |\mathbf{x} - \mathbf{x}'|^2 / \mathcal{E}$.

**Genie-Aided MMSE-DFE Gaussian approximation (MMSE-DFE-GA).** Here we assume that a genie removes only the causal ISI (i.e., the MMSE-DFE works under the ideal feedback assumption). The channel presented to the VA can be modeled as

$$y^{(GAB)}[j] = \sqrt{\beta}x[j] + w[j] \tag{2.31}$$

where $E[|w[j]|^2] = 1$, and $\mathcal{E}\beta$ is the signal to interference plus noise ratio (SINR) at the output of the MMSE-DFE under the ideal feedback assumption, given by [79]

$$\beta\mathcal{E} = \exp\left\{\int_{-1/2}^{1/2} \ln\left(1 + \frac{\mathcal{E}}{N_0}\Gamma(f)\right) df\right\} - 1 \tag{2.32}$$

where $\Gamma(f) \triangleq \sum_{i=1}^{M} G_i(f)$ and where $G_i(f)$ is the discrete-time Fourier transform of the symbol-rate sampled autocorrelation function of the $i$-th channel impulse response $\mathbf{g}_i$. The SINR expression (2.32) is obtained in the limit for large block length ($N \to \infty$), that makes the vector model (2.9) stationary. Since the term $w[j]$ in (2.31) contains both noise and anti-causal ISI, we make a Gaussian approximation and let $w[j] \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. The approximated error probability for this model can be derived exactly in the same manner as for the MFB, by replacing the SNR $\gamma_0\mathcal{E}/N_0$ in (2.30) by $\beta\mathcal{E}$. Unfortunately, the expectation with respect to $\beta$ must be evaluated by Monte-Carlo average, since the pdf of $\beta$ cannot be given in closed form. Remarkably, simulations show that this approximation is very tight and predicts very accurately the WER of the TCM-TR-STBC scheme under the actual joint equalization and decoding scheme (i.e., without ideal decision feedback).

**Achievable diversity.** The maximum achievable diversity in the MISO channel with $M$ independent antennas and $P$ separable paths is obviously given by $d_{\max} = MP$. Consider the single-input single-output channel with ISI obtained by including the TR-STBC encoding (at the transmitter) and combining (at the receiver) as part of the channel. Standard results of information theory show that the maximum information rate achievable by signals with frequency-flat power spectral density is given by[80]

$$I_G(\mathcal{E}/N_0) \triangleq \int_{-1/2}^{1/2} \log_2\left(1 + \frac{\mathcal{E}}{N_0}\Gamma(f)\right) df \tag{2.33}$$

For the quasi-static fading model considered in this Chapter, it follows that the best possible WER for any code, in the limit of large block length, is given by the *information outage probability*

$$P_{\text{out}}(\mathcal{E}/N_0, \eta) = \Pr\left(I_G(\mathcal{E}/N_0) \leq \eta\right) \tag{2.34}$$

and, by following the argument of [62, 63], that the high-SNR slope of the outage probability curve, defined by the limit

$$\lim_{\mathcal{E}/N_0 \to \infty} \frac{-\log P_{\text{out}}(\mathcal{E}/N_0, \eta)}{\log \mathcal{E}/N_0} \tag{2.35}$$

is given by $d_{\max} = MP$.

It is also well-known that the information rate (2.33) can be achieved by Gaussian codes, block interleaving, and by joint MMSE-DFE equalization and decoding (see for example the tutorial presentation in [81, Sect. VII.B] and references therein). We conclude that, in the limit of large interleaving depth $D$ and $N \gg L$, MMSE-DFE equalization and decoding with ideal Gaussian (capacity achieving) codes achieves maximum diversity. Our low-complexity decoding scheme can be seen as a practical version of the asymptotically optimal scheme and differs in two key aspects that make it practical: 1) it uses a very short interleaving depth $D$; 2) it uses very simple off-the-shelf TCM codes. Short $D$ implies unreliable feedback decision. Simulations show that the PSP approach is able to mitigate this effect and that full diversity is easily achieved by our scheme under no ideal feedback assumption.

## 2.7 Numerical Results

In order to evaluate the performance of the proposed scheme, simulations have been performed in the following conditions. Two ISI channel models are considered: a symbol-spaced $P$-path channel with the equal strength paths and the pedestrian channel B [82] for the TD-SCDMA 3rd generation standard [83]. Classical Ungerboeck TCM codes are used with different signal constellations and spectral efficiencies. WER curves are plotted versus either $E_b/N_0$ or $\mathsf{SNR}$ in dB, where we define $\mathsf{SNR} \stackrel{\Delta}{=} \frac{M\mathcal{E}}{N_0}$ as the total transmit energy per channel use over the noise power spectral density or, equivalently, as the SNR at the receiver antenna, in agreement with standard STC literature. In the following, the simulated WER curves for the actual per-survivor

processing decoder are denoted by "PSP" with an interleaving depth $D$ , the simulated WER curves for a genie-aided decoder that makes use of ideal feedback decisions are denoted by "Genie", the MFB approximation is denoted by "MFB" and the MMSE-DFE Gaussian Approximation is denoted by "MMSE-DFE-GA".

**Comparison with other schemes** Figure 2.5 compares the TCM-TR-STBC scheme with previously proposed schemes for $\eta = 2$[bit/channel use], $M = 2$ and $P = 2$ equal strength ISI channel. The corresponding information outage probability is shown for comparison. The ST-BICM schemes of [16, 19], employing turbo equalization and decoding based on a BCJR algorithm for the ISI channel and for the trellis code, yield performance similar to ours. However, these schemes have much higher receiver complexity.[4] In the case of [16], the memory-one ISI channel with 8PSK modulation has trellis complexity 64 and the 16-state convolutional code of rate-2/3 used in the BICM scheme has trellis complexity 64. Five iterations are required, yielding a total complexity of $5 \times 128 = 640$ branches per coded symbol. In the case of [20], the memory-two MISO ISI channel with 4PSK modulation has trellis complexity 256 and the 16-state TCM space-time code used has trellis complexity 64. Five iterations are required, yielding a total complexity of $5 \times 320 = 1600$ branches per coded symbol. Our scheme, with a 64-state rate-2/3 8PSK TCM code and no iterative processing, has trellis complexity of 256 branches per coded symbol.

**Some aspects of the TCM-TR-STBC scheme.** In Figure 2.6, we evaluate the impact of the number of separable paths on the WER with $M = 2$ for a spectral efficiency of 2[bit/channel use]. A 4-state 8PSK Ungerboeck TCM is used. As the number of paths increases, the slope of the curves becomes steeper and gets closer and closer to that of an unfaded ISI-free AWGN channel (TCM performance in standard AWGN). Since Ungerboeck TCM codes are optimized for the AWGN channel, this fact justifies the choice of these codes for the concatenated scheme. The performance of the actual PSP decoder lies in between the MFB and the MMSE-DFE-GA approxima-

[4] In order to obtain an implementation-free complexity estimate, we assume that the complexity of the BCJR and of the PSP algorithms are essentially given by their trellis complexity (number of branches per coded symbol). Hence, we evaluate the receiver complexity as the overall trellis complexity times the number of equalizer/decoder iterations.

tions. We have also simulated the performance of a genie-aided decoder that makes use of ideal feedback decisions. We notice that the performance of the PSP decoder coincide with that of the genie-aided decoder, showing that the effect of non-ideal decisions in the MMSE-DFE is negligible in the proposed PSP scheme already for interleaving depth $D = 4$.

In Figure 2.7, we investigated the effect of the number of transmit antennas for the 4-path equal strength ISI channel. The 4-state 8PSK Ungerboeck TCM is used, which yields a spectral efficiency of 2 [bit/channel use] for $M = 1, 2$. Since a full-rate GOD does not exist for $M = 4, 8$, the corresponding spectral efficiency is 1.5, 1[bit/channel use] respectively. By increasing the number of the transmit antennas, the actual WER performance gets closer to the MFB approximation and for 4 and 8 antennas the system achieves the MFB. This shows that the effect of ISI is reduced by increasing the system transmit diversity. In fact, the matrix $\mathbf{\Gamma}$ defined in (2.8) is given by the sum of $M$ independent Toeplitz matrices $\mathcal{H}(\mathbf{g}_i)^{\mathsf{H}}\mathcal{H}(\mathbf{g}_i)$ where the diagonal terms are real and positive while the off-diagonal terms are complex and added non-coherently with different phases. Figure 2.7 shows the averaged gain of $\gamma_l$, i.e. $\mathbb{E}[\sqrt{|\gamma_l|^2}]$ for different $l$ for $M = 1, 2, 4, 8$. As clearly shown in Figure, as $M$ increases $\mathbf{\Gamma}$ becomes more and more diagonally dominated and $\gamma_0$ gets closed to $M$.

Figure 2.9 shows the performance of our PSP scheme compared to the information outage probability for different modulation schemes (increasing spectral efficiency) over 4-path equal-strength ISI channel for $M = 2$. The 4-state Ungerboeck TCM codes are used over different constellations and the resulting spectral efficiencies are 1, 2, 3, 4 [bit/channel use] for QPSK, 8PSK, 16QAM, 32 cross respectively. For all spectral efficiencies the gap between the outage probability and the WER of the actual schemes is almost constant. This fact is due to the optimality of the underlying Alamouti code for the 2-antennas MISO channel in the sense of the diversity-multiplexing tradeoff of [62].

Figure 2.10 shows a $M = 4$ antenna system over the pedestrian channel B. The TCM-TR-STBC scheme is obtained by concatenating a 16-state Ungerboeck TCM code with the TR-STBC obtained from the rate-3/4 GOD with parameters $[T = 8, M = 4, k = 6]$ [72]. The spectral efficiencies for QPSK, 8PSK, 16QAM, 32cross are 0.75, 1.5, 2.25, 3[bit/channel use]. Even on a realistic channel model where the number of separable paths $P$ is much smaller than the length of the channel impulse response, the proposed scheme shows the same slope of the information outage probability at high SNR, which

shows that the maximum diversity $d_{\max} = MP$ is achieved. However, unlike the result in Figure 2.9, the gap to outage probability increases as the spectral efficiency becomes large. This fact is well-known and it is due to the non-optimality of GODs for $M > 2$ [62].

## 2.8 Conclusion

We proposed a concatenated TCM-TR-STBC scheme for single-carrier transmission over frequency selective MISO fading channels. Thanks to a reduced-state joint equalization and decoding approach, our scheme achieves much lower complexity with similar/superior performance than previously proposed schemes for the same spectral efficiency. Moreover, since the receiver complexity is independent of the modulation constellation size and Ungerboeck TCM schemes implement very easily different spectral efficiencies with the same encoder, by introducing parallel transitions and expanding the signal constellation, our scheme is suitable for implementing adaptive modulation with low complexity. This is a key component in high-speed downlink transmission with transmitter feedback information.

We wish to conclude with a simple numerical example inspired by a 3G system setting, showing that very high data rates with high diversity can be easily achieved with the proposed scheme. Consider a MISO downlink scenario such as TD-SCDMA [83]. This system is based on slotted quasi-synchronous CDMA at 1.28 Mchip/s ($\sim$ 2MHz bandwidth). A slot, of duration 675 $\mu$s, is formed by two data-bearing blocks of 352 chips that are separated by 144 chips of midamble for channel estimation. At the end of the second block, 16 chips of guard interval are added for slot separation. With 128 chips plus 16 chips of guard interval (total 144 chips) we can estimate easily 4 channels of length 16 chips in the frequency domain, using an FFT of length 128 samples. We can use the rate-3/4 TR-STBC for $M = 4$ antennas with a 8-PSK TCM code. Using blocks of $N = 76$ [symbols], $L = 17$, and $R_{tcm} = 2$ [bit/channel use], the resulting spectral efficiency is: $\eta = \frac{3}{4}\frac{76\times 8}{864}R_{tcm} = 1.056$ [bit/chip]. This yields 1.35 Mb/s on a single carrier. On three carriers (equivalent to the 5 MHz of the European UMTS), we obtain 4.05 Mb/s, well beyond the "dream" target of 2Mb/s of high-speed links in 3rd generation systems. We conclude that the TCM-TR-STBC scheme represents a valid candidate for the high data rate downlink of TD-SCDMA.

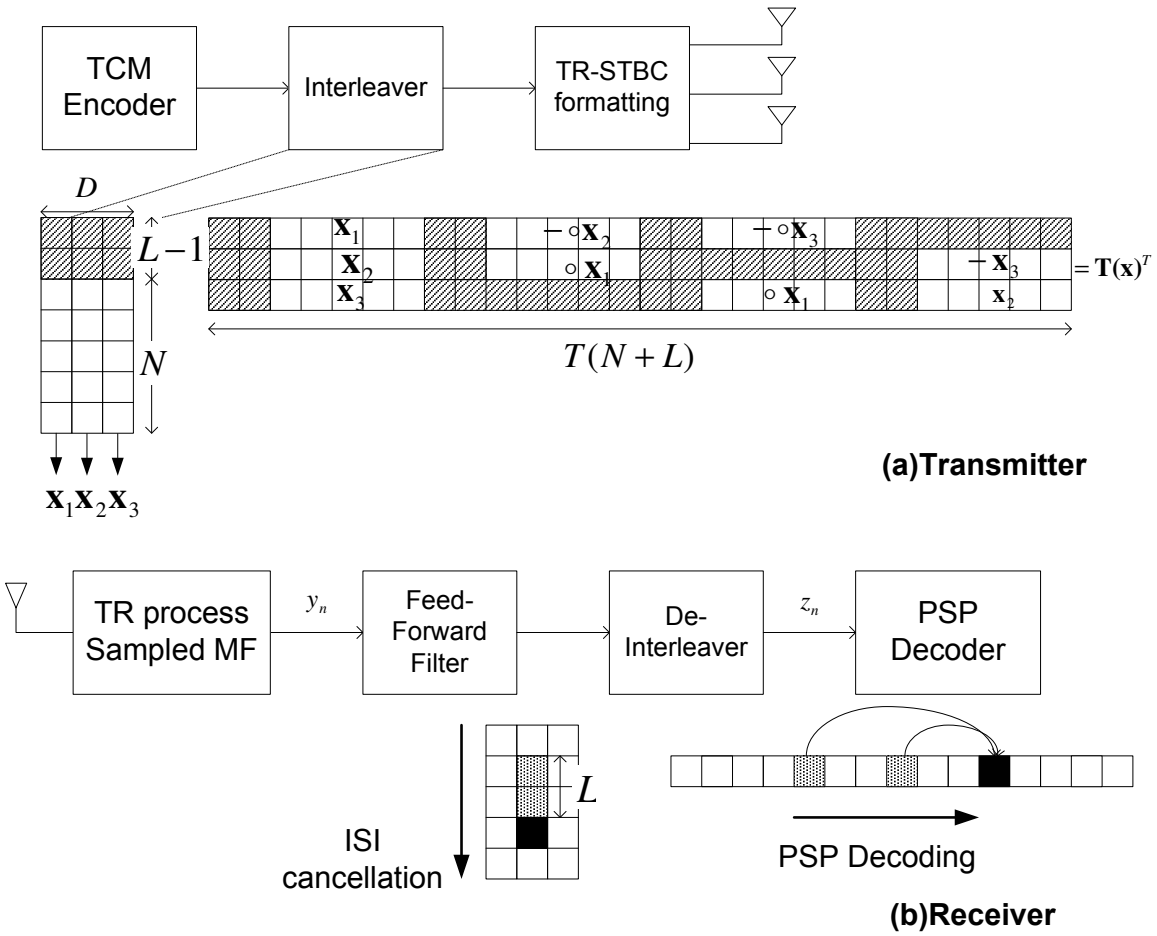


Figure 2.4: Block diagram of the TCM-TR-STBC scheme for $M = 3$.

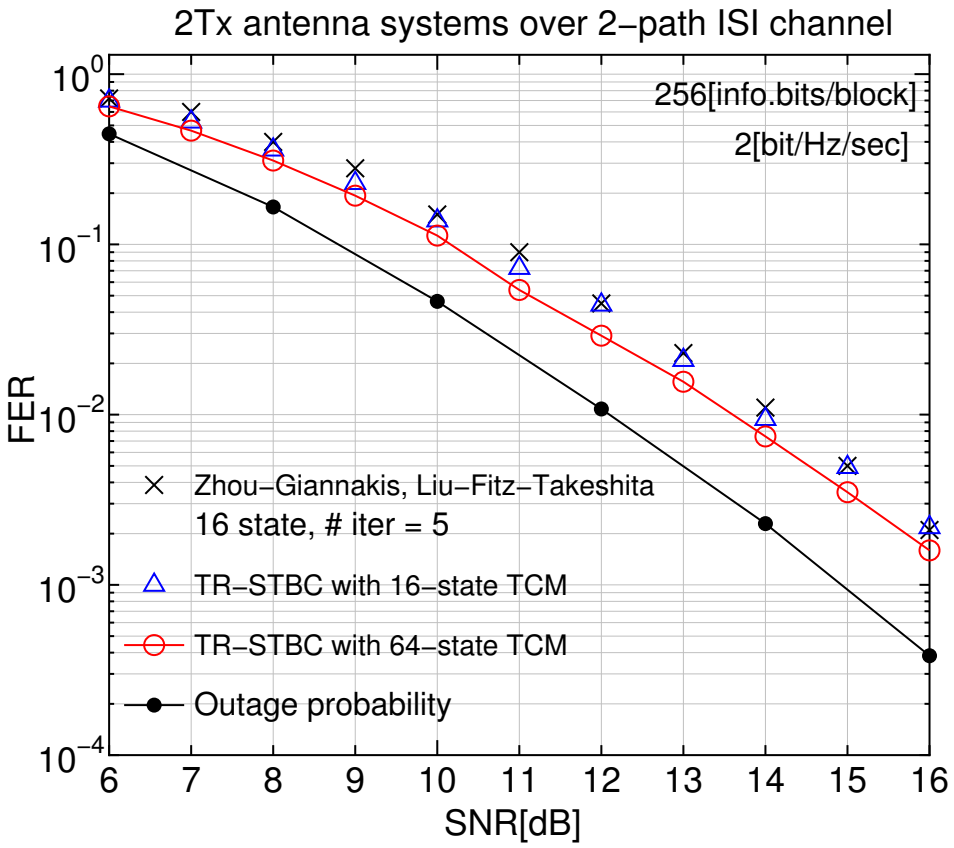


Figure 2.5: Comparison with previously proposed STC schemes.

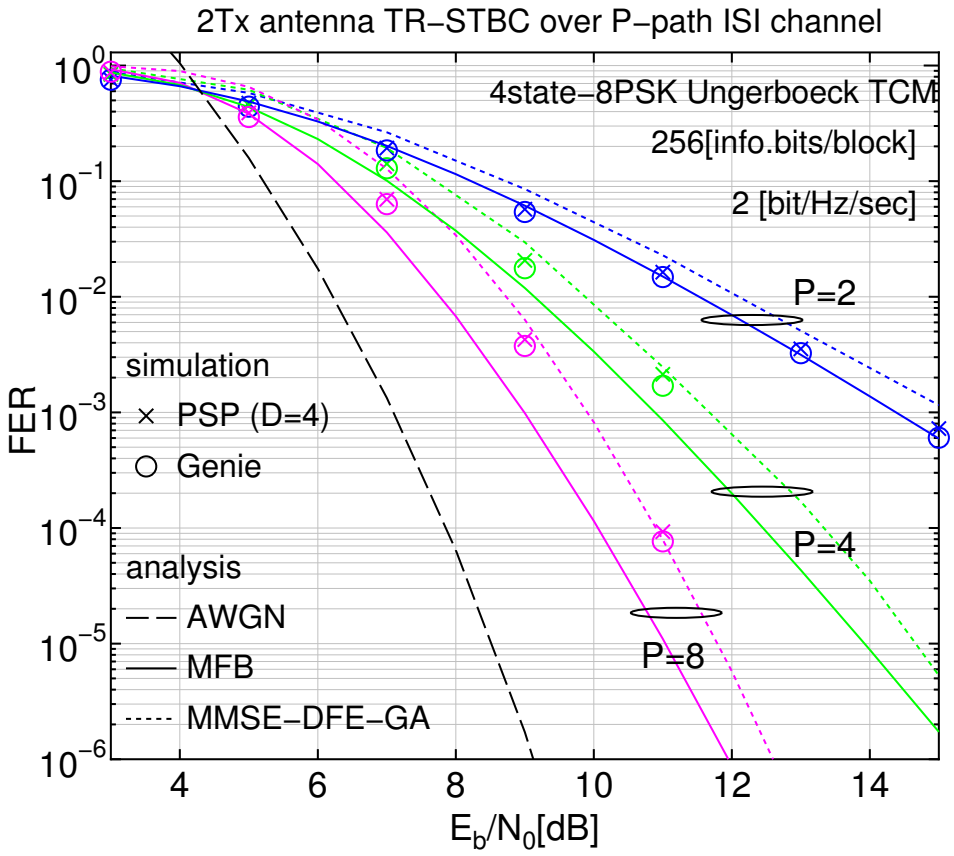


Figure 2.6: Performance of the TCM-TR-STBC scheme for $M = 2$ and increasing number of paths.

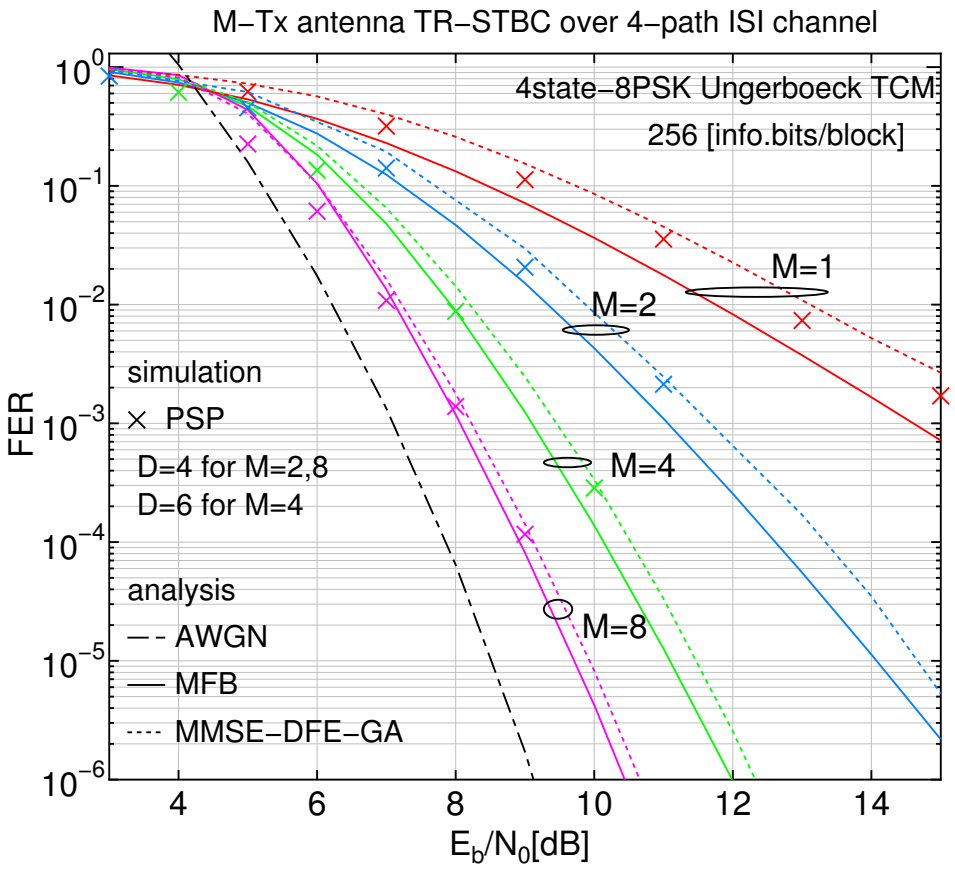


Figure 2.7: Performance of the TCM-TR-STBC scheme for $P = 4$ and increasing number of transmit antennas.

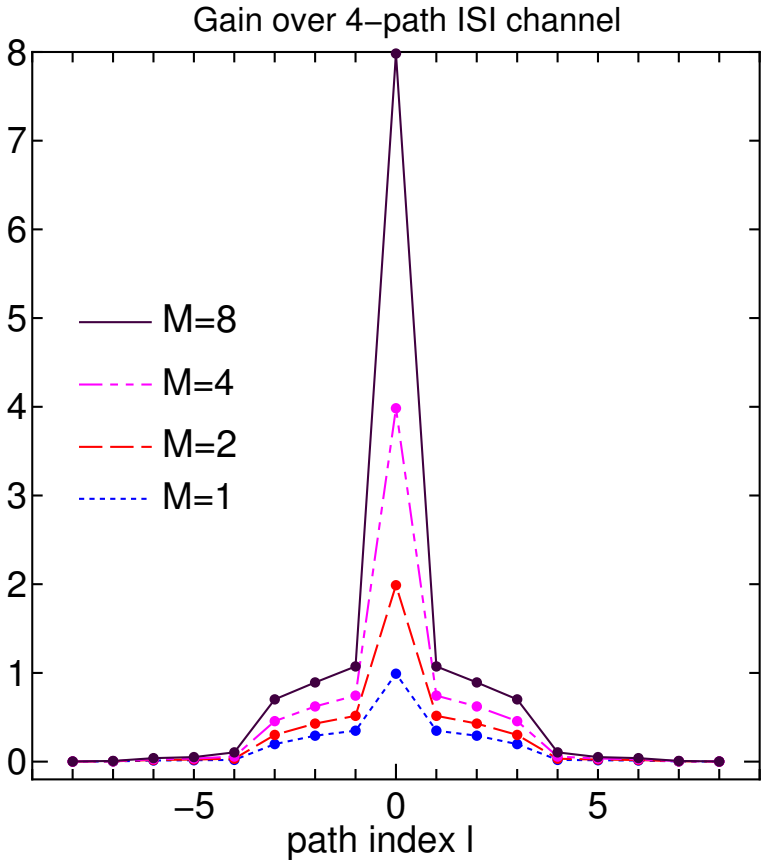


Figure 2.8: Averaged gain over 4-path ISI channel

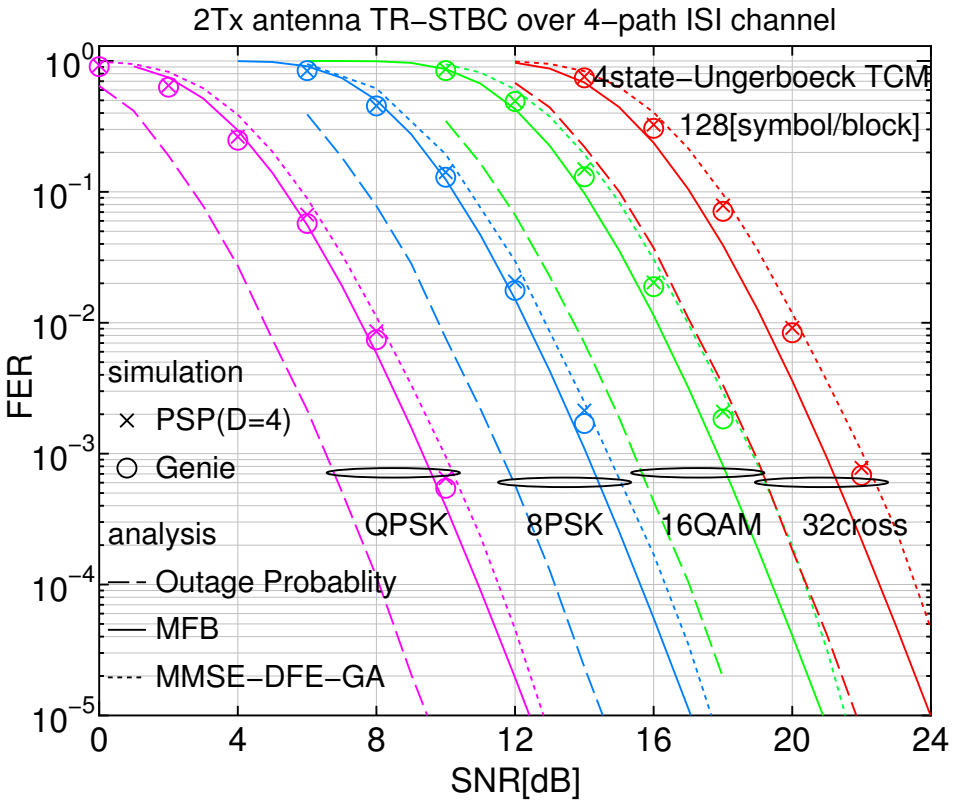


Figure 2.9: Comparison with outage probability for $M = 2$ and $P = 4$.

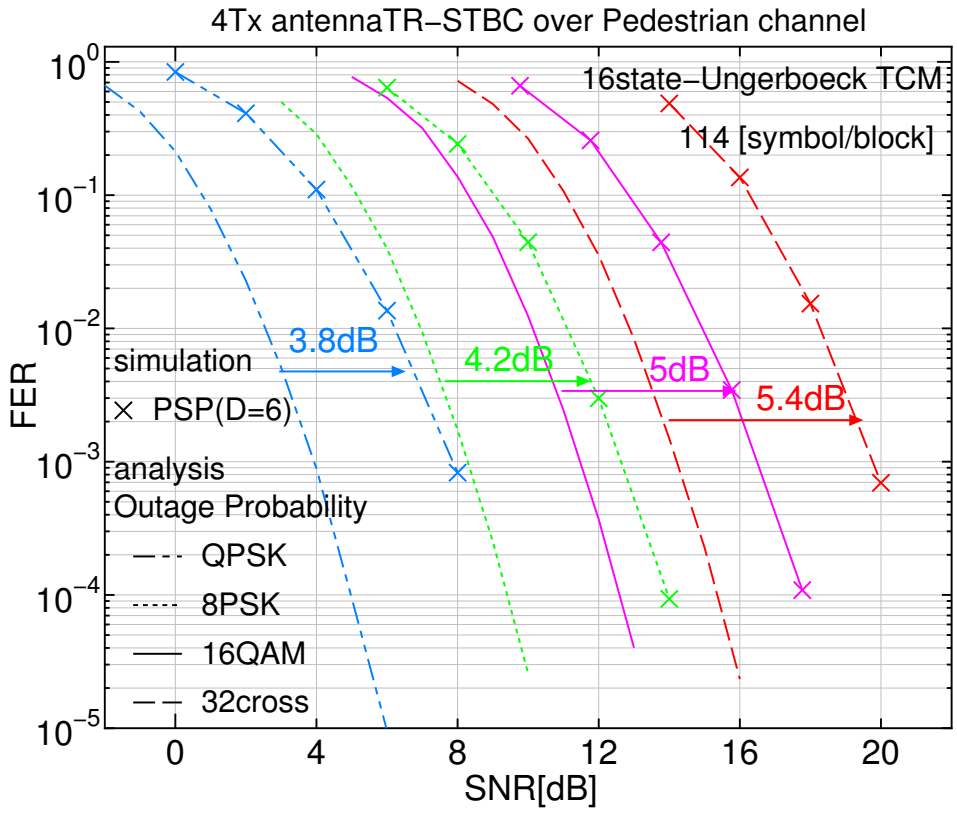


Figure 2.10: Performance over the pedestrian B channel, with $M = 4$ transmit antennas.

# Chapter 3

# MIMO Broadcast Channel with SNR Feedback

*In this chapter, we investigate the use of multiple antennas in a point-to-multipoint downlink system. We consider a MIMO broadcast channel where a transmitter equipped with $M$ antennas serves $K$ users with a single antenna each. We compare two practical signaling schemes based on a simple SNR feedback : STC and opportunistic beamforming. As seen in Chapter 2, STC (transmit diversity) improves the reliability of each user's link and equalizes the users' channels in a multiuser setting. Hence, it has been considered that transmit diversity is harmful for the system throughput under multiuser diversity based scheduling. In opportunistic beamforming, multiuser diversity is amplified because multiple transmit antennas are used to induce faster and larger fading fluctuation. We take a deeper look on the interaction between transmit diversity achieved by STC and multiuser diversity achieved by opportunistic beamforming, by taking into account the fundamental aspects that were neglected in existing works : random packet arrivals with finite transmission buffers and time-varying fading channels with a delay in the feedback link. The realistic assumption of non-perfect CSIT yields a non-trivial tradeoff between transmit diversity and multiuser diversity. For a given delay feedback, there is no clear ranking between these two schemes and the relative merit depends strongly on the channel Doppler bandwidth.*

# 3.1 Introduction

## 3.1.1 Motivation

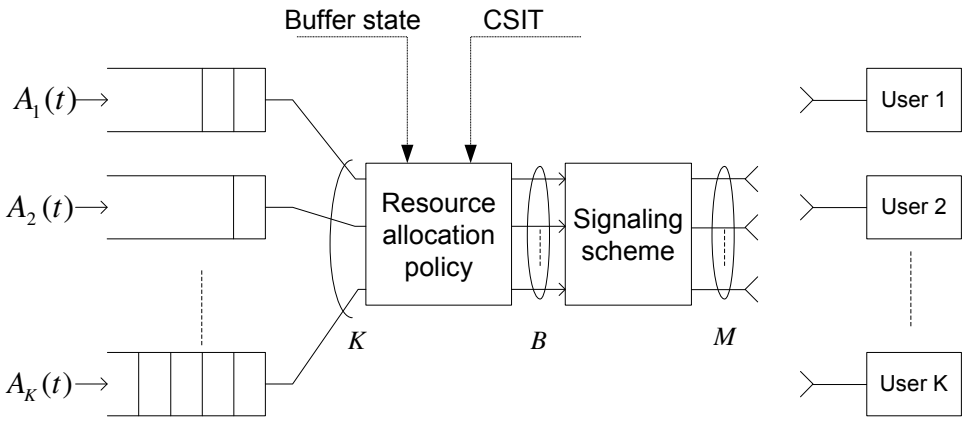


Figure 3.1: Block diagram of the SDMA/TDMA downlink

We consider the system shown in Figure 3.1 where a base station with $M$ antennas serves $K$ users, each one equipped with a single antenna. Transmission is slotted and each slot comprises $T$ channel uses (complex dimensions). Information bits arrive randomly at the transmitter and are locally stored into $K$ queues, each associated to one user.

The base station operates in SDMA/TDMA mode: at each slot, $B \leq M$ *streams* of coded signal are generated by encoding packets of information bits from the $K$ queues. Each stream is destined to one user. Hence, the system serves simultaneously $B$ users at any point in time. The $B$ streams are transmitted by using some beamforming algorithm, that is generally referred to as the *signaling strategy*. For a given signaling strategy, the resource-allocation policy formed by queue selection (scheduling) and rate allocation is referred to as a SDMA/TDMA policy.

Several results on SDMA/TDMA downlink schemes exist in the literature. Driven by the information theoretic analysis of fading broadcast channels [21, 22], systems that serve the user enjoying the best instantaneous channel conditions have been proposed for high-data rate data packet downlink in evolutionary 3G system standardization [2, 3, 1]. When the base station has multiple antennas, random "opportunistic" beamforming has been proposed

in [23, 26]. These systems generate $1 \leq B \leq M$ random time-varying beams, such that each user can measure the rate that can be reliably received on each beam and feeds back this information to the transmitter. The scheduling algorithm allocates the best user on each beam at any point in time. The ability of a system to serve a user at its peak rate conditions is generally referred to as "multiuser diversity" [24, 25, 23, 26].

A rather different use of the $M$ antennas consists of improving the reliability of transmission for each user by space-time coding (STC) (see for example [9, 7, 6] and references therein). In this case, we have $B = 1$ (TDMA) since all antennas are used to provide *transmit diversity* to a single user. Scheduling can also be applied on top of STC and it is natural to investigate the relative merit of transmit diversity and multiuser diversity. Partial answers to this question have been provided, for example, in [41, 42, 43]. These works, as well as many others, indicate essentially that transmit diversity always *decreases* the throughput of a downlink system under SDMA/TDMA scheduling. In fact, multiuser diversity benefits from highly variable user channel conditions. This effect is amplified, in some sense, by random beamforming. On the contrary, transmit diversity tends to equalize the channel of each user as the number of antennas increases (the "channel hardening" effect [42]). Therefore, while the link of each single user becomes individually more reliable, the total system throughput is decreased.

The assumptions underlying these results are: 1) channel errors never occur, i.e., once a user is scheduled and is allocated a given (channel dependent) rate, the message will be successfully received with probability 1; 2) transmission queues have infinite buffers and all users are backlogged. Assumption 1) implies that the rate that can be supported by the channel can be perfectly known to the transmitter. Hence, it does not take into account the possibility of channel measurement errors and of feedback delay which causes the transmitter to allocate rates based on outdated information. It is obviously clear that, under this assumption, STC is of no help. In fact, STC is geared to cope with bad fading channel conditions unknown to the transmitter [84] as seen in Chapter 2. Assumption 2) implies that there are always bits available for transmission for any user at any point in time. This, in turns, implies that the scheduling policy should try to maximize the system throughput subject to some *fairness* criterion. Unfortunately, "fairness" is not a well-defined concept and many criteria exist. Indeed, several different scheduling algorithms have been proposed that are optimal subject to different fairness criteria (see for example [23, 57, 5, 58, 59] and refer-

ences therein). The lack of a uniquely defined criterion to rank the different SDMA/TDMA policies makes this comparison very difficult.

In order to take a deeper look on the interaction of transmit diversity and multiuser diversity, we take into account the fundamental aspects that were neglected in works such as [41, 42, 23, 26]: random packet arrivals with finite transmission buffers and time-varying fading channels with a delay in the feedback link. Under these assumptions, the questions that we pose are the following.

- What is the notion of fairness ?
- Is STC always harmful ?
- Can the opportunistic $M$-beamforming exploit the multiplexing gain ?
- How does the Doppler effect impact the performance on STC and opportunistic-beamforming ?

### 3.1.2 Contribution

Under random bits arrival, the notion of fairness is replaced by the notion of *stability* [55, 56]. In this setting, achieving any point in the system stability region (defined formally later) subsumes any reasonable fairness criterion and may be considered as the single most important goal of a downlink resource-allocation policy.

In this work, we build on the framework of [55] and find the SDMA/TDMA stability region under non-perfect channel state information at the transmitter (CSIT), that yields non-zero decoding error probability. In our case, non-perfect CSIT is due to a delay in the feedback and to the fact that channels are time-varying. Then, we find the adaptive transmission policy that stabilizes the system for all arrival rates inside the system *stability region*. Finally, we apply the stability framework to provide a fair comparison between STC (transmit diversity) and random beamforming schemes. The realistic assumption of non-perfect CSIT yields a non-trivial tradeoff between the transmit diversity achieved by STC and the multiuser diversity achieved by opportunistic beamforming schemes. This tradeoff is missed by the analysis of [41, 42, 43, 23, 26, 57, 5, 58, 59] that neglects the fact that decoding errors may occur with non-zero probability. In our more refined and realistic setting, the ability of accurately predicting the channel SNR clearly emerges as

one of the main limiting factors of multiuser diversity-based schemes. Therefore, we also propose an improvement of the random beamforming scheme where each user is synchronized with a common random number generator that produces the random beamforming matrices. Hence, the matrices can be considered *a priori* known[1] and this information is used for channel estimation. This makes the rate of variation of the random beams and the ability of estimating the channels unrelated, whereas the latter depends only on the physical channel rate of variation.

Even with the proposed scheme, that represents a "best case" for opportunistic beamforming, our results show that for a given feedback delay the relative merit of opportunistic beamforming versus STC strongly depends on the channel Doppler bandwidth. In particular, for slowly-varying channels opportunistic beamforming with $B = M$ beams achieves the best average delay, while for faster channels STC is best. In light of these results, the utility of random beamforming with $B = 1$ is fully questionable.

### 3.1.3 Related works

The stability theory of a queued multiuser system has been investigated in [85, 55, 56, 86]. In [85, 55] the authors developed the stability theory and proposed the optimal power allocation for satellite downlink systems. The authors also generalized the Lyapunov drift over time-varying process. The optimal power/rate allocation policy for single-input single-output (SISO) multiple access channel (MAC) and broadcast channel was established in [56] and extended to MIMO-MAC in [86]. All these works consider perfect CSIT.

## 3.2 System model

We assume a frequency non-selective block-fading channel where the signal received at user $k$ terminal in slot $t$ is given by

$$\mathbf{y}_k(t) = \mathbf{X}(t)\mathbf{h}_k(t) + \mathbf{w}_k(t) \tag{3.1}$$

where $\mathbf{X}(t) \in \mathbb{C}^{T\times M}$ is the transmitted codeword, $\mathbf{h}_k(t) \in \mathbb{C}^{M\times 1}$ denotes the $M$-input 1-output channel response, assumed constant in time and frequency

[1]Notice that in practical CDMA systems each user is synchronized with the random spreading/scrambling code of the base station. Therefore, this assumption is fully realistic.

over each slot, and $\mathbf{w}_k(t) \in \mathbb{C}^{T\times 1}$ is a complex circularly symmetric AWGN with components $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. The base station has fixed transmit energy per channel use denoted by $\gamma$, that is, $\mathrm{tr}(\mathbf{X}(t)\mathbf{X}(t)^H) \leq \gamma T$, for all $t$. Due to the noise variance normalization, $\gamma$ takes on the meaning of *transmit* SNR.

Coding and decoding is performed on a slot-by-slot basis. We assume that $T$ is large enough such that powerful codes exist whose error probability is characterized by a threshold behavior: letting $R(t)$ denote the transmission rate on a given slot $t$ and $\overline{R}(t)$ denote the supremum of the coding rates supported by the channel, which is a random variable because of fading, the decoding error probability in slot $t$ is equal to $1\{\overline{R}(t) \leq R(t)\}$, i.e., the decoder makes an error with very large probability if the transmission rate is above the maximum achievable rate of the channel, while error probability is negligibly small if it is below.[2] We shall refer to the probability $\mathbb{P}(\overline{R}(t) \leq R(t))$ as the *outage probability*. In order to handle decoding errors, we assume an ideal ARQ protocol such that the unsuccessfully decoded information bits are left in the transmission buffer and shall be re-scheduled for transmission at a later time.

A SDMA/TDMA policy is generally a function of Channel State Information at the Transmitter (CSIT) and of the state of the transmitter queues. CSIT can be obtained in several ways depending on the system. For the systems considered in this work, specific idealized models for CSIT and for the signaling strategies will be specified in Section 3.5. In general, we assume that at the beginning of each slot $t$, both a CSIT signal $\boldsymbol{\alpha}(t) \in \mathcal{A}$ and the queue buffer states are revealed to the transmitter, where $\mathcal{A}$ denotes the CSIT signal alphabet. The arrival process of queue $k$ is denoted by $A_k(t)$, an ergodic process with *arrival rate* $\lambda_k \stackrel{\Delta}{=} \frac{1}{T}\mathbb{E}[A_k(t)]$ (bit/channel use). The buffer size of queue $k$ is denoted by $S_k(t)$ (bit). We let $\boldsymbol{\alpha}_1^t = \{\boldsymbol{\alpha}(\tau) : \tau = 1,\ldots,t\} \in \mathcal{A}^t$ denote the sequence of CSIT signals up to time $t$, and $\mathbf{S}_1^t = \{S_1(\tau),\ldots,S_K(\tau) : \tau = 1,\ldots,t\} \in \mathbb{R}_+^{Kt}$ denote the queue buffer state sequence up to time $t$. An SDMA/TDMA resource allocation

[2] This threshold behavior of block error probability can be approached in practice by concatenated coding schemes such as inner trellis codes with Reed-Solomon outer coding, turbo codes and low-density parity-check codes, and it is closely related to the concept of information outage probability [12] and of $\epsilon$-capacity, or outage capacity [61], which are widely used in the information-theoretic analysis of block-fading channels.

policy is formally defined by two sequences of functions

$$\begin{aligned} \mathbf{P}^{(t)} &: \mathcal{A}^t \times \mathbb{R}_+^{Kt} \to [0,1]^{K\times B} \\ \mathbf{R}^{(t)} &: \mathcal{A}^t \times \mathbb{R}_+^{Kt} \to \mathbb{R}_+^{K\times B} \end{aligned} \tag{3.2}$$

for $t = 1, 2, \ldots$, such that each user $k$ on each slot $t$ is given a fraction $p_{k,j}^{(t)}(\boldsymbol{\alpha}_1^t, \mathbf{S}_1^t)$ of dimensions of stream $j = 1, \ldots, B$, and transmits at rate $R_{k,j}^{(t)}(\boldsymbol{\alpha}_1^t, \mathbf{S}_1^t)$ bit/channel use. The function $\mathbf{P}^{(t)}$ must satisfy the SDMA/TDMA feasibility constraint

$$\sum_{k=1}^{K} p_{k,j}^{(t)} \leq 1 \tag{3.3}$$

for each $j = 1, \ldots, B$ and for all $t$. Notice also that the functions $\mathbf{P}^{(t)}$ and $\mathbf{R}^{(t)}$ depend *causally* on the CSIT and queue state sequences.

For a given signaling strategy and SDMA/TDMA policy $\{\mathbf{P}^{(t)}, \mathbf{R}^{(t)}\}$, let $\overline{R}_{k,j}(t)$ denote the supremum of the coding rates supported by the channel for user $k$ on stream $j$ in slot $t$. Notice that $\overline{R}_{k,j}(t)$ depends on the instantaneous channel realization, and is therefore a random variable. Under the assumptions given above, the queue buffer states evolve in time according to the stochastic difference equation

$$S_k(t+1) = \left[ S_k(t) - T \sum_{j=1}^{B} p_{k,j}^{(t)} R_{k,j}^{(t)} \, 1\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\} \right]_+ + A_k(t) \tag{3.4}$$

for all $k = 1, \ldots, K$, where $[\cdot]_+ \stackrel{\Delta}{=} \max\{\cdot, 0\}$. The presence of the indicator function $1\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\}$ is due to the ARQ mechanism illustrated before: only if decoding is successful, i.e., if the scheduled rate $R_{k,j}^{(t)}$ is indeed achievable, the corresponding information bits are removed from the transmission buffer.

## 3.3 Preliminaries on stability

**Definition of stability** Following [55], we define the buffer overflow function

$$g_k(S) = \limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} 1\{S_k(\tau) > S\}$$

We say that the queued downlink system is *stable* if $\lim_{S\to\infty} g_k(S) = 0$ for all $k = 1, \ldots, K$. We say that the queued downlink system is *stabilizable* if there exists a SDMA/TDMA policy $\{\mathbf{p}(t), \mathbf{R}(t)\}$ satisfying the feasibility constraint (3.3) for which the queues are stable.

The lim sup definition of stability is carefully chosen because the lim inf of any bounded function exists and therefore it can be applied to arbitrary arrivals and control policies[85].

**Stability region** For a given signaling strategy, we define the system SDMA/TDMA stability region $\Omega$ as the set of all arrival rates $K$-tuples $\boldsymbol{\lambda} \in \mathbb{R}_+^K$ such that there exists a feasible policy in the form (3.2) for which the system is stable.

**Necessary condition for stability** Let $\mu_k(t)$ denote the instantaneous service rate in slot $t$ for user $k$ in bit and $\overline{\mu}_k$ denote the long-term average service rate of user $k$ in bit/channel use, given by

$$\overline{\mu}_k = \liminf_{t\to\infty} \frac{1}{tT} \sum_{\tau=1}^{t} \mu_k(\tau) \tag{3.5}$$

For the queued downlink system with general arrival $\{A_k(t)\}$ and service rates $\{\mu_k(t)\}$, a necessary condition for stability is [55, Lemma 1]

$$\lambda_k \leq \underline{\mu}_k, \quad k = 1, \ldots, K. \tag{3.6}$$

If both the arrival process and the rate processes evolve according to an ergodic Markov chain, then a sufficient condition for stability is also $\lambda_k \leq \underline{\mu}_k$ for $k = 1, \ldots, K$.

The necessary condition of the stability will be used to establish the stability region.

**Sufficient condition for stability in Markov chains –Lyapunov drift** For the stability analysis, we use a well-developed theory of stability in Markov chains based on negative Lyapunov drift [55, 56, 85]. Let define a non-negative Lyapunov function $L(\mathbf{S}(t)) = \sum_{k=1}^{K} \theta_k S_k^2(t)$ for arbitrary positive weights $\{\theta_k\}$. The followings are the sufficient conditions for the stability of a Markov chain [55, Theorem 2].

If there exists a compact set $\Lambda \in \mathbb{R}^K$ and some $\epsilon > 0$ such that

1. $\mathbb{E}\left[L\left(\mathbf{S}(t+1)\right)|\mathbf{S}(t)\right] < \infty$, for all $\mathbf{S} \in \mathbb{R}^K$
2. $\mathbb{E}\left[L(\mathbf{S}(t+1)) - L(\mathbf{S}(t))|\mathbf{S}(t)\right] \leq -\epsilon$, whenever $\mathbf{S}(t) \notin \Lambda$
3. whenever $\mathbf{S}(t) \in \Lambda$, there is a nonzero probability that $\mathbf{S}(t+m) = 0$ for some finite integer $m$

then, a steady-state distribution for $\mathbf{S}$ exists and hence the system is stable. Moreover, if the third condition holds and there exists a positive constant $\mathcal{C}$ such that

$$\mathbb{E}[L\left(\mathbf{S}(t+1)\right) - L\left(\mathbf{S}(t)\right)|\mathbf{S}(t)] \leq \mathcal{C} - \sum_{k=1}^{K} \theta_k S_k(t) \tag{3.7}$$

then there exists a steady-state distribution for $\mathbf{S}$ with bounded first moment $\mathbb{E}[S_k]$, such that

$$\sum_{k=1}^{K} \theta_k \mathbb{E}[S_k] \leq \mathcal{C} \tag{3.8}$$

By Little's theorem [87], we let $\mathbb{E}[D_k] = \frac{\mathbb{E}[S_k]}{T\lambda_k}$ denote average delay of user $k$ expressed in slots. Then, (3.8) can be re-written by

$$\sum_{k=1}^{K} \theta_k \lambda_k \mathbb{E}[D_k] \leq \frac{\mathcal{C}}{T} \tag{3.9}$$

## 3.4 Stability region and max-stability adaptive policy

In this section we make some key simplifying assumptions that allow us to obtain a very simple characterization of $\Omega$ and a simple adaptive SDMA/TDMA policy that achieves any $\boldsymbol{\lambda} \in \Omega$. We assume that: 1) the channel vectors $\{\mathbf{h}_k(t) : k = 1, \ldots, K\}$, the CSIT signal $\boldsymbol{\alpha}(t)$ and the arrival processes $\{A_k(t) : k = 1, \ldots, K\}$ evolve according to a jointly stationary ergodic Markov process; 2) the arrival processes have $\mathbb{E}[A_k(t)^2] \leq \infty$; 3) for a given signaling strategy, for every sufficiently large $t$, the following Markov chain holds:

$$\boldsymbol{\alpha}_1^{t-1} \rightarrow \boldsymbol{\alpha}(t) \rightarrow \{\overline{R}_{k,j}(t) : k = 1, \ldots, K, j = 1, \ldots, B\} \tag{3.10}$$

In particular, the third condition implies that for all $k, j$ and sufficiently large $t$, the conditional outage probability depends only on the current CSIT value, i.e., it satisfies

$$\mathbb{P}(\overline{R}_{k,j}(t) \leq R|\boldsymbol{\alpha}_1^t) = \mathbb{P}(\overline{R}_{k,j}(t) \leq R|\boldsymbol{\alpha}(t)) \tag{3.11}$$

We can cast the stability problem for the general SDMA/TDMA system into a stability-wise equivalent, but simpler, problem for which the results of [55] are almost immediately applicable. In a new (virtual) system, $\boldsymbol{\alpha}(t)$ takes on the role of the channel state, and the instantaneous *effective* rate $R_{k,j}^{(t)} 1\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\}$ is replaced by the conditional outage rate, defined by

$$R_{k,j}^{\text{out}}(\mathbf{a}) = \sup_{R \geq 0} \; R(1 - \mathbb{P}(\overline{R}_{k,j} \leq R|\boldsymbol{\alpha} = \mathbf{a})) \tag{3.12}$$

which is a function of the current CSIT value $\boldsymbol{\alpha} = \mathbf{a} \in \mathcal{A}$.[3] In Appendix 4.A we prove:

**Theorem 1 [Stability region].** Under the above system assumptions, for any fixed signaling strategy, the system stability region is given by

$$\Omega = \text{coh} \bigcup_{\mathbf{P} \in \mathcal{P}} \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^K \;\; : \;\; \lambda_k \leq \sum_{j=1}^{B} \mathbb{E}\left[ p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\text{out}}(\boldsymbol{\alpha}) \right] \right\} \tag{3.13}$$

where "coh" means *closure of the convex hull* and where $\mathcal{P}$ is the set of *stationary* SDMA/TDMA policies $\mathbf{P} : \mathcal{A} \to [0,1]^{K,B}$ that map the current CSIT $\boldsymbol{\alpha} = \mathbf{a} \in \mathcal{A}$ into an array $[p_{k,j}(\mathbf{a})]$ satisfying (3.3). This stability region is achieved by the stationary rate allocation policy $\mathbf{R}^\star$, defined by

$$R_{k,j}^\star(\mathbf{a}) = \arg \sup_{R \geq 0} \; R(1 - \mathbb{P}(\overline{R}_{k,j} \leq R|\boldsymbol{\alpha} = \mathbf{a})) \tag{3.14}$$

□

**Proof:** see 3.A.

The following comments are in order: 1) Under the conditions assumed in this Chapter, we have that the rate allocation function $\mathbf{R}^\star$ in (3.14) is optimal for any stationary $\mathbf{P} \in \mathcal{P}$ and any SDMA/TDMA signaling scheme. This reduces the problem of the stability-optimal resource allocation policy

[3] Here and in the following we let $\boldsymbol{\alpha}$ denote a random vector over $\mathcal{A}$ with the same distribution of $\boldsymbol{\alpha}(t)$, which is independent of $t$ because of stationarity.

to the determination of $\mathbf{P} \in \mathcal{P}$ alone; 2) under any stationary policy $(\mathbf{P}, \mathbf{R}^\star)$, the user average service rates $\widetilde{\mu}_k(t) = T \sum_{j=1}^{B} p_{k,j}(\boldsymbol{\alpha}(t)) R_{k,j}^{\text{out}}(\boldsymbol{\alpha}(t))$ (bit/slot) are linear functions of $\mathbf{P} \in \mathcal{P}$ and $\mathcal{P}$ is a convex set. It follows that no convex-hull operation is needed in (3.13); 3) we can introduce the class $\mathcal{P}_{\text{on-off}}$ of *randomized* on-off SDMA/TDMA policies such that for every $\mathbf{P} \in \mathcal{P}$, there exists $\mathbf{P'} \in \mathcal{P}_{\text{on-off}}$ defined as follows: for all $\mathbf{a} \in \mathcal{A}$ and $j = 1, \ldots, B$, let user $\kappa_j$ be served on the whole slot on stream $j$ where $\kappa_j \in \{1, \ldots, K\}$ is a random variable generated according to the probability mass function $(p_{1,j}(\mathbf{a}), \ldots, p_{K,j}(\mathbf{a}))$. Clearly, $\Omega$ is achieved by restricting the union to the policies in $\mathcal{P}_{\text{on-off}}$. In practice, randomized on-off policies are preferable since handling a single user per slot per stream is much easier. Therefore, we shall restrict to these policies in the following; 4) from the convexity of $\Omega$, the region boundary $\partial\Omega$ can be obtained by letting $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K) \in \mathbb{R}_+^K$ and finding $\lambda_k(\boldsymbol{\theta}) = \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\text{out}}(\boldsymbol{\alpha})\right]$, for the policy $\mathbf{P} \in \mathcal{P}$ solution of the maximization problem

$$\max_{\mathbf{P} \in \mathcal{P}} \sum_k \theta_k \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\text{out}}(\boldsymbol{\alpha})\right] \tag{3.15}$$

Then, $\partial\Omega$ is the convex upper envelope of the points $\{\boldsymbol{\lambda}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{R}_+^K, \sum_k \theta_k = 1\}$.

For given values $\boldsymbol{\theta} \in \mathbb{R}_+^K$, the solution of (3.15) is readily obtained as

$$\widehat{p}_{k,j}(\mathbf{a}) = \begin{cases} 1 & k = \arg\ \max_{k'} \theta_{k'} R_{k',j}^{\text{out}}(\mathbf{a}) \\ 0 & k \neq \arg\ \max_{k'} \theta_{k'} R_{k',j}^{\text{out}}(\mathbf{a}) \end{cases} \tag{3.16}$$

This means that, as expected, the points on $\partial\Omega$ are achieved by time-sharing (convex-hull operation) of *deterministic* on-off policies that schedule the best user, given by $k = \arg\ \max_{k'} \theta_{k'} R_{k',j}^{\text{out}}(\mathbf{a})$, on each stream $j$. Notice that time-sharing of deterministic on-off policies yields the class of randomized on-off policies $\mathcal{P}_{\text{on-off}}$, as expected.

For given $\boldsymbol{\lambda} \in \Omega$ there exists some memoryless stationary policy $\mathbf{P}_{\boldsymbol{\lambda}} \in \mathcal{P}_{\text{on-off}}$ that stabilizes the system. However, in order to determine $\mathbf{P}_{\boldsymbol{\lambda}}$ the a priori knowledge of $\boldsymbol{\lambda}$ is generally required. This might not be available in practice. Hence, a policy that achieves stability for all $\boldsymbol{\lambda} \in \Omega$ *adaptively* (i.e., without prior knowledge of the arrival rates) is of great practical interest [55, 56]. We shall refer to this policy as the "max-stability" adaptive policy. We have,

**Theorem 2 [Max-stability adaptive policy].** Under the above system assumptions, for any fixed signaling strategy, the max-stability adaptive policy is given by

$$\widehat{p}_{k,j}(\mathbf{a},\mathbf{S}) = \begin{cases} 1 & k = \arg\ \max_{k'} \theta_{k'} S_{k'} R^{\text{out}}_{k',j}(\mathbf{a}) \\ 0 & k \neq \arg\ \max_{k'} \theta_{k'} S_{k'} R^{\text{out}}_{k',j}(\mathbf{a}) \end{cases} \tag{3.17}$$

for any strictly positive weights $\theta_k > 0$. □

**Proof:** see 3.B.

The max-stability adaptive policy allocates on each stream $j$ in slot $t$ the user maximizing the weighted outage rate $\theta_{k'} S_{k'}(t) R^{\text{out}}_{k',j}(\boldsymbol{\alpha}(t))$. The *adaptive* nature of the max-stability policy is evident from the fact that $\widehat{\mathbf{P}}$ is a memoryless stationary function of the current CSIT $\boldsymbol{\alpha}(t)$ and of the current queues state $\mathbf{S}(t)$, rather than of $\boldsymbol{\alpha}(t)$ alone. Intuitively, the buffer state represents an empirical observation of the arrival rate.

## 3.5 Space-time coding vs. opportunistic beamforming

In this section we make use of the max-stability adaptive policy obtained before and apply it to STC (transmit diversity) and random opportunistic beamforming. We assume that the channel vectors $\mathbf{h}_k(t)$ are mutually statistically independent for different index $k$ and i.i.d. for different antennas. Focusing without loss of generality on a scalar channel coefficient $h(t)$, where we drop the user and antenna index because of the i.i.d. assumption, we assume that $h(t)$ evolves from slot to slot according to a stationary ergodic $L$-order Gauss-Markov process, given by

$$h(t) = -\sum_{\ell=1}^{L} g_\ell h(t-\ell) + \nu(t) \tag{3.18}$$

where $\nu(t) \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ is an i.i.d. process. We make the optimistic assumption that the receivers can estimate *exactly* their channel vector and feed back the corresponding value without distortion (unquantized and noiseless). However, due to a fixed delay of $d$ slots in the feedback link, the CSIT is given by $\boldsymbol{\alpha}(t) = \{\mathbf{h}_k(t-\ell-d) : k = 1,\ldots,K, \ell = 0,\ldots,L-1\}$. We hasten to

say that this CSIT model is just a *convenient idealization* for which the assumptions yielding the stability region and max-stability policy obtained in Section 3.4 hold exactly. In practice, many other sources of uncertainty are present, such as noisy channel observations [44], a rate-constrained feedback link [88, 2], or an unquantized but noisy feedback link [44]. As a matter of fact, for a fixed signaling strategy and a fixed delay $d$ in the feedback, any *degraded version* of $\boldsymbol{\alpha}(t)$ defined above can only worsen the performance. This is an immediate consequence of the Markov nature of the channel processes. Hence, our assumption yields an optimistic scenario in favor of the opportunistic schemes.

Notice also that we consider only fixed signaling strategies that do not make full use of the CSIT $\boldsymbol{\alpha}(t)$. In fact, the goal of this Chapter is not to find the best possible signaling strategy for a given type of CSIT, but compare given fixed strategies under the max-stability policy. It is clear that neither opportunistic beamforming nor STC would be the best strategy under the assumed CSIT if one has full freedom of optimizing the system with respect to the signaling strategy. We compare the following signaling strategies.

**Space Time Coding.** In this case, $\mathbf{X}(t) \in \mathbb{C}^{T\times M}$ denotes the transmitted space-time codeword. One user is served in each slot, i.e., $B = 1$ (in this case we drop the stream index $j$ for notation simplicity). The system cannot exploit spatial multiplexing since the user terminals have only one antenna. Hence, STC yields only $M$-fold transmit diversity. The supremum of the rates supported by the channel under any coding scheme is given by the channel instantaneous capacity $\overline{R}_k(t) = \log_2\left(1+\gamma\frac{1}{M}|\mathbf{h}_k(t)|^2\right)$. The conditional outage probability is given by

$$\mathbb{P}(\overline{R}_k(t) \leq R|\boldsymbol{\alpha}(t)=\mathbf{a}) = 1 - \mathcal{Q}_M\left(\sqrt{\frac{2|\mathbf{g}_k(t)|^2}{\sigma_e^2}}, \sqrt{\frac{2M(2^R-1)}{\gamma\sigma_e^2}}\right) \quad (3.19)$$

where $\mathcal{Q}_M$ denotes generalized Marcum's Q function, $\sigma_e^2$ denotes the prediction MMSE [4] of $\mathbf{h}_k(t)$ from $\boldsymbol{\alpha}(t)$ and where $\mathbf{g}_k(t)$ is the MMSE predictor of $\mathbf{h}_k(t)$ from $\boldsymbol{\alpha}(t)$. Expression (3.19) is easily obtained by using the fact that $\mathbf{h}_k(t)$ and $\boldsymbol{\alpha}(t)$ are jointly Gaussian. In a practical implementation, each user sends back (a suitably quantized version of) the estimated channel gain $|\mathbf{g}_k(t)|^2/M$. Hence, the amount of feedback required by this scheme is very small.

[4] We define $\sigma_e^2 = \mathbb{E}[|h(t) - \mathbb{E}[h(t)|h(t-d),\ldots,h(t-d-L+1)]|^2]$, which is the same for all components of $\mathbf{h}_k(t)$ and all users, under the symmetry assumptions considered.

**Improved Opportunistic Beamforming.** We consider opportunistic beamforming using $1 \leq B \leq M$ mutually orthogonal pseudorandom beams, as proposed in [23, 26]. The transmitted signal is given by

$$\mathbf{X}(t) = \sum_{j=1}^{B} \mathbf{s}_j(t)\boldsymbol{\phi}_j^T(t) \tag{3.20}$$

where $\mathbf{s}_j(t) \in \mathbb{C}^{T\times 1}$ is the signal associated to beam (stream) $j$, $\boldsymbol{\phi}_j(t) \in \mathbb{C}^{M\times 1}$ is the beamforming vector for beam $j$ in slot $t$, and it is assumed that $\boldsymbol{\phi}_j^H(t)\boldsymbol{\phi}_m(t) = \delta_{j,m}$. The signal-to-interference plus noise ratio (SINR) of user $k$ in beam $j$ is equal to

$$\mathsf{SINR}_{k,j}(t) = \frac{|\boldsymbol{\phi}_j^T(t)\mathbf{h}_k(t)|^2}{B/\gamma + \sum_{m\neq j} |\boldsymbol{\phi}_m^T(t)\mathbf{h}_k(t)|^2} \tag{3.21}$$

Assuming user codes drawn from an i.i.d. Gaussian distribution (or making a Gaussian approximation of interference), the supremum of the rates supported by beam $j$ for user $k$ is given by $\overline{R}_{k,j}(t) = \log_2\left(1 + \mathsf{SINR}_{k,j}(t)\right)$. In the schemes analyzed in [23, 26], each user measures its SINR for each of the $B$ transmit beams and feeds back its best SINR and the index of the beam achieving it. The SINR values are considered instantaneously and perfectly known to the transmitter. This is clearly not a realistic assumption. Moreover, in these works it is not clear how the SINRs are estimated and at which rate the random beamforming matrix changes. The next result establishes a simple relationship between the average transmission delays and the speed of variation of the random beamformer, in the ideal case of perfect CSIT (zero outage probability).

**Theorem 3 [Bound on the average delay].** Assume that CSIT is perfect, i.e., $R_{k,j}^{\text{out}}(t) = \overline{R}_{k,j}(t)$, and that the channels change independently every $N$ slots. Then, the average user delays $\mathbb{E}[D_k]$ (expressed in slots) obtained by the max-stability policy satisfies

$$\sum_{k=1}^{K} \theta_k \lambda_k \mathbb{E}[D_k] \leq \frac{N\mathcal{K}}{2\delta} \tag{3.22}$$

where $\mathcal{K} = \sum_{k=1}^{K} \theta_k \mathbb{E}\left[\left(\frac{A_k(t)}{T}\right)^2\right] + B\sum_{k=1}^{K}\theta_k \sum_{j=1}^{B} \mathbb{E}\left[(\overline{R}_{k,j}(t))^2\right]$, and where $\delta > 0$ indicates the distance of the arrival rate vector from the stability region

boundary. □

**Proof:** see 3.C.

Under perfect CSIT and static physical channels, the average delays are bounded by a function that increases linearly with $N$, the number of slots over which the combination of physical channels and random beamformers remain constant. Hence, it is clear that under perfect CSIT it is convenient to make the random beamformers vary as fast as possible (i.e., change independently at each slot). In practice, however, if the beamformers change too rapidly the users are not able to reliably estimate their SINRs. We conclude that, under any realistic scheme to obtain CSIT, there must be a tradeoff between the goal of letting the channels rapidly varying and the need of estimating the SINRs accurately.

Motivated by this consideration, we propose the following improvement: each user in the system is synchronized with a common random number generator that generates the random beamforming matrices. Hence, the matrices can be considered *a priori* known by all users logged into the system. Moreover, since they are unitary, they have no impact on the estimation of the underlying physical channel that can be achieved by usual pilot-aided schemes. In this way, the rate of variation of the random beams is independent of the ability of estimating the channels, that depends uniquely on the physical channel Doppler bandwidth. In the numerical results in Section 3.6 , we let $N = 1$, which can be regarded as a best case for the average transmission delay.

By using standard methods of characteristic functions of Hermitian quadratic forms of Gaussian random variables [89] it is possible to compute numerically the conditional outage probability, which is needed to compute the instantaneous rate request $R^{\star}_{k,j}(t)$ and the corresponding outage rate. It can be shown that this probability depends on $\boldsymbol{\alpha}(t)$ only through the two real numbers $|\boldsymbol{\phi}_j^T \mathbf{g}_k(t)|^2$ and $\sum_{m \neq j} |\boldsymbol{\phi}_m^T \mathbf{g}_k(t)|^2$, where again $\mathbf{g}_k(t)$ denotes the MMSE prediction of $\mathbf{h}_k(t)$ from $\boldsymbol{\alpha}(t)$. The details of this calculation are given in 3.D. In a practical implementation, each user feeds back (a suitably quantized version of) the $B$ real numbers $\{|\boldsymbol{\phi}_j^T \mathbf{g}_k(t)|^2 : j = 1, \ldots, B\}$, from which the max-stability policy can be computed. Hence, also in this case the amount of feedback required by this scheme is moderately small.

## 3.6 Numerical Results

**Simulation setting** We considered mutually independent arrival processes such that $A_k(t) = \sum_{j=1}^{M_k(t)} b_{k,j}(t)$, where $M_k(t)$ is an i.i.d. Poisson distributed sequence that counts the number of packets arrived to the $k$-th buffer at the beginning of slot $t$ and $\{b_{k,j}(t)\}$ are i.i.d. exponentially distributed random variables expressing the number of bits per packet. We take $\mathbb{E}[b_{k,j}(t)] = T$ ($T = 2000$ in our simulations), so that $\lambda_k$ coincides with the average number of packets arrived in a slot ($T$ channel uses). We considered a Gauss-Markov process of order $L = 5$ where the coefficients in (3.18) are chosen in order to approximate (see [90, 91]) Jakes' autocorrelation model [92], typical of wireless mobile channels. Inspired by the HDR system [2], we let the duration of a slot be 1.67 msec, and the feedback delay $d = 2$ slots. Under this setting, the mobile speeds $v = 0$, 25, 40, 60, 80 km/h yield a channel prediction MMSE $\sigma_e^2 = 0.00$, 0.05, 0.10, 0.40, 0.60 respectively. The average SNR is set equal to $\gamma = 10$ dB. For opportunistic beamforming, we generate a new independent set of random beamforming vectors in every slot that are assumed to be known a priori by the users as explained before.

**Maximum sum rate** We evaluate the maximum sum rate of STC and opportunistic beamforming. Since the maximum sum-rate is given by the intersection between the boundary of the stability region and the symmetric arrival vector $\lambda_1 = \cdots = \lambda_K$, this allows us to know exactly the total arrival rate where the buffer size diverges under the max-stability policy. The maximum sum-rate is obtained by scheduling at slot and on each stream the user with the largest outage rate, irrespectively of the buffers state. Namely, for STC, since $\alpha_k$ is chi-squared variable with degree of freedom $2M$, the maximal sum-rate is given by

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\alpha}}\left[\max_{k=1,\ldots,K} R_k^{\text{out}}(\alpha_k)\right] &\overset{(a)}{=} \mathbb{E}_{\boldsymbol{\alpha}}\left[R^{\text{out}}(\max_{k=1,\ldots,K}\alpha_k)\right] \\ &= \int_0^\infty R^{\text{out}}(x)K\cdot\left(\frac{(\frac{M}{1-\sigma_e^2})^M e^{-\frac{Mx}{1-\sigma_e^2}}x^{M-1}}{(M-1)!}\right) \\ &\quad\cdot\left(1-e^{-\frac{Mx}{1-\sigma_e^2}}\sum_{k=0}^{M-1}\frac{\left(\frac{Mx}{1-\sigma_e^2}\right)^k}{k!}\right)^{K-1} dx\end{aligned} \tag{3.23}$$

where equality $(a)$ follows from the fact that all users have the same fading statistics and that $R^{\text{out}}(x)$ is a monotonic increasing function of $x$.

For opportunistic beamforming with $B = M$ beams, the maximum sum-rate is given by

$$\sum_{j=1}^{B} \mathbb{E}\left[\max_{k=1,\ldots,K} R_{k,j}^{\text{out}}(\boldsymbol{\alpha})\right]$$

and can be evaluated only by Monte-Carlo simulation for the non-ideal CSIT case (for the ideal CSIT case, the numerical integration given in [26] can be used).

Figures 3.2 and 3.3 show the maximum sum-rate vs. the number of users for $\sigma_e^2 = 0.0, 0.05, 0.40$ by using STC and opportunistic beamforming respectively. In Figure 3.3 we let $B = M$. In both figures, the case $M = 1$ denotes a standard single-antenna TDMA performance and coincides with the performance of opportunistic single-beamforming [23] with $M > 1$.

In Figure 3.2 we observe that number of users after which transmit diversity becomes harmful depends heavily on the CSIT quality. We have $K = 2$ for ideal CSIT ($\sigma_e^2 = 0$), and $K = 5, 28$ for $\sigma_e^2 = 0.05, 0.40$, respectively. In Figure 3.3 we observe large gain with $M = 2, 4$ beams especially for $K \geq 10, 15$, for the case of perfect CSIT. Unfortunately, the advantage of multiple beams decreases dramatically as the quality of the CSIT worsens. For $\sigma_e^2 = 0.40$, multiple beams seem to be harmful even for a large number of users in the system, i.e., multiuser diversity has completely disappeared.

**Two-user stability region** Figure 3.4 shows two-user stability region of STC for $\sigma_e^2 = 0.0, 0.40$. $M = 1$ denotes the standard single antenna performance and coincides with opportunistic single beamforming for $M > 1$. For non-ideal CSIT, since multiuser diversity can not be fully exploited, transmit diversity enlarges the whole stability region including the maximal sum-rate point. On the other hand, for perfect CSIT transmit diversity improves only the vertices of the stability region. This implies that transmit diversity is beneficial especially in asymmetric traffic conditions, however it can be harmful for symmetric traffic in a system with a large number of users and ideal CSIT.

Figure 3.5 shows the two-user stability region of opportunistic beamforming with $M = B = 2$ for $\sigma_e^2 = 0.0, 0.05$. With such a small number of users, opportunistic beamforming with 2 beams decreases the stability region even under perfect CSIT.

**Impact of the beamforming variation speed on delay** We consider opportunistic beamforming with $B = M = 4$ over a static channel for a system with 50 users with symmetric arrival. Figure 3.6 shows the average delay as well as its upper bound derived in Theorem 3 for different $N$. The upper bound becomes tighter and tighter as the arrival rate gets closer and closer to the stability region boundary. Clearly, at arrival rates between zero and the stability boundary, $N = 1$ yields the smallest average delay.

**Average buffer size vs. sum arrival rate** The average buffer size denotes the time- and user-averaged buffer size expressed in bits, For STC, we consider a 20-user system with $\sigma_e^2 = 0.40$ (60 km/h). Figure 3.7 shows the average buffer size of STC as a function of the sum arrival rate under symmetric arrival process, i.e., $\lambda_1 = \cdots = \lambda_{20}$. The max-stability scheduling policy makes the buffer bounded whenever the arrival rate is inside the stability region. Under the symmetric arrival, the buffers overflow when the sum arrival rate achieves the max sum-rate given in Figure 3.2. Transmit diversity decreases the average buffer size thanks to so called "channel hardening effect" [42], i.e., with larger $M$ the channels of all users become closed to each other and the max-stability policy essentially chooses to serve the user with the longest queue. We consider also an asymmetric arrival case where we partition 20 users into four different arrival classes with 5 users each and arrival rates $\lambda_2 = 8\lambda_1, \lambda_3 = 32\lambda_1, \lambda_4 = 64\lambda_1$. Figure 3.8 shows the average buffer size as a function of the sum arrival rate. Compared with Figure 3.7, the boundary point at which the buffer diverges is rather insensitive to large traffic imbalances. In other words, transmit diversity can accommodate for very different rates without suffering too much in terms of the sum rate.

For opportunistic beamforming, we consider a 50-user system with $\sigma_e^2 = 0.05$ (25 km/h) under the symmetric traffic condition. Figure 3.9 shows the average buffer size of opportunistic beamforming with $B = M$ beams as a function of the sum arrival rate. For $M = 1$, the performance of the single-antenna TDMA system is given as a reference. We observe that the opportunistic beamforming with $B = M$ beams keeps the buffer smaller than $M = 1$ by making the channel vary i.i.d. from one slot to another.

**Average delay vs. mobile speed** We evaluated the average delay of STC and opportunistic beamforming as a function of the mobile speed by letting the total arrival rate fixed to 2.5 bit/channel use. We consider the symmetric

arrival case. The average delay expressed in slot is given by $\mathbb{E}[D_k]/K = \mathbb{E}[S_k]/(\lambda TK)$ where $\lambda = 0.05$ in bit/channel use/user denotes the arrival bit rate per user. Figures 3.10,3.11 and 3.12 show the average delay vs. the mobile speed $v$ for a system with 50 users with STC, random beamforming with $B = M$ and random beamforming with $B = 1$, respectively. Clearly, the case $M = 1$ is the same in all three figures and refers to a standard single-antenna system.

For a very slowly-varying channel (close to $v = 0$ km/h) the STC system becomes non-ergodic and there is a positive probability of buffer overflow. This probability is reduced by increasing transmit diversity, thanks to channel-hardening effect: ergodicity which is lost in time is eventually recovered in the spatial domain by increasing the number of transmit antennas.

As seen from Figures 3.11 and 3.12, opportunistic random beamforming decreases the average delay by making the channel vary almost i.i.d.. When the channel is slow (below 40 km/h), opportunistic beamforming with $M$ beams achieves the smallest delay. As $v$ increases (i.e., the quality of CSIT becomes worse), STC outperforms random beamforming due to its better outage rate. Interestingly, the opportunistic beamforming systems become unstable (the average delay diverges) with $M = 2, 4$ and $v$ larger than 60 km/h. This means that at this speed users are essentially allocated on the wrong beam with high probability. It is also noticed that, with the parameters of this simulation, the $M$-antenna $B = 1$ random beamforming scheme proposed in [23] is outperformed by the $B = M$ random beamforming scheme for low mobile speed and by STC for larger mobile speed. Therefore, it is never useful.

## 3.7 Conclusions

We have compared two simple SDMA/TDMA schemes for downlink transmission in a mobile wireless system where the base station has multiple antennas and the user terminals have one antenna: space-time coding and random "opportunistic" beamforming. Beyond their simplicity, these schemes are also relevant since they are currently considered for standardization in evolutionary 3G systems. Our comparison is made under a general max-stability SDMA/TDMA scheduling and rate allocation policy that is relevant and more meaningful than "fairness" policies in the case of random arrivals and finite-length transmission queues. Moreover, unlike previous works, we took

into account the key aspect of non-perfect CSIT, which allows for decoding errors, and a simple ARQ protocol that retransmits packets that are not successfully decoded.

Our results evidenced that the ability of accurately estimate the channel (or beams) SINRs has a fundamental impact on the performance of opportunistic beamforming schemes. In the case of mobile communications, with a delay in the feedback, the transmitter cannot have perfect CSIT and there exists a non-trivial tradeoff between multiuser diversity and transmit diversity. It clearly appears that the multiuser diversity gain disappears as soon as the channels change too rapidly. Hence, while random beamforming with $B = M$ beams should be chosen for very slow channels, space-time coding should be chosen for faster mobility terminals. This result cannot be observed under the somehow naive assumption of no feedback delay made by other works.

We has also proposed and demonstrated an improvement of the random beamforming system such that the users are synchronized and have a priori knowledge of the (pseudo-) random beamforming matrices. This decouples the problem of channel estimation and prediction from the speed of variation of the beamformers.

Curiously, it appears that the opportunistic single beamforming scheme is not attractive since its performance is dominated by either STC for large Doppler bandwidth or by the opportunistic $M$-beam scheme for small Doppler bandwidth.

# APPENDIX

## 3.A Proof of Theorem 1

For a fixed policy $\{\mathbf{P}^{(t)}, \mathbf{R}^{(t)}\}$, the instantaneous service rate (information bits per slot) for user $k$ is given by

$$\mu_k(t) = T \sum_{j=1}^{B} p_{k,j}^{(t)} R_{k,j}^{(t)} \, 1\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\} \tag{3.24}$$

The necessary condition (3.6) says that under the assumption of Section 3.4 the system is stable if and only if $\lambda_k \leq \underline{\mu}_k$ for $k = 1, \ldots, K$, where $\underline{\mu}_k$ denotes the long-term average service rate defined in (3.5).

Let $\widetilde{\Omega}$ denote the stability region of a new system with channel state $\boldsymbol{\alpha}(t)$ and feasible rate $\mathbf{R}^{\text{out}}(t)$. The instantaneous service rate (information bits per slot) of user $k$ in the new system is given by $\widetilde{\mu}_k(t) = T \sum_{j=1}^{B} p_{k,j}^{(t)} R_{k,j}^{\text{out}}(\boldsymbol{\alpha}(t))$, which is linear (hence concave) and non-decreasing in $\mathbf{P}^{(t)}$. Theorem 1 of [55] yields that the stability region $\widetilde{\Omega}$ of the new system is given by (3.13). Hence, the theorem is proved if we show that the two stability regions coincide.

By restricting the resource allocation policies for the original system to the stationary policies $\{\mathbf{P}, \mathbf{R}^{\star}\}$, for $\mathbf{P} \in \mathcal{P}$ and $\mathbf{R}^{\star}$ defined by (3.14), the arrival processes $\{A_k(t)\}$ and the service rates $\{\mu_k(t)\}$ become jointly Markov modulated [55] and it is immediate to show, by ergodicity, that $\liminf_{t\to\infty} \frac{1}{tT} \sum_{\tau=1}^{t} \mu_k(\tau) = \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\text{out}}(\boldsymbol{\alpha})\right]$. Hence, any point $\boldsymbol{\lambda} \in \widetilde{\Omega}$ is also in $\Omega$ and it is stabilized by a policy $\{\mathbf{P}, \mathbf{R}^{\star}\}$ for some $\mathbf{P} \in \mathcal{P}$. This shows that $\widetilde{\Omega} \subseteq \Omega$.

In order to show that $\Omega \subseteq \widetilde{\Omega}$, let's assume that the channels and CSIT signals take on values in finite discrete sets (the proof can be extended for well-behaved continuous processes by using standard discretization and continuity arguments). By ergodicity and from the definition of liminf, for any

$\epsilon > 0$ there exists a sufficiently large $t_\epsilon$ such that, simultaneously,

$$\begin{aligned} \frac{|N_{\mathbf{a}}(t_\epsilon)|}{t_\epsilon} &\leq P_\alpha(\mathbf{a}) + \epsilon \\ \frac{1}{t_\epsilon T}\sum_{\tau=1}^{t_\epsilon}\sum_{j=1}^{B}\mu_{k,j}(\tau) &\geq \underline{\mu}_k - \epsilon \\ \frac{1}{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|}\sum_{\tau\in N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)} 1\{r_{k,j} < \overline{R}_{k,j}(\tau)\} &\leq 1 - \mathbb{P}(\overline{R}_{k,j} \leq r_{k,j}|\boldsymbol{\alpha} = \mathbf{a}) \end{aligned} \quad (3.25)$$

where we define the sets $N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t) = \left\{\tau \in \{1,\ldots,t\} \ : \ \boldsymbol{\alpha}(\tau) = \mathbf{a}, \mathbf{P}^{(\tau)} = \mathbf{p}, \mathbf{R}^{(\tau)} = \mathbf{r}\right\}$, $N_{\mathbf{a},\mathbf{p}}(t) = \bigcup_{\mathbf{r}} N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t)$ and $N_{\mathbf{a}}(t) = \bigcup_{\mathbf{p}} N_{\mathbf{a},\mathbf{p}}(t)$, and where $P_\alpha(\mathbf{a}) = \mathbb{P}(\boldsymbol{\alpha} = \mathbf{a})$ denotes the stationary probability of $\boldsymbol{\alpha}(t)$, which exists by assumption, and where we use the short-hand notation $\mu_{k,j}(t) = Tp_{k,j}^{(t)}R_{k,j}^{(t)}1\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\}$ to denote the instantaneous service rate for user $k$ on stream $j$ in slot $t$.

The last inequality in (3.25) follows since by assumption any feasible resource allocation policy is a causal function of the CSIT process, the CSIT process is ergodic, and $\overline{R}_{k,j}(\tau)$ is independent of $\boldsymbol{\alpha}_1^{\tau-1}$ given $\boldsymbol{\alpha}(\tau)$.

Consider any stable arrival rate vector $\boldsymbol{\lambda}$. By the necessary condition of [55, Lemma 1](see above), there exists some not necessarily stationary policy $\{\mathbf{P}^{(t)}, \mathbf{R}^{(t)}\}$ such that, for all $k = 1,\ldots,K$,

$$\lambda_k \leq \underline{\mu}_k \leq \frac{1}{t_\epsilon T}\sum_{\tau=1}^{t_\epsilon}\sum_{j=1}^{B}\mu_{k,j}(\tau) + \epsilon \quad (3.26)$$

By using (3.25), we have

$$\begin{aligned} \lambda_k &\leq \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a})\frac{1}{|N_{\mathbf{a}}(t_\epsilon)|}\sum_{\tau\in N_{\mathbf{a}}(t_\epsilon)}\sum_{j=1}^{B} p_{k,j}^{(\tau)}R_{k,j}^{(\tau)}1\{R_{k,j}^{(\tau)} < \overline{R}_{k,j}(\tau)\} + \epsilon' \\ &\leq \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a})\sum_{\mathbf{p},\mathbf{r}}\frac{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|}\frac{1}{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|}\sum_{\tau\in N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)}\sum_{j=1}^{B} p_{k,j}r_{k,j}1\{r_{k,j} < \overline{R}_{k,j}(\tau)\} + \epsilon' \\ &\leq \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a})\sum_{\mathbf{p},\mathbf{r}}\frac{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|}\sum_{j=1}^{B} p_{k,j}r_{k,j}(1 - \mathbb{P}(\overline{R}_{k,j} \leq r_{k,j}|\boldsymbol{\alpha} = \mathbf{a})) + \epsilon'' \\ &\overset{(a)}{\leq} \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a})\sum_{j=1}^{B}\widetilde{p}_{k,j}(\mathbf{a})R_{k,j}^{\text{out}}(\mathbf{a}) + \epsilon'' \\ &= \widetilde{\lambda}_k + \epsilon'' \end{aligned} \quad (3.27)$$

where (a) follows from the definition of $R_{k,j}^{\text{out}}(\mathbf{a})$ (see (3.12)) and by letting

$$\widetilde{p}_{k,j}(\mathbf{a}) = \sum_{\mathbf{p}} \frac{|N_{\mathbf{a},\mathbf{p}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|} p_{k,j}$$

Since by assumption $\sum_{k=1}^{K} p_{k,j} \leq 1$ for all $j = 1, \ldots, B$ on every slot, and since $\sum_{\mathbf{p}} \frac{|N_{\mathbf{a},\mathbf{p}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|} = 1$, it follows that $\sum_{k=1}^{K} \widetilde{p}_{k,j}(\mathbf{a}) \leq 1$ for all $j$ and $\mathbf{a} \in \mathcal{A}$, i.e., the stationary policy $\{\widetilde{p}_{k,j}\}$ defined above is feasible. Hence, $\widetilde{\boldsymbol{\lambda}}$ with $k$-th component given by the last line of (3.27) is a point in $\widetilde{\Omega}$. Since $\epsilon > 0$ is arbitrary, and $\epsilon'' \to 0$ as $\epsilon \to 0$, we have that $\boldsymbol{\lambda} \leq \widetilde{\boldsymbol{\lambda}} \in \widetilde{\Omega}$, which eventually implies that $\widetilde{\Omega}$ and $\Omega$ coincide.

## 3.B Proof of Theorem 2

Following [55, Theorem 3], we demonstrate that max-stability policy stabilizes the system whenever the arrival rate vector is inside the stability region. For the sake of analysis simplicity, we assume that both the underlying channel process $\{\mathbf{h}(t)\}$ and the arrival process $\{\mathbf{A}(t)\}$ vary i.i.d. from one slot to another.

The first condition of the Lyapunov drift is satisfied because of the assumption $\mathbb{E}[|A_k(t)|^2] < \infty$. The third condition is also satisfied because we assume that there is non-zero probability that no arrivals occur to any queue during a time slot. Hence, we focus on the second condition of Lyapunov drift. Let's consider one-slot buffer dynamic under max-stability policy.

$$\begin{aligned}
S_k^2(t+1) &= \{\max[0, S_k(t) - \mu_k(t)] + A_k(t)\}^2 \\
&\leq (S_k(t) - \mu_k(t))^2 + A_k(t)^2 + 2A_k(t)\max[0, S_k(t) - \mu_k(t)\}] \\
&\leq (S_k(t) - \mu_k(t))^2 + A_k(t)^2 + 2A_k(t)S_k(t) \\
&= S_k(t)^2 - 2TS_k(t)\left(\frac{\mu_k(t)}{T} - \frac{A_k(t)}{T}\right) + T^2\left\{\left(\frac{\mu_k(t)}{T}\right)^2 + \left(\frac{A_k(t)}{T}\right)^2\right\}
\end{aligned}$$

where the instantaneous service rate under the max-stability policy is given by

$$\mu_k(t) = T\sum_{j=1}^{B} p_{k,j}(t) R_{k,j}^{\star}(t)\ 1\{R_{k,j}^{\star}(t) < \overline{R}_{k,j}(t)\}$$

Taking expectations conditioned on $\mathbf{S}(t)$, scaling by weights $\theta_k$ and summing over all $k$, we have

$$\mathbb{E}\left[L(\mathbf{S}(t+1)) - L(\mathbf{S}(t))|\mathbf{S}(t)\right] \leq T^2\tilde{\mathcal{C}} - 2T\sum_{k=1}^{K}\theta_k S_k(t)\left(\sum_{j=1}^{B}\mathbb{E}[R_{k,j}^{\text{out}}(t)|\mathbf{S}(t)] - \lambda_k\right) \quad (3.28)$$

where we used $\lambda_k = \mathbb{E}[\frac{A_k}{T}|\mathbf{S}_t]$ expressed in bit per channel use from i.i.d. arrival process and used $\mathbb{E}[\frac{\mu_k}{T}|\mathbf{S}(t)] = \sum_{j=1}^{B}\mathbb{E}[R_{k,j}^{\text{out}}(t)|\mathbf{S}(t)]$ from the definition and where have $\tilde{\mathcal{C}} = \sum_k \theta_k \mathbb{E}\left[(\mu_k(t)/T)^2|\mathbf{S}(t)\right] + \sum_k \theta_k \mathbb{E}\left[(A_k(t)/T)^2\right]$. $\tilde{\mathcal{C}}$ is upper bounded by $\mathcal{C}$ given by

$$\mathcal{C} = B\sum_{k=1}^{K}\theta_k\sum_{j=1}^{B}\mathbb{E}\left[R_{k,j}^{\star}(t)^2\right] + \sum_{k=1}^{K}\theta_k\mathbb{E}\left[\left(\frac{A_k(t)}{T}\right)^2\right]$$

In order to derive the first term of $\mathcal{C}$, we used the following

$$\begin{aligned}
&\sum_{k=1}^{K}\theta_k\mathbb{E}\left[\left(\frac{\mu_k(t)}{T}\right)^2\right] \\
&\overset{(c)}{\leq} \sum_{k=1}^{K}\theta_k\mathbb{E}\left[\left(\sum_{j=1}^{B}p_{k,j}(t)^2\right)\left(\sum_{j=1}^{B}R_{k,j}^{\star}(t)^2\,1\{R_{k,j}^{\star}(t) < \overline{R}_{k,j}(t)\}^2\right)|\mathbf{S}(t)\right] \\
&\overset{(d)}{\leq} B\sum_{k=1}^{K}\theta_k\sum_{j=1}^{B}\mathbb{E}\left[R_{k,j}^{\star}(t)^2\,1\{R_{k,j}^{\star}(t) < \overline{R}_{k,j}(t)\}^2\right] \\
&\leq B\sum_{k=1}^{K}\theta_k\sum_{j=1}^{B}\mathbb{E}\left[R_{k,j}^{\star}(t)^2\right]
\end{aligned}$$

where inequality (c) follows from Cauchy-Schwartz inequality and inequality (d) follows from $\sum_{j=1}^{B}p_{k,j}(t)^2 \leq B$ for any $k$.

Notice that the term $\sum_{k=1}^{K}\theta_k S_k(t)\sum_{j=1}^{B}\mathbb{E}[R_{k,j}^{\text{out}}(t)|\mathbf{S}(t)]$ in (3.28) maximizes the value of $\sum_{k=1}^{K}\theta_k S_k(t)\eta_k$ over all the vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K) \in \Omega$. Hence,

$$\sum_{k=1}^{K}\theta_k S_k(t)\sum_{j=1}^{B}\mathbb{E}[R_{k,j}^{\text{out}}(t)|\mathbf{S}(t)] \geq \sum_{k=1}^{K}\theta_k S_k(t)\eta_k \quad (3.29)$$

For any $\boldsymbol{\lambda}$ inside the stability region $\Omega$, there exists $\boldsymbol{\delta} = (\delta, \ldots, \delta)$ such that $\boldsymbol{\lambda} + \boldsymbol{\delta} \in \Omega$. Then, the one step drift (3.28) is rewritten by

$$\mathbb{E}\left[L(\mathbf{S}(t+1)) - L(\mathbf{S}(t))|\mathbf{S}(t)\right] \leq T^2\mathcal{C} - 2T\delta \sum_{k=1}^{K} \theta_k S_k(t) \tag{3.30}$$

Choose any $\epsilon > 0$ and define the compact region $\Lambda$

$$\Lambda \triangleq \left\{ \mathbf{S} \in \mathbb{R}^K : S_k > 0, \sum_{k=1}^{K} \theta_k S_k \leq \frac{T^2\mathcal{C} + \epsilon}{2T\delta} \right\}. \tag{3.31}$$

Then, for $\mathbf{S} \notin \Lambda$, the RHS of (3.30) is strictly less than $-\epsilon$. □

# 3.C Proof of Theorem 3.

The feasible rates $\overline{R}_{k,j}(t) = \log_2(1 + \mathsf{SINR}_{k,j}(t))$ change randomly and independently every $N$ slots. We shall refer to a block of $N$ slots over which the feasible rates are constant as a *frame*. Theorem 3 can be proved by considering the stability of a frame-based modification of max-stability policy, which maximizes $\sum_k \theta_k S_k(t)\widetilde{\mu}_k(t)$ subject to the SDMA/TDMA feasibility constraint (3.3) at every frame and keeps the same policy during all $N$ slots in a frame irrespectively of the arrival processes.

Applying (3.7) to a $N$-slot Lyapunov drift, if for some constants $\mathcal{K}_1$ and $\nu_k$ the condition

$$\mathbb{E}[L(\mathbf{S}(t+N)) - L(\mathbf{S}(t))|\mathbf{S}(t)] \leq \mathcal{K}_1 - \sum_{k=1}^{K} \nu_k S_k(t) \tag{3.32}$$

on the Lyapunov drift holds, then $\sum_k \nu_k \mathbb{E}[S_k(t)] \leq \mathcal{K}_1$. By Little's theorem, letting $\mathbb{E}[D_k] = \mathbb{E}[S_k]/T\lambda_k$ denote average delay of user $k$ expressed in slots, we obtain $\sum_k \nu_k \lambda_k \mathbb{E}[D_k] \leq \frac{\mathcal{K}_1}{T}$. Let consider an arbitrary time $t_0$ and the $N$-slot buffer evolution

$$S_k(t_0 + N) = \left[ S_k(t_0) - NT \sum_{j=1}^{B} p_{k,j}(t_0)\overline{R}_{k,j}(t_0) \right]_+ + \sum_{\tau=t_0}^{t_0+N-1} A_k(\tau) \tag{3.33}$$

It is immediate to show that the Lyapunov drift can be bounded as

$$\mathbb{E}\left[L(\mathbf{S}(t_0+N)) - L(\mathbf{S}(t_0))|\mathbf{S}(t_0)\right] \leq N^2T^2\widetilde{\mathcal{K}} - 2NT \sum_k \theta_k S_k(t_0) \left(\mathbb{E}[\overline{\mu}_k(t_0)|\mathbf{S}(t_0)] - \lambda_k\right) \tag{3.34}$$

where we define $\overline{\mu}_k(t_0) = \sum_{j=1}^{B} p_{k,j}(t_0) R_{k,j}(t_0)$ and $A_k = \frac{1}{N}\sum_{\tau=t_0}^{t_0+N-1} A_k(\tau)$ and where $\widetilde{\mathcal{K}} = \sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k}{T}\right)^2\right] + \sum_k \theta_k \mathbb{E}\left[\overline{\mu}_k^2(t_0)|\mathbf{S}(t_0)\right]$. In order to bound this term by a constant, we have

$$\begin{aligned}\sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k}{T}\right)^2\right] &\overset{(a)}{\leq} \sum_k \theta_k \frac{1}{N}\sum_{\tau=t_0}^{t_0+N-1} \mathbb{E}\left[\left(\frac{A_k(\tau)}{T}\right)^2\right] \\ &\overset{(b)}{=} \sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k(\tau)}{T}\right)^2\right]\end{aligned}$$

where inequality (a) follows from Jensen's inequality and equality (b) follows from the i.i.d. arrival assumption.

Then, we also have that

$$\begin{aligned}\sum_{k=1}^{K} \theta_k \mathbb{E}\left[\overline{\mu}_k^2(t_0)|\mathbf{S}(t_0)\right] &\overset{(c)}{\leq} \sum_{k=1}^{K} \theta_k \mathbb{E}\left[\left(\sum_{j=1}^{B} p_{k,j}^2(t_0)\right)\left(\sum_{j=1}^{B} \overline{R}_{k,j}^2(t_0)\right)|\mathbf{S}(t_0)\right] \\ &\overset{(d)}{\leq} B\sum_{k=1}^{K}\theta_k \sum_{j=1}^{B} \mathbb{E}\left[\overline{R}_{k,j}^2(t_0)\right]\end{aligned}$$

where inequality (c) follows from Cauchy-Schwartz inequality and inequality (d) follows from $\sum_{j=1}^{B} p_{k,j}^2 \leq B$ for any $k$. By letting $\mathcal{K} = \sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k(t)}{T}\right)^2\right] + B\sum_{k=1}^{K}\theta_k \sum_{j=1}^{B}\mathbb{E}\left[\overline{R}_{k,j}^2(t)\right]$ we have $\widetilde{\mathcal{K}} \leq \mathcal{K}$. It follows that Theorem 3 holds for the frame-based policy. Finally, we argue that the frame-based policy is a suboptimal policy that deliberately ignores the buffer state at times $t \neq t_0 + mN$, for $m = 1, 2, \ldots$. With little effort it is possible to show that the frame-based policy also achieves the system stability region for any finite $N$ and that if the max-stability (slot-based) policy and the frame-based policy are applied to the same initial state $\mathbf{S}(t_0)$ then $\sum_k \theta_k \mathbb{E}[S_k^{slot}(t)] \leq \sum_k \theta_k \mathbb{E}[S_k^{frame}(t)]$ for all $t \geq t_0$. Then, Theorem 3 applies also to the max-stability policy.

## 3.D Conditional cdf of SINR

We derive the cdf of $\mathsf{SINR}_{k,j}$ conditioned on $\mathbf{g}_k$, i.e. the MMSE prediction of $\mathbf{h}_k$. Assuming that $\mathbf{h}_k$ and $\mathbf{g}_k$ are jointly Gaussian, the conditional pdf of $\mathbf{h}_k$

given $\mathbf{g}_k$ is also Gaussian $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{g}_k, \sigma_e^2\mathbf{I})$ where $\sigma_e^2$ denotes the prediction MMSE. Then we have the following

$$z_{k,j} = \hat{z}_{k,j} + e_{k,j} \tag{3.35}$$

where we denote for notation simplicity $z_{k,j} = \boldsymbol{\phi}_j^T \mathbf{h}_k$, $\hat{z}_{k,j} = \boldsymbol{\phi}_j^T \mathbf{g}_k$, and $e_{k,j} = \boldsymbol{\phi}_j^T \mathbf{e}_k$ for $k = 1, \ldots, K$ and $j = 1, \ldots, B$. The conditional pdf of $z_{k,j}$ given $\hat{z}_{k,j}$ is Gaussian $z_{k,j} \sim \mathcal{N}_{\mathbb{C}}(\hat{z}_{k,j}, \sigma_e^2)$. Remark that $e_{k,j}^* e_{k,m} = 0$ for any $m \neq j$ because of orthogonality between different beams. Under the assumption that all users have the same fading statistics and the average SNR, we can remove the user index $k$. Then, without loss of generality, we consider the conditional cdf of SINR for beam 1 given $\hat{z}_1, \ldots, \hat{z}_B$.

$$\begin{aligned} \Pr\left(\mathsf{SINR}_1 \leq x\right) &= \Pr\left(\frac{|z_1|^2}{B/\gamma + \sum_{m\neq 1} |z_m|^2} \leq x\right) \\ &= \Pr\left(\Delta \leq Bx/\gamma\right) \end{aligned} \tag{3.36}$$

where we let $\Delta = |z_1|^2 - x\sum_{m\neq 1} |z_m|^2$. We can further express $\Delta$ as a Hermitian quadratic form of complex Gaussian random variable (HQF-GRV)[93].

$$\begin{aligned} \Delta &= (z_1^* \ldots z_B^*) \begin{pmatrix} 1 & \mathbf{0}_{B-1}^T \\ \mathbf{0}_{B-1} & -x\mathbf{I}_{B-1} \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_B \end{pmatrix} \\ &= \mathbf{z}^H \mathbf{F} \mathbf{z} \end{aligned} \tag{3.37}$$

where $\mathbf{0}_{B-1}, \mathbf{I}_{B-1}$ denotes a column zero vector, an identity vector of size $B-1$ respectively, and we denote the complex Gaussian vector $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\hat{\mathbf{z}}, \mathbf{R})$ of length $B$ with the covariance matrix $\mathbf{R} = \sigma_e^2\mathbf{I}$ and the $B \times B$ Hermitian matrix $\mathbf{F}$. Then, the characteristic function of the HQF-GRV is given by [93]

$$\begin{aligned} \Phi_\Delta(s) &= \mathbb{E}[\exp(-\Delta s)] = \frac{\exp\left(-s\hat{\mathbf{z}}^H \mathbf{F}(\mathbf{I} + s\mathbf{R}\mathbf{F})^{-1}\hat{\mathbf{z}}\right)}{\det(\mathbf{I} + s\mathbf{R}\mathbf{F})} \\ &= \frac{\exp\left\{-s\left(\frac{|\hat{z_1}|^2}{1+s\sigma_e^2} + \frac{-x\sum_{m\neq 1}|\hat{z_m}|^2}{1-sx\sigma_e^2}\right)\right\}}{(1+s\sigma_e^2)(1-sx\sigma_e^2)^{B-1}} \end{aligned} \tag{3.38}$$

It is worth noticing that the characteristic function is conditioned on the two values, $|\hat{z_1}|^2$ and $\sum_{m\neq 1} |\hat{z_m}|^2$. Then, the conditional cdf can be evaluated by

Laplace transform inversion formula,

$$\Pr(\Delta \leq Bx/\gamma) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} \Phi_\Delta(s)e^{sBx/\gamma}\frac{ds}{s} \tag{3.39}$$

where the vertical integration path $s = c + j\omega$ $\omega \in \mathbb{R}$ must belong to the region of convergence (ROC) of $\Phi_\Delta(s)e^{sBx/\gamma}/s$. The ROC is the vertical strip $0 < c < \frac{1}{x\sigma_e^2}$ where $\frac{1}{x\sigma_e^2}$ corresponds to the positive real part of the singularities of $\Phi_\Delta(s)e^{sBx/\gamma}/s$ [89]. The numerical technique for evaluating (3.39) is given in [89] based on Gauss-Chebyshev quadrature. Using the Gauss-Chebyshev quadrature, the conditional cdf is given by [89]

$$\begin{aligned}\Pr(\Delta \leq Bx/\gamma) = \frac{1}{\nu}\sum_{i=1}^{\nu/2}\{&\mathrm{Re}[\Phi_\Delta(c + jc\tau_i)e^{(c+jc\tau_i)Bx/\gamma}] + \\ &\tau_i \mathrm{Im}[\Phi_\Delta(c + jc\tau_i)e^{(c+jc\tau_i)Bx/\gamma}]\} + E_\nu\end{aligned} \tag{3.40}$$

where $\tau_i = \tan((2i-1)\pi/(2\nu))$ and the error term $E_\nu$ vanishes as $\nu \to \infty$.

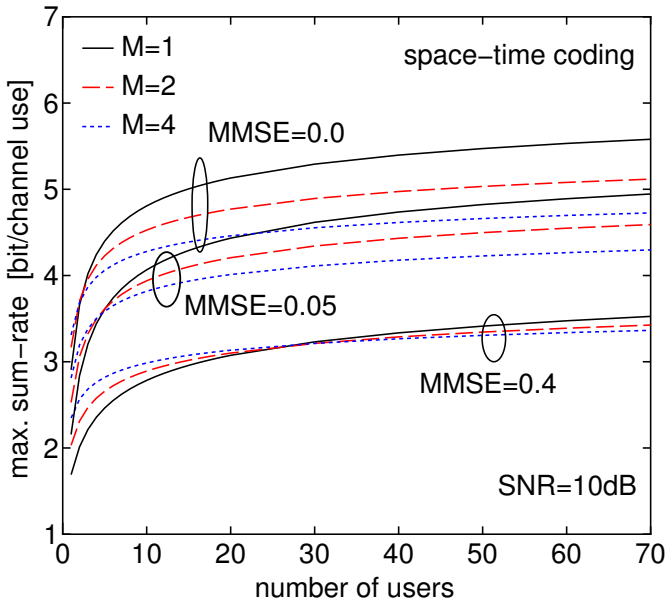


Figure 3.2: Max sum-rate vs. number of users (STC)

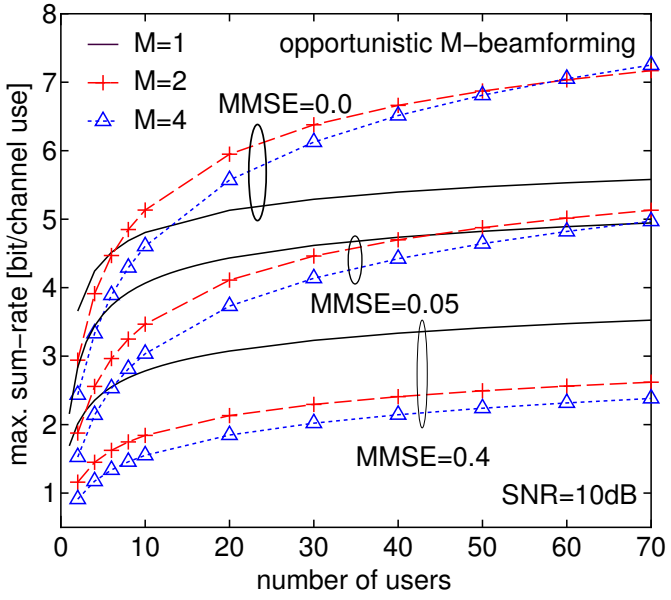


Figure 3.3: Max sum-rate vs. number of users (beamforming with $B = M$)

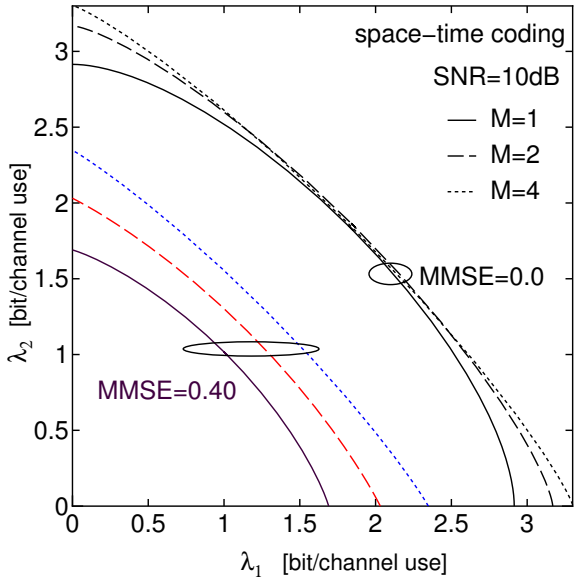


Figure 3.4: STC stability region for $\sigma_e^2 = 0.40$

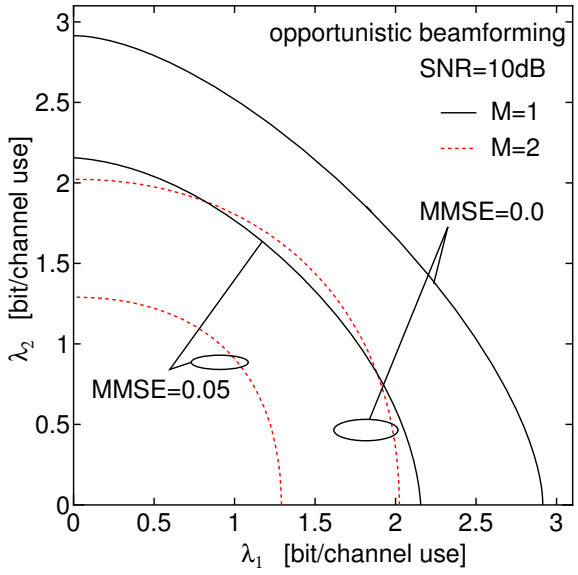


Figure 3.5: Beamforming stability region for $\sigma_e^2 = 0.05$

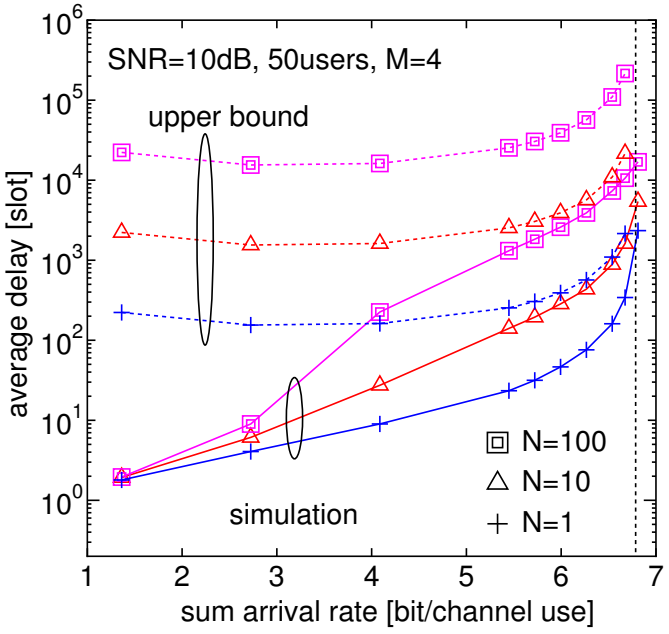


Figure 3.6: Impact of beamforming variation speed

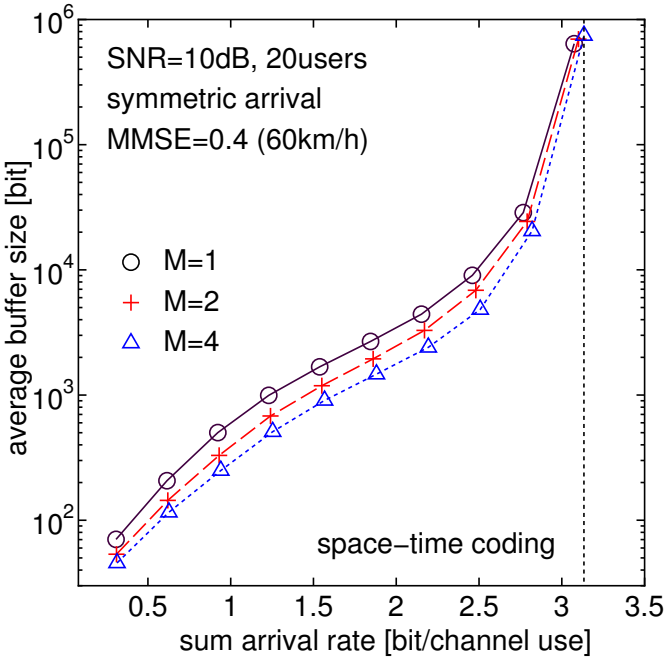


Figure 3.7: Average buffer size for symmetric arrival (STC)

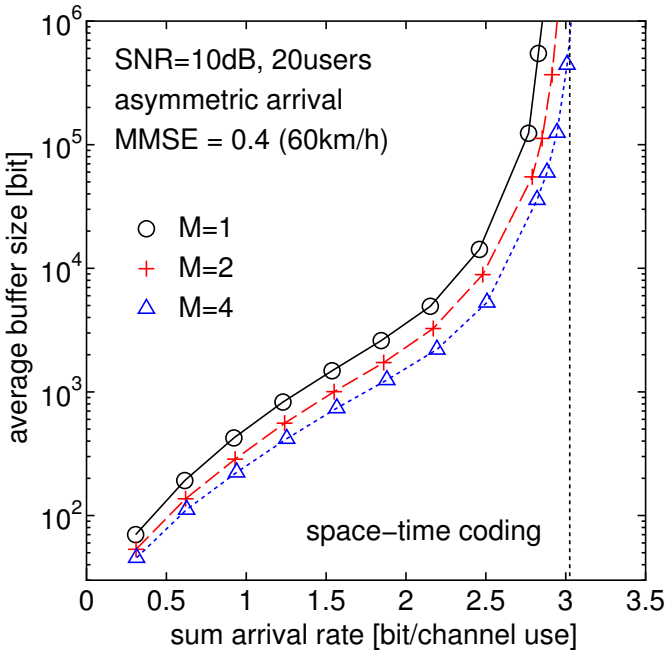


Figure 3.8: Average buffer size for asymmetric arrival (STC)

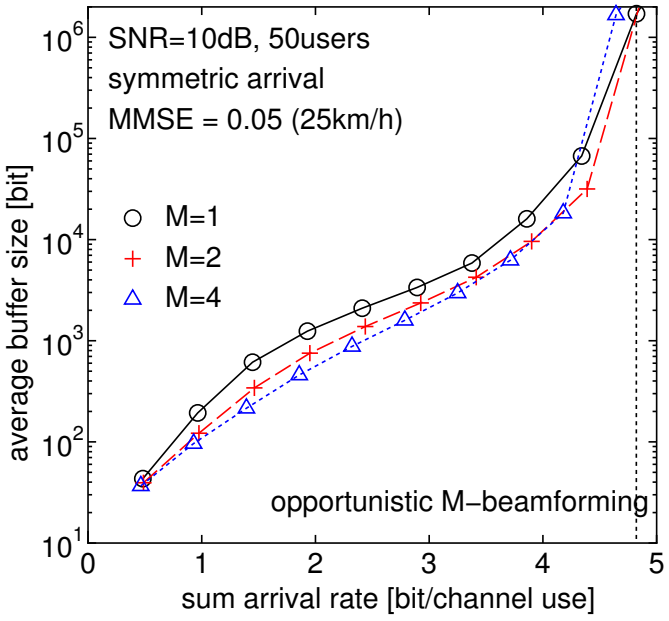


Figure 3.9: Average buffer size for symmetric arrival ($B = M$ beamforming)

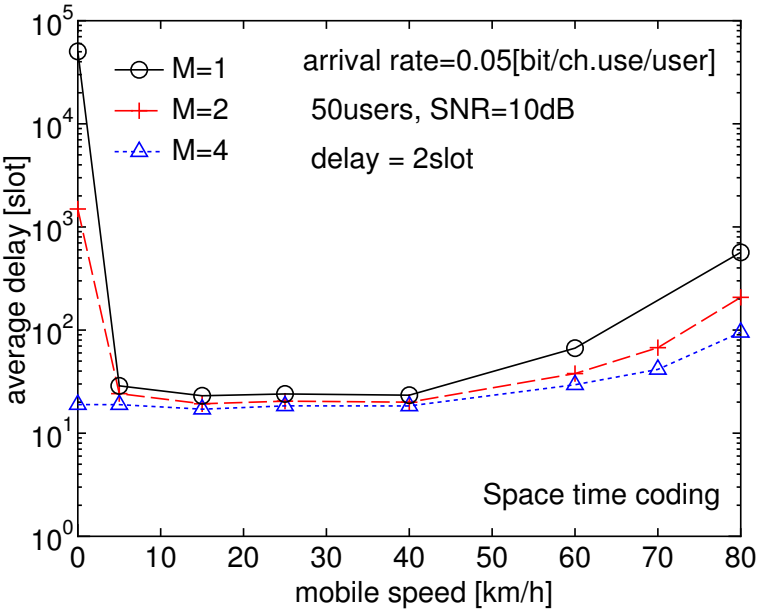


Figure 3.10: Average delay vs. speed (STC)

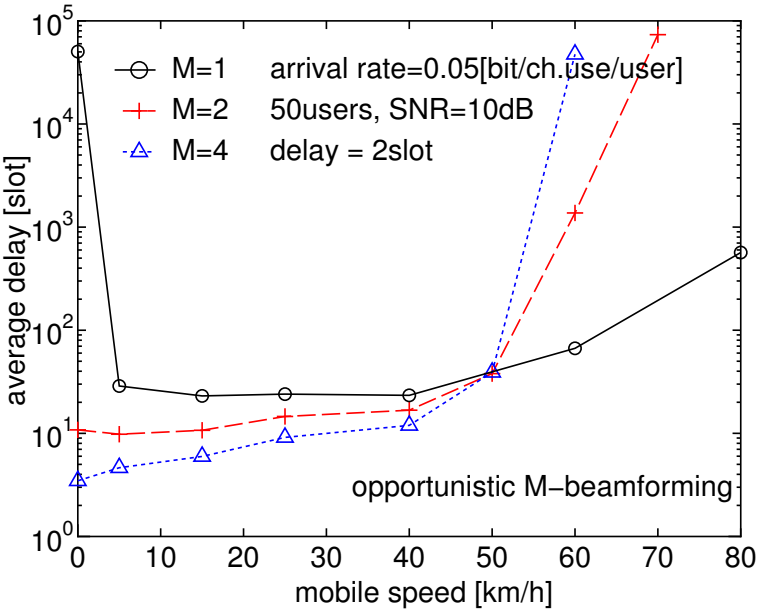


Figure 3.11: Average delay vs. speed ($B = M$ beamforming)

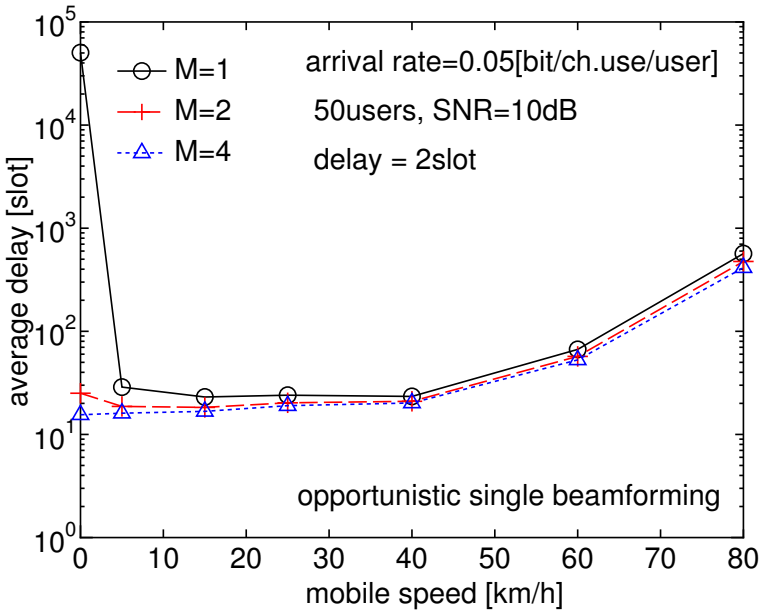


Figure 3.12: Average delay vs. speed ($B = 1$ beamforming)

# Chapter 4

# MIMO Broadcast Channel with Analog Feedback

*The opportunistic beamforming strategy of Chapter 3 based on SNR feedback makes a very simple use of multiple transmit antennas in a MIMO broadcast channel. In this Chapter, we consider a more informative type of feedback that enables a smarter use of the multiple antennas at the transmitter. In particular, we consider a distortion-wise optimal "analog feedback" scheme in which each user sends back its estimated channel vector without quantizing and coding over the feedback link. Then, we propose a novel method for user selection and linear beamforming that maximizes the weighted sum rate over a beamforming matrix. Under perfect CSIT, the proposed scheme achieves near dirty-paper coding (DPC) performance by exploiting multiplexing gain, power gain, and multiuser diversity. Under non-perfect CSIT, the proposed scheme adjusts the number of beams so that it should not be interference-limited. Over a time-varying channel with a given feedback delay, our proposed scheme outperforms the previously considered space-time coding and opportunistic beamforming for any channel Doppler bandwidth.*

# 4.1 Introduction

## 4.1.1 Motivation

We consider the same downlink system given in Figure 3.1 of Chapter 3. The transmitter equipped with $M$ transmit antennas serves $K$ user terminals with a single antenna. The transmitter has $K$ transmission queues each of which is destined to one user. The base-station operates in SDMA mode: at each slot $B \leq M$ streams of coded signals are generated by encoding the packets of information bits from the $K$ queues. Under the assumptions of random packet arrivals and non-ideal CSIT due to a delay on the feedback link, we wish to find the better signaling scheme than opportunistic beamforming by smartly exploiting CSIT with a reasonable amount of feedback.

In [28, 29, 30], it has been shown that the sum-rate capacity of a MIMO broadcast channel can be achieved by dirty paper coding (DPC)[31]. Since DPC requires the transmitter to have perfect CSI of all users, it is difficult to implement in practice especially when the number of users is large. Hence, various practical precoding strategies [32] have been proposed for a MIMO broadcast channel including linear beamforming [33, 34, 35, 36, 37], non-linear precoding [38, 39]. Among these precoding strategies for a MIMO broadcast channel, we focus on linear beamforming because it is simple and can be generalized to non-perfect CSIT case, although it is generally sub-optimal. Moreover, linear beamforming has been shown to asymptotically approach DPC performance [40, 32, 37] and references therein. The sum rate scales $M \log \mathsf{SNR}$ as $\mathsf{SNR} \to \infty$ for fixed $K >> M$, and scales $M \log \log K$ as $K \to \infty$ and for fixed $\mathsf{SNR}$. The recent results [35, 36] show that the sum-rate scaling $M \log \log K$ is achieved even for a practical number of users.

In order to perform linear beamforming, the feedback from each user should represent the statistics of a channel vector instead of a digitalized SNR value. In [44, 45], the authors study a distortion-wise optimal "analog feedback" of CSI where the $K$ users' channel coefficients are directly sent over the feedback link without quantizing and coding. Moreover, it is shown that if the MIMO downlink channel is an independent Rayleigh fading channel and the uplink feedback channel is AWGN, the unquantized and uncoded feedback is optimal in that it achieves the same MMSE distortion as the optimum scheme[44, 45, 94]. Motivated by these results, we consider that each user sends back the estimated channel coefficient vector over the feedback link without any quantization and coding.

Under the considered queued downlink with random packet arrivals, the main goal is to the stabilize the system whenever the arrival rate is inside the system stability region. As argued in Chapter 3, achieving any point in the system stability region subsumes any reasonable fairness criterion. Therefore, for a given linear beamforming strategy, we wish to derive the stability region and an adaptive policy that stabilizes the system based on the framework of [55]. We take into account the non-perfect CSIT due to a delay in the feedback channel. The challenge is how to maximize for each CSIT the weighted sum-rate over a beamforming matrix, because the sum-rate maximization with respect to the beamforming matrix generally falls into a non-convex optimization problem [32]. Then, we address the following the questions:

- What is the optimal scheduling under linear beamforming strategy ?
- Can linear beamforming outperform STC and opportunistic beamforming ?
- How does a time-varying channel impact the performance of linear beamforming under non-perfect CSIT ?

### 4.1.2 Contribution

We derive max-stability stationary policies under linear beamforming that stabilize the system whenever the arrival rate is inside the stability region. Unfortunately, such policies are difficult to implement in practice. Thus, we propose a novel heuristic method that jointly optimizes user (queue) selection and the linear beamforming matrix so that the weighted sum-rate should be maximized. A user (queue) selection consists of choosing a subset of $1 \leq B \leq M$ queues among the non-empty queues. The proposed method can be seen as a practical version of the max-stability stationary policies. Interestingly, the numerical example shows that our proposed scheme stabilizes the system until the sum arrival rate achieves the boundary of the stability region.

Under perfect CSIT, the numerical results show that for a fixed number of users our proposed scheme achieves near DPC performance by exploiting both multiplexing gain and power gain and that the sum rate increases as $M \log \mathsf{SNR}$ where $\mathsf{SNR}$ is expressed in dB. Under non-perfect CSIT, our proposed scheme adjusts the number of simultaneously served users so that it should not be interference-limited. However, it loses full multiplexing gain.

We evaluate the relative merit of our linear beamforming scheme and the previously considered schemes, space-time coding and opportunistic beamforming with $M$ beams, over the time-varying channel. At a given feedback delay, it is found that our proposed scheme achieves the best average delay over any range of mobile speed. Specifically, for slow varying channels, our proposed scheme allocates accurate beams and powers to users with large weighted rates by exploiting multiplexing gain and power gain, thus balances the queues very well. For faster channels, while the opportunistic $M$-beamforming becomes strongly interference-limited and then becomes unstable, the proposed scheme tends to allocate the whole power to the user with the largest weighted rate at each slot. Compared to STC that also operates in TDMA mode, our proposed scheme achieves better average delay because of the larger outage rate.

Therefore, our scheme offers an attractive solution for a next generation of high-speed downlink system exploiting MIMO technology such as HSDPA Release 6 or 1xEV-DO Rev.A.

### 4.1.3 Related works

There exist several methods for linear beamforming and user selection over the MIMO broadcast channel in the literature [36, 35, 46, 37]. In those works, the authors consider zero-forcing beamforming to construct $M$-parallel fading channels by choosing the best subset of $M$ users whose channels are quasi-orthogonal. Since the selection of the users for maximizing the sum rate results in a non-convex optimization problem and it is difficult to handle, several heuristic approaches are proposed [36, 35, 46, 37]. The optimality criteria is sum-rate maximization since they assume the infinite backlog for transmission. In all these works [36, 35, 37], it is assumed that channel errors never occur meaning that the instantaneous rate that can be supported by the channel is perfectly known to the transmitter, even though the quantization error in the CSIT is taken into account in [46].

The stability-optimal policy for MIMO multi-access channel (MAC) under individual power constraint and total power constraint has been established in [86, 95]. The optimal scheduling policy determines the optimal successive interference cancellation (SIC) order and the rate so that the users with large queues receive high rates. This policy for MIMO-MAC is related to the stability optimal policy for MIMO broadcast channels (BC) because of the duality between MIMO-BC and MIMO-MAC [28]. The duality says that

the same rates can be achieved for both downlink and uplink with successive encoding/decoding under the same power constraint. Hence, we can obtain the stability optimal policy for MIMO-BC under the total power constraint by successive encoding (DPC) with the reverse encoding order as MIMO-MAC.

## 4.2 Preliminary on analog feedback

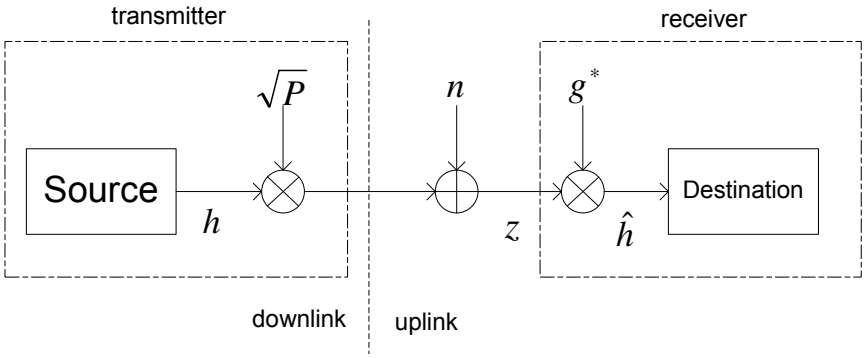


Figure 4.1: Analog feedback that achieves the MMSE distortion

The "analog feedback" of CSI has been studied and shown to be distortion-wise optimal in [44, 45, 94]. Consider the CSI transmission system illustrated in Figure 4.1. The source symbol is the downlink channel coefficient $h$ which is complex, continuous in amplitude and discrete in time. We assume that the source is Gaussian with $h \sim \mathcal{N}_{\mathbb{C}}(0,1)$. The average transmit power is $P$ and the uplink feedback link introduces AWGN $n \sim \mathcal{N}_{\mathbb{C}}(0, N_0^{\text{up}})$. At the receiver (located in the base-station), the received signal $z$ is scaled by $g^*$ so that the estimated symbol at the destination is given by

$$\hat{h} = g^* z = g^*(\sqrt{P}h + n) \tag{4.1}$$

We choose $g$ to minimize the mean squared error (MSE) between $h$ and $\hat{h}$.

$$g = \arg_f \min \mathbb{E}[|h - \hat{h}|^2] = \arg_f \min \mathbb{E}[|h - f^*(\sqrt{P}h + n)|^2]$$

This is given by

$$g = \frac{\sqrt{P}}{P + N_0^{\text{up}}}$$

and the resulting MMSE is given by

$$\min \mathbb{E}[|h - \hat{h}|^2] = \frac{1}{1 + P/N_0^{\text{up}}}$$

The rate distortion theory [80] states that for a given distortion $0 \leq D \leq 1$ the rate distortion function, i.e. the averaged number of bits per symbol is given by

$$R = \log_2 \left( \frac{1}{D} \right)$$

The distortion can be expressed in terms of the rate

$$D = 2^{-R}$$

Since the capacity of AWGN channel is $C_{\text{awgn}} = \log_2 \left( 1 + \frac{P}{N_0^{\text{up}}} \right)$, the optimal distortion in MMSE is given by

$$D = \frac{1}{1 + P/N_0^{\text{up}}}$$

which is precisely the MMSE achieved by the analog feedback. This result motivates us to use the analog feedback where each user directly sends back its estimated channel vector without quantizing and coding.

## 4.3 System Model

We assume the same downlink model illustrated in 3.1 of Chapter 3, where the transmitter equipped with $M$ antennas serves $K$ users with a single antenna. We consider a slotted frequency flat block-fading channel. One slot comprises $T$ channel uses. The received signal at channel use $i$ of slot $t$ is given by

$$\mathbf{y}_i(t) = \mathbf{H}(t)\mathbf{x}_i(t) + \mathbf{w}_i(t) \quad \text{for } i = 1, \ldots, T \tag{4.2}$$

where $\mathbf{H}(t) = [\mathbf{h}_1(t)^T, \ldots, \mathbf{h}_K(t)^T]^T \in \mathbb{C}^{K \times M}$ is a $M$-input $K$-output channel response assumed constant over a slot in time and in frequency, $\mathbf{x}_i(t) \in \mathbb{C}^{M \times 1}$ is the transmitted vector that forms a codeword $(\mathbf{x}_1(t), \ldots, \mathbf{x}_T(t)) \in \mathbb{C}^{M \times T}$, and $\mathbf{w}_i(t) \in \mathbb{C}^{K \times 1}$ is a complex circularly symmetric AWGN with components $\sim \mathcal{N}_{\mathbb{C}}(0, N_0^{\text{down}})$. The base station has unit normalized transmit energy per channel use, that is, $\text{tr}(\mathbf{x}_i(t)\mathbf{x}_i(t)^H) \leq 1$, for all $i$ and $t$ so that

$1/N_0^{\text{down}}$ denotes downlink *transmit* SNR. For notation simplicity, we remove the channel use index $i$ in the following. Under a given linear beamforming strategy, the transmitted vector $\mathbf{x}(t)$ can be further expressed by

$$\mathbf{x}(t) = \frac{1}{\sqrt{\text{tr}(\mathbf{G}(t)^H \mathbf{G}(t))}} \mathbf{G}(t)\mathbf{u}(t) \tag{4.3}$$

where $\mathbf{G}(t) = (\mathbf{g}_1(t), \ldots, \mathbf{g}_K(t)) \in \mathbb{C}^{M \times K}$ is a unnormalized linear beamforming matrix and $\mathbf{u}(t) \in \mathbb{C}^{K \times 1}$ is the information vector such that $\mathbb{E}[\mathbf{u}(t)\mathbf{u}^H(t)] = \mathbf{I}_K$ for any $t$. Then, the equivalent received of user $k$ at any given channel use is given by

$$y'_k(t) = \mathbf{h}_k^T(t)\mathbf{g}_k(t)u_k(t) + \mathbf{h}_k^T(t) \sum_{j \neq k} \mathbf{g}_j(t)u_j(t) + w'_k(t) \tag{4.4}$$

where $w'_k(t) \sim \mathcal{N}_{\mathbb{C}}(0, N_0^{\text{down}} \text{tr}(\mathbf{G}^H(t)\mathbf{G}(t)))$ is an i.i.d. process. The corresponding SINR of user $k$ is given by

$$\mathsf{SINR}_k(t) = \frac{\phi_{k,k}(t)}{N_0^{\text{down}} \text{tr}(\mathbf{G}^H(t)\mathbf{G}(t)) + \sum_{j \neq k} \phi_{k,j}(t)} \tag{4.5}$$

where we let $\phi_{k,j}(t) = |[\mathbf{H}(t)\mathbf{G}(t)]_{k,j}|^2$.

Coding and decoding is performed on a slot-by-slot basis. We assume that $T$ is large enough such that powerful codes exist whose error probability is characterized by a threshold behavior: letting $R(t)$ denote the transmission rate on a given slot $t$ and $\overline{R}(t)$ denote the supremum of the coding rates supported by the channel, which is a random variable because of fading, the decoding error probability in slot $t$ is equal to $1\{\overline{R}(t) \leq R(t)\}$, i.e., the decoder makes an error with very large probability if the transmission rate is above the maximum achievable rate of the channel, while error probability is negligibly small if it is below. In order to handle decoding errors, we assume an ideal ARQ protocol such that the unsuccessfully decoded information bits are left in the transmission buffer and shall be re-scheduled for transmission at a later time.

For a given linear beamforming strategy, the resource-allocation policy formed by beam and rate allocation is referred to as a SDMA scheduling policy. A SDMA scheduling policy is a function of CSIT and of the state of the transmitter queues. Let denote the buffer size of queue $k$ by $S_k(t)$ expressed in bit. We assume that at the beginning of each slot $t$, both

a CSIT signal $\boldsymbol{\alpha}(t) = (\boldsymbol{\alpha}_1^T(t), \ldots, \boldsymbol{\alpha}_K^T(t))^T \in \mathbb{C}^{K \times M}$ and the queue buffer states $\mathbf{S}(t) = \{S_1(t), \ldots, S_K(t)\} \in \mathbb{R}_+^K$ are revealed to the transmitter. Considering the analog feedback link, we define the CSIT $\boldsymbol{\alpha}_k(t)$ of user $k$ as the MMSE prediction of $\mathbf{h}_k(t)$ given the observation (specified later). The arrival process of queue $k$ is denoted by $A_k(t)$, an ergodic process with *arrival rate* $\lambda_k \triangleq \frac{1}{T}\mathbb{E}[A_k(t)]$ (bit/channel use). In the following, we restrict ourselves to a stationary SDMA scheduling policy for simplicity. A SDMA stationary scheduling policy is defined by a sequence of functions $\mathcal{F} = (\mathbf{G}(t), \mathbf{R}(t))$ such that

$$\mathbf{G}(t) : \mathbb{C}^{K \times M} \times \mathbb{R}_+^K \mapsto \mathbb{C}^{M \times K} \tag{4.6}$$

$$\mathbf{R}(t) : \mathbb{C}^{K \times M} \times \mathbb{R}_+^K \mapsto \mathbb{R}_+^K \tag{4.7}$$

for any $t$ and $\mathbf{R}(t) = (R_1(t), \ldots, R_K(t))$ denotes the set of transmission rates. Notice that the scheduling policy does not depend on $t$ from stationarity. The rate set is given as a function of CSIT and of the resulting $\mathbf{G}$. Since the conditional outage probability given the CSIT $\boldsymbol{\alpha} = \mathbf{a}$ and $\mathbf{G}$, i.e., $\mathbb{P}(\overline{R}_k \leq R | \boldsymbol{\alpha} = \mathbf{a}, \mathbf{G})$ is difficult to evaluate, the rate allocation that maximizes the instantaneous outage rate can not be obtained easily. Therefore, we introduce a rate margin $\rho \in (0, 1)$ so as to assign a fraction of the estimated rate given the current CSIT value $\boldsymbol{\alpha} = \mathbf{a}$ and $\mathbf{G}$, which enables the rate allocation to be conservative. The rate allocation to user $k$ in slot $t$ is given by

$$R_k(t) = \rho \hat{R}_k(t) \tag{4.8}$$

where $\hat{R}_k(t)$ denotes the estimated rate as a function of $\boldsymbol{\alpha} = \mathbf{a}$ and $\mathbf{G}$ given by

$$\hat{R}_k(t) = \log_2 \left( 1 + \frac{\psi_{k,k}(t)}{N_0^{\text{down}} \text{tr}(\mathbf{G}^H(t)\mathbf{G}(t)) + \sum_{j \neq k} \psi_{k,j}(t)} \right) \tag{4.9}$$

where $\psi_{k,j}(t) = \mathbf{g}_j^H(t) \boldsymbol{\Sigma}_k(t) \mathbf{g}_j(t)$ denotes the estimated cross-talk between channel $k$ and beam $j$ and $\boldsymbol{\Sigma}_k(t) = \mathbb{E}[\mathbf{h}_k(t)\mathbf{h}_k^H(t) | \boldsymbol{\alpha}(t)]$ is the conditional covariance matrix of user $k$ given the current CSIT value. We remark that this rate allocation is not optimal but heuristic.

For a SDMA scheduling policy, we let $\overline{R}_k(t) = \log_2(1 + \mathsf{SINR}_k(t))$ denote the supremum of the coding rates supported by the channel for user $k$ in slot $t$. Notice that $\overline{R}_k(t)$ depends on the instantaneous channel realization, and is therefore a random variable. Under the assumptions given above, the queue

buffer states evolve in time according to the stochastic difference equation

$$S_k(t+1) = \left[S_k(t) - TR_k(t)\, 1\{R_k(t) < \overline{R}_k(t)\}\right]_+ + A_k(t) \tag{4.10}$$

for all $k = 1, \ldots, K$, where $[\cdot]_+ \stackrel{\Delta}{=} \max\{\cdot, 0\}$. The presence of the indicator function $1\{R_k(t) < \overline{R}_k(t)\}$ is due to the ARQ mechanism illustrated before: only if decoding is successful, i.e., if the scheduled rate $R_k(t)$ is indeed achievable, the corresponding information bits are removed from the transmission buffer.

In order to define stability, we follow [55] and define the buffer overflow function $g_k(S) = \limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} 1\{S_k(\tau) > S\}$. We say that the system is stable if $\lim_{S\to\infty} g_k(S) = 0$ for all $k = 1, \ldots, K$. The stability region $\Omega$ of a SDMA scheduling policy is the set of all vectors of bit arrival rates $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$ such that there exists a stationary scheduling policy $\mathcal{F}$ which achieves stability of all $K$ buffers whenever the bit arrival vector is inside the stability region.

## 4.4 Stability region and max-stability policy under linear beamforming

We make the same key assumptions as Chapter 3 that allow us to obtain a very simple characterization of $\Omega$ and a simple SDMA policy that achieves any $\boldsymbol{\lambda} \in \Omega$. We assume that: 1) the channel matrix $\mathbf{H}(t)$, the CSIT signal $\boldsymbol{\alpha}(t)$ and the arrival processes $\{A_k(t) : k = 1, \ldots, K\}$ evolve according to a jointly stationary ergodic Markov process; 2) the arrival processes have $\mathbb{E}[A_k(t)^2] \leq \infty$; 3) for a given signaling strategy, for every sufficiently large $t$, the following Markov chain holds:

$$\boldsymbol{\alpha}_1^{t-1} \to \boldsymbol{\alpha}(t) \to \{\overline{R}_k(t) : k = 1, \ldots, K\} \tag{4.11}$$

In particular, the third condition implies that for all $k$ and sufficiently large $t$, the conditional outage probability given the linear beamforming strategy depends only on the current CSIT value, i.e., it satisfies

$$\mathbb{P}(\overline{R}_k(t) \leq R | \mathbf{G}(t), \boldsymbol{\alpha}_1^t) = \mathbb{P}(\overline{R}_k(t) \leq R | \mathbf{G}(t), \boldsymbol{\alpha}(t)) \tag{4.12}$$

In order to derive the stability region of a given linear beamforming strategy, we have to generalize the stability theory given in Chapter 3. Again $\boldsymbol{\alpha}(t)$

takes on the role of the virtual channel state, and the instantaneous *effective* rate $R_k(t)1\{R_k(t) < \overline{R}_k(t)\}$ is replaced by the conditional outage rate. Since we consider $R_k(t)$ given in (4.8), the conditional outage rate is defined by

$$R_k^{\text{out}}(\mathbf{G}, \mathbf{a}) = \rho \hat{R}_k(\mathbf{G}, \mathbf{a})(1 - \mathbb{P}(\overline{R}_k(\mathbf{G}, \mathbf{a}) \leq \rho \hat{R}_k(\mathbf{G}, \mathbf{a}))) \tag{4.13}$$

This is a function of the current CSIT value $\boldsymbol{\alpha} = \mathbf{a}$ and $\mathbf{G}$. [1] Unfortunately, the outage rate function $R_k^{\text{out}}(\mathbf{G}, \mathbf{a})$ under a linear beamforming strategy is not generally concave in $\mathbf{G}$, we can not apply the Theorem 1 [55] directly. However, using the necessary and sufficient condition of a Markov chain in 3.3 of Chapter 3, we have

**Theorem 1 : stability region** The stability region $\Omega$ of stationary SDMA scheduling policies is given by

$$\Omega = \text{coh} \bigcup_{\mathcal{F}} \{\lambda_k \leq \mathbb{E}[R_k^{\text{out}}(\mathbf{G}, \boldsymbol{\alpha})]\} \tag{4.14}$$

where "coh" denotes convex-hull operation and the union is over all possible stationary SDMA policies. □

**Proof :** see Appendix 4.A.

The boundary surface $\partial\Omega$ of the stability region is given by $\boldsymbol{\lambda}$ solution of

$$\boldsymbol{\lambda} = \arg \max_{\boldsymbol{\lambda}' \in \Omega} \sum_{k=1}^{K} \theta_k \lambda_k' \tag{4.15}$$

for any non-negative weight vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K) \in \mathbb{R}_+^K$ such that $\sum_k \theta_k = 1$. The boundary surface can be obtained by maximizing the weighted average outage rate rate over all $\mathcal{F}$ for each $\boldsymbol{\theta}$. Let define $\boldsymbol{\lambda}(\boldsymbol{\theta})$ the boundary point that corresponds to a given weight vector $\boldsymbol{\theta}$. This is given by

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \{\lambda_k(\boldsymbol{\theta}) \leq \mathbb{E}\left[R_k^{\text{out}}(\mathbf{G}, \boldsymbol{\alpha})\right]\} \tag{4.16}$$

Finally the boundary surface $\partial\Omega$ is obtained by a convex-hull operation over $\boldsymbol{\lambda}(\boldsymbol{\theta})$ by varying $\boldsymbol{\theta}$, given by

$$\partial\Omega = \text{coh} \bigcup_{\boldsymbol{\theta}} \boldsymbol{\lambda}(\boldsymbol{\theta}) \tag{4.17}$$

[1] Here and in the following we let $\boldsymbol{\alpha}$ denote a random vector with the same distribution of $\boldsymbol{\alpha}(t)$, which is independent of $t$ because of stationarity.

For a given weight vector $\boldsymbol{\theta} \in \mathbb{R}_+^K$ and for each $\boldsymbol{\alpha} = \mathbf{a}$, the solution of (4.15) is given by

$$\hat{\mathcal{F}}_{\boldsymbol{\theta}}(\mathbf{a}) = \arg\max_{\mathcal{F}} \sum_k \theta_k R_k^{\text{out}}(\mathbf{G}, \mathbf{a})$$

For later use, we also define the instantaneous stability region $\Omega(\boldsymbol{\alpha})$ as the set of arrival rates such that there exists a stationary policy $\mathcal{F}$ for which the system is stable for a given CSIT $\boldsymbol{\alpha}$. This is given by

$$\Omega(\boldsymbol{\alpha}) = \bigcup_{\mathcal{F}} \{\lambda_k \leq R_k^{\text{out}}(\mathbf{G}, \boldsymbol{\alpha})\} \tag{4.18}$$

For a given $\boldsymbol{\lambda} \in \Omega$ there exists some memoryless stationary policy $\mathcal{F}_{\boldsymbol{\lambda}}$ that stabilizes the system. However, in order to determine $\mathcal{F}_{\boldsymbol{\lambda}}$ the a priori knowledge on the arrival rate vector $\boldsymbol{\lambda}$ is generally required. Since this might not be available in practice, it is interesting to have an adaptive policy that achieves stability for all $\boldsymbol{\lambda} \in \Omega$ *adaptively* without prior knowledge of the arrival rates [55, 56]. We refer such an adaptive policy to max-stability adaptive policy. Under the assumption that the arrival process, the channel process, and the CSIT evolve according to a jointly stationary ergodic Markov chain, we have

**Theorem 2 : max-stability adaptive policy** The stationary SDMA scheduling policy choosing the beamforming matrix $\mathbf{G}$

$$\hat{\mathcal{F}} = \arg\max_{\mathcal{F}} \sum_k \theta_k S_k R_k^{\text{out}}(\mathbf{G}, \mathbf{a}) \tag{4.19}$$

stabilizes the system whenever the arrival rate vector is inside $\Omega$. □

**Proof** See 4.B.

## 4.5 Proposed method for user selection and linear beamforming

In this section, we propose a novel method for user selection and linear beamforming that intends to stabilize the system whenever the arrival rate vector is inside the stability region given in (4.14). The proposed scheme can be regarded as a practical version of the max-stability adaptive policy derived in the previous section.

We assume that the channel vectors $\mathbf{h}_k(t) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$ are mutually statistically independent for different index $k$ and i.i.d. for different antennas. We consider that $\{\mathbf{h}_k(t)\}$ evolves from slot to slot according to a stationary ergodic Gauss-Markov process of order 1, given by

$$\mathbf{h}_k(t) = r\mathbf{h}_k(t-1) + \boldsymbol{\nu}_k(t) \tag{4.20}$$

where $\boldsymbol{\nu}_k(t) \sim \mathcal{N}_{\mathbb{C}}(0, (1-r^2)\mathbf{I}_M)$ is an i.i.d. process. We make the assumption that the receivers can estimate *exactly* their channel vector and send it back over the analog feedback link without quantizing and coding. We suppose that the feedback link is noisy link given by

$$\mathbf{z}_k(t) = \mathbf{h}_k(t) + \mathbf{n}_k(t)$$

where $\mathbf{n}_k(t) \sim \mathcal{N}_{\mathbb{C}}(0, N_0^{\text{up}}\mathbf{I}_M)$ is an i.i.d. process. The uplink SNR is equal to $1/N_0^{\text{up}}$. The sequences $\{\boldsymbol{\nu}_k(t)\}$ and $\{\mathbf{n}_k(t)\}$ are uncorrelated. Due to a fixed delay of $d$ slots in the feedback link, the CSIT $\boldsymbol{\alpha}_k(t+d)$ of user $k$ in slot $t+d$ is defined by the MMSE *prediction* of $\mathbf{h}_k(t+d)$ given the delayed and noisy observation up to slot $t$. The CSIT $\boldsymbol{\alpha}_k(t+d)$ is given by

$$\boldsymbol{\alpha}_k(t+d) = \mathbb{E}[\mathbf{h}_k(t+d)|\mathbf{z}_k(t), \ldots, \mathbf{z}_k(0)] = r^d\hat{\mathbf{h}}_k(t|t) \tag{4.21}$$

where $\hat{\mathbf{h}}_k(t|t) = \mathbb{E}[\mathbf{h}_k(t)|\mathbf{z}_k(t), \ldots, \mathbf{z}_k(0)]$ denotes the MMSE *estimation* of $\mathbf{h}_k(t)$ given the observation up to slot $t$ and this can be obtained recursively via Kalman filter [96] as we explain in 4.C. Notice that this CSIT model as well as the underlying channel model is to validate the Markov assumptions mentioned in the previous section. As a result, $\mathbf{h}_k(t)$ and $\boldsymbol{\alpha}_k(t)$ are jointly Gaussian with covariance

$$\begin{bmatrix} \mathbf{I} & (1-\sigma_e^2)\mathbf{I} \\ (1-\sigma_e^2)\mathbf{I} & (1-\sigma_e^2)\mathbf{I} \end{bmatrix} \tag{4.22}$$

where $\sigma_e^2 = \mathbb{E}[|\mathbf{h}_k(t) - \boldsymbol{\alpha}_k(t)|^2]$ denotes the *prediction* MMSE for the delayed observation, given by

$$\sigma_e^2 = r^{2d}\xi_{mmse} + (1-r^2)\sum_{l=0}^{d-1} r^{2l} \tag{4.23}$$

where $\xi_{mmse} = \mathbb{E}[|\mathbf{h}_k(t) - \hat{\mathbf{h}}_k(t|t)|^2]$ denotes the *estimation* MMSE for the delay-free observation, given by

$$\xi_{mmse} = \frac{1}{2\pi}\int_{-\pi}^{\pi} [P_h(e^{jw}) - P_{\text{kalman}}(e^{jw})P_{hz}(e^{jw})]dw \tag{4.24}$$

where $P_h(e^{jw})$, $P_{\text{kalman}}(e^{jw})$, $P_{hz}(e^{jw})$ denotes the power spectrum density (psd) of the fading channel, the psd of the Kalman filter, the psd of correlation between the fading channel and the observation. Noticing that $\{\mathbf{h}_k(t)\}$ and $\{\mathbf{n}_k\}$ are uncorrelated, we have $P_{hz}(e^{jw}) = P_h(e^{jw})$. $P_h(e^{jw})$, $P_{\text{kalman}}(e^{jw})$ are respectively given by

$$\begin{aligned} P_h(e^{jw}) &= \frac{1-r^2}{(1-re^{-jw})(1+re^{jw})} \\ P_{\text{kalman}}(e^{jw}) &= \frac{1}{\sigma_z^2 Q(e^{jw})}\left[\frac{P_h(e^{jw})}{Q(e^{-jw})}\right]_+ \end{aligned}$$

where $\sigma_z^2 Q(e^{jw})$ is obtained by the factorization on the psd of the observation by letting $P_z(e^{jw}) = N_0^{\text{up}} + P_h(e^{jw}) = \sigma_z^2 Q(e^{jw})Q(e^{-jw})$. Finally, the conditional covariance of user $k$ at slot $t$ given $\boldsymbol{\alpha}_k(t)$ is

$$\boldsymbol{\Sigma}_k(t) = \mathbb{E}[\mathbf{h}_k(t)\mathbf{h}_k^H(t)|\boldsymbol{\alpha}_k(t)] = \boldsymbol{\alpha}_k(t)\boldsymbol{\alpha}_k^H(t) + \sigma_e^2\mathbf{I}_M.$$

In order to exploit the analog feedback, the amount of feedback required per user is $M$ complex values and thus reasonable.

The main challenge is to find for each CSIT $\boldsymbol{\alpha}$ the beamforming matrix $\mathbf{G}$ that maximizes the weighted sum-rate. As remarked in [32], the sum rate maximization over $\mathbf{G}$ is not easily amenable because it falls into a non-convex optimization problem.

### 4.5.1 Perfect CSIT

When perfect CSIT is available, i.e. the underlying channel is static and the observation is noise and delay free, the feasible rate is given by $R_k^{\text{out}}(t) = R_k(t) = \log_2(1 + \mathsf{SINR}_k(t))$. In the following we remove the slot index $t$ for notation simplicity. Given the perfect CSIT $\boldsymbol{\alpha}(t) = \mathbf{H}(t)$, a linear beamforming matrix that maximizes the weighted sum-rate can be found by solving the following problem.

$$\max_{\mathbf{G}} \sum_k \theta_k \log_2\left(1 + \frac{\psi_{k,k}}{N_0^{\text{down}}\text{tr}(\mathbf{G}^H\mathbf{G}) + \sum_{j\neq k}\psi_{k,j}}\right) \tag{4.25}$$

where we let $\psi_{k,j} = \mathbf{g}_j^H\boldsymbol{\Sigma}_k\mathbf{g}_j$ and the sum is over the set of the users with positive weight, i.e. $\mathcal{K} = \{k : \theta_k > 0\}$ with $K_0 = |\mathcal{K}|$. Notice that for perfect CSIT case, since $\boldsymbol{\Sigma}_k = \mathbf{h}_k\mathbf{h}_k^H$, we can further write $\psi_{k,j} = |[\mathbf{HG}]_{k,j}|^2$.

In [34], Stojnic et al. proposed a heuristic approach to solve the sum-rate maximization problem for a $M \times M$ system where the transmitter with $M$ transmit antennas transmits to $M$ users with a single antenna each. We will extend the Stojnic et al. algorithm to the weighted sum-rate maximization for a $M \times K_0$ system. Let denote the argument of (4.25) by $f(\mathbf{G})$. Differentiating $f(\mathbf{G})$ with respect to $\mathbf{g}_k$, we obtain the following necessary condition for all $k \in \mathcal{K}$ so that $\mathbf{G}$ should be a stationary point.

$$\frac{\partial f}{\partial \mathbf{g}_k} = \frac{\theta_k}{d_k}\mathbf{\Sigma}_k \mathbf{g}_k - N_0^{\text{down}} \sum_j [\mathbf{D}]_{j,j}\mathbf{g}_k - \sum_j [\mathbf{D}]_{j,j}\mathbf{\Sigma}_j \mathbf{g}_k = \mathbf{0} \tag{4.26}$$

where $[\mathbf{D}]_{j,j}$ denotes the $j$-th diagonal element of $\mathbf{D}$ given by

$$\mathbf{D} = \text{diag}\left(\frac{\theta_k \psi_{k,k}}{d_k(d_k + \psi_{k,k})}\right) \tag{4.27}$$

and $d_k = N_0^{\text{down}}\text{tr}(\mathbf{G}^H\mathbf{G}) + \sum_{j \neq k} \psi_{k,j}$ denotes the denominator of the estimated SINR of user $k$. Noticing that under perfect CSIT we have $\mathbf{\Sigma}_j = \mathbf{h}_j\mathbf{h}_j^H$ for all $j$, (4.26) can be further written by

$$\frac{\theta_k}{d_k}\mathbf{h}_k^*[\mathbf{HG}]_{k,k} - N_0^{\text{down}} \sum_j [\mathbf{D}]_{j,j}\mathbf{g}_k - \sum_j [\mathbf{D}]_{j,j}\mathbf{\Sigma}_j \mathbf{g}_k = \mathbf{0} \tag{4.28}$$

By letting (4.28) for any $k \in \mathcal{K}$, we have

$$\mathbf{G} = \left(\mathbf{H}^H\mathbf{D}\mathbf{H} + N_0^{\text{down}}\text{tr}(\mathbf{D})\mathbf{I}_M\right)^{-1}\mathbf{H}^H\mathbf{\Delta} \tag{4.29}$$

where we let

$$\mathbf{\Delta} = \text{diag}\left(\frac{\theta_k[\mathbf{HG}]_{k,k}}{d_k}\right)$$

The Stojnic et.al algorithm can be suitably applied to a system with any $K_0$ while taking into account the weights $\{\theta_k\}$. The following summaries the algorithm to maximize the weighted sum rate that we apply every slot $t$ by replacing $\theta_k(t)$ with the normalized buffer queue size $S_k(t)$.

**Modified Stojnic algorithm**

1. Initialize $\mathbf{D}^{(0)} = \mathbf{I}_{K_0}, \mathbf{\Delta}^{(0)} = \mathbf{I}_{K_0}$
2. At iteration $n$, update $\mathbf{G}^{(n)}$ such that

$$\mathbf{G}^{(n)} = \left(\mathbf{H}^H\mathbf{D}^{(n-1)}\mathbf{H} + N_0^{\text{down}}\text{tr}(\mathbf{D}^{(n-1)})\mathbf{I}_M\right)^{-1}\mathbf{H}^H\mathbf{\Delta}^{(n-1)} \tag{4.30}$$

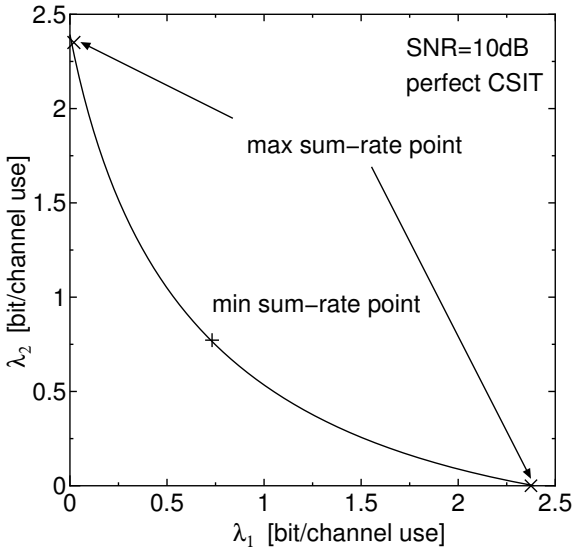


Figure 4.2: Two user instantaneous stability region

3. Update $\mathbf{D}^{(n)}, \mathbf{\Delta}^{(n)}$ by

$$\begin{aligned} \mathbf{\Psi}^{(n)} &= \mathbf{H}\mathbf{G}^{(n)} \\ d_k^{(n)} &= \mathrm{tr}(\mathbf{G}^{(n)^H}\mathbf{G}^{(n)}) + \sum_{j\neq k} \phi_{k,j}^{(n)} \quad \forall k \\ \mathbf{D}^{(n)} &= \mathrm{diag}\left(\frac{\theta_k \psi_{k,k}^{(n)}}{d_k^{(n)}(d_k^{(n)} + \psi_{k,k}^{(n)})}\right), \quad \mathbf{\Delta}^{(n)} = \mathrm{diag}\left(\frac{\theta_k [\mathbf{H}\mathbf{G}^{(n)}]_{k,k}}{d_k}\right) \end{aligned} \tag{4.31}$$

4. Repeat $n = n + 1$

After convergence, $\mathbf{G}$ has the columns proportional to the normalized dual uplink MMSE filter with uplink power

$$q_k = \frac{1}{N_0^{\mathrm{down}} \mathrm{tr}(\mathbf{D})} \frac{\theta_k \psi_{k,k}}{d_k(d_k + \psi_{k,k})}, \quad k = 1, \ldots, K$$

satisfying the same total power constraint $\sum_k q_k \leq 1$. Hence, the algorithm converges to a point on the boundary of the instantaneous stability region. However, the convergence to the global optimum is not guaranteed. Here is

an example that shows that the algorithm can converge to the minimum sum-rate point instead of the maximum sum-rate point. Figure 4.2 shows the two-user instantaneous stability region given the CSIT $\boldsymbol{\alpha} = \mathbf{H} = [2 - 0.5; 2 - 0.5]$ where the SNR is set to be $1/N_0^{\text{down}} = 0$ [dB] and the weight vector is set to $\theta_1 = \theta_2 = 1/2$. Under this channel matrix whose rank is 1, the Stojnic et al. algorithm finds the minimum sum-rate point instead of the maximum sum-rate point. Nevertheless, numerical experiences show that the algorithm improves the objective function when $\mathbf{H}$ has a full rank.

At each iteration of the modified Stojnic algorithm, the following computations are required. In (4.30), the matrix inverse of a $M \times M$ square matrix $\mathcal{O}(M^3)$ and $MK_0(M + K_0 + 1)$ complex multiplications are necessary. In (4.31), $2MK_0^2$ complex multiplications are required. This results in roughly $\mathcal{O}(K^2 + M^3)$ computations per iteration. Denoting the number of iterations by $N_{\text{iter}}$, we need $\mathcal{O}(N_{\text{iter}}(K_0^2 + M^3))$ computations in total. The numerical experiments show that we need roughly $N_{\text{iter}} = 15$ iterations so that the algorithm should converge.

### 4.5.2 Imperfect CSIT

When the transmitter has imperfect CSIT, the outage rate $R_k^{\text{out}}(\mathbf{G}, \mathbf{a})$ in (4.13) can not be easily computed because the conditional outage probability depends on the joint pdf of $\mathbf{H}, \mathbf{a}$ that is a function of $\mathbf{G}$ and $\mathbf{a}$. Hence, we consider the following maximization problem that can be easily solved by the transmitter.

$$\max_{\mathbf{G}} \sum_k \theta_k \hat{R}_k(\mathbf{G}, \mathbf{a}) = \max_{\mathbf{G}} \sum_k \theta_k \log_2 \left( 1 + \frac{\psi_{k,k}}{N_0^{\text{down}} \text{tr}(\mathbf{G}^H \mathbf{G}) + \sum_{j \neq k} \psi_{k,j}} \right) \tag{4.32}$$

where $\psi_{k,j} = \mathbf{g}_j^H \boldsymbol{\Sigma}_k \mathbf{g}_j$ denotes cross-talk between conditional covariance of $k$ and beam $j$. Then, using the $\mathbf{G}$ solution of (4.32) we propose the following rate allocation. Let $B$ denote the number of columns with positive power in $\mathbf{G}$. $B$ is the number of beams (or streams) corresponding to the number of simultaneously served users given the CSIT $\boldsymbol{\alpha} = \mathbf{a}$. We let the optimal rate margin $\rho_{M,B}^{\star}$ for a system with $M$ transmit antennas and $B$ beams such that

$$\rho_{M,B}^{\star} = \arg \max_{0 < \rho < 1} \sum_{k=1}^{B} \mathbb{E} \left[ \rho \hat{R}_k(\mathbf{G}, \mathbf{a}) \mathbf{1}\{ \rho \hat{R}_k(\mathbf{G}, \mathbf{a}) < \overline{R}_k(\mathbf{G}, \mathbf{a}) \} \right] \tag{4.33}$$

We allocate the following rate to any $k \in \mathcal{K}$

$$R_k(\mathbf{G}, \mathbf{a}) \;=\; \rho^{\star}_{M,B} \hat{R}_k(\mathbf{G}, \mathbf{a}) \tag{4.34}$$

The weighted sum-rate maximization of (4.32) can be considered as a novel generalization of the approach for perfect CSIT in [34]. The difference with respect to the perfect CSIT case is that the transmitter knows only the statistics of the channel, i.e. the conditional covariance $\mathbf{\Sigma}_k(t)$ for all users $k$. Differentiating the objective function $f(\mathbf{G})$ with respect to $\mathbf{g}_k$, we obtain (4.26) for any $k \in \mathcal{K}$. Therefore, $\mathbf{G}$ is a stationary point of $f(\mathbf{G})$ if the columns of $\mathbf{G}$ satisfy the following necessary condition

$$\mathbf{g}_k = \frac{\theta_k}{d_k} \left[ N_0^{\text{down}} \sum_j [\mathbf{D}]_{j,j} \mathbf{I}_M + \sum_j [\mathbf{D}]_{j,j} \mathbf{\Sigma}_j \right]^{-1} \mathbf{\Sigma}_k \mathbf{g}_k \tag{4.35}$$

By generalizing the Stojnic et al. algorithm [34], we propose a novel iterative algorithm that maximizes the weighted sum rate for a $M \times K_0$ system.

**Linear beamforming (LB) algorithm**

1. Initialize $\mathbf{G}^{(0)} = \frac{\boldsymbol{\alpha}^H}{\sqrt{\text{tr}(\boldsymbol{\alpha}\boldsymbol{\alpha}^H)}}$

2. At iteration $n$, update $\mathbf{G}^{(n)}$ such that for $k \in \mathcal{K}$

$$\mathbf{g}_k^{(n)} = \frac{\theta_k}{d_k^{(n-1)}} \left[ N_0^{\text{down}} \sum_j [\mathbf{D}]_{j,j}^{(n-1)} \mathbf{I}_M + \sum_j [\mathbf{D}]_{j,j}^{(n-1)} \mathbf{\Sigma}_j \right]^{-1} \mathbf{\Sigma}_k \mathbf{g}_k^{(n-1)} \tag{4.36}$$

3. Repeat $n = n + 1$

The initial $\mathbf{G}^{(0)}$ coincides with the matched filtering. Numerical examples show that this choice of initialization allows the algorithm to converge after roughly 15 iterations. The LB algorithm does not guarantee the convergence to the global optimum point as the modified Stojnic algorithm for the perfect CSIT case. Here is an example to show that the stationary point achieved by the LB algorithm corresponds to the minimum sum-rate point. Figure 4.3 shows the two-user instantaneous stability region given the CSIT $\boldsymbol{\alpha} = [2 \;\; -0.5; -0.5 \;\; 2]$ where the downlink SNR is set to be $1/N_0^{\text{down}} = 50$ [dB] and the weight vector is set to $\theta_1 = \theta_2 = 1/2$. Under prediction MMSE $\sigma_e^2 = 0.05$, the covariance of two users are given by $\mathbf{\Sigma}_1 = [4.05 \;\; -1; -1 \;\; 0.30]$

and $\mathbf{\Sigma}_2 = [0.3 \ \ -1; -1 \ 4.05]$. At such a high SNR and under non-perfect CSIT, the optimum point is given by a vertex point of the instantaneous stability region at large probability, i.e. only one user should be served at each slot so that the weighted sum-rate should not be interference-limited.

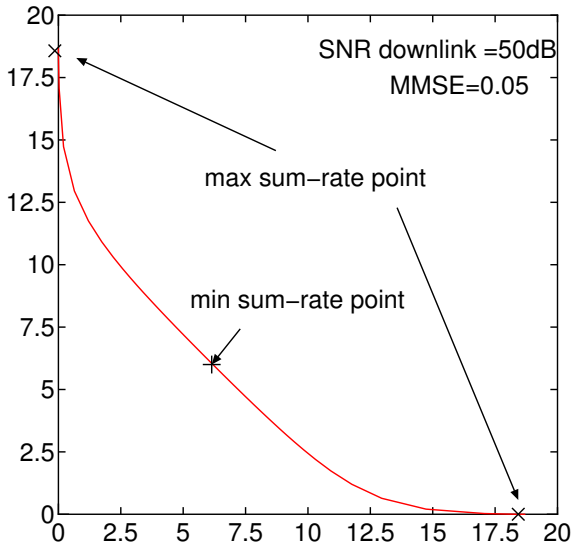


Figure 4.3: Two user instantaneous stability region

In order to avoid the convergence to local optimum points, we will not apply the LB algorithm directly for a $M \times K_0$ system. However, exhaustive search for optimal user selection would require searching for all $\binom{m}{K_0}$ subsets of $m$ out of $K_0$ users for $m = 1, \ldots, M$. The problem of efficient optimal user selection is still open and then we propose the following greedy search algorithm to choose a subset of $1 \leq B \leq M$ users. Let $f_m(\mathcal{S})$ be the weighted sum-rate achieved by the LB algorithm for an arbitrary set of $m$ users $\mathcal{S}$.

**User selection algorithm**

1. Initialize $\mathcal{K}_0(0) = \mathcal{K}$ and $\mathcal{K}_1(0) = \emptyset$

2. For $m = 1, \ldots, M$ repeat

$$k = \arg \max_{j \in \mathcal{K}_0(m-1)} f_m(j, \mathcal{K}_1(m-1)) \tag{4.37}$$
$$\mathcal{K}_1(m) = \mathcal{K}_1(m-1) + \{k\}$$
$$\mathcal{K}_0(m) = \mathcal{K}_0(m-1) - \{k\}$$
$$f_m^\star(\mathcal{K}_1(m)) = \rho_{M,m}^\star f_m(\mathcal{K}_1(m)) \tag{4.38}$$

End

3. Let $B = \arg\max_m f_m^\star(\mathcal{K}_1(m))$ and the final set $\mathcal{K}_1 = \mathcal{K}_1(B)$

where $\rho_{M,m}^\star$ denotes the rate margin optimized for a system with $M$ antennas and $m$ beams defined in (4.33). Notice that for $m = 1$, the algorithm simply chooses the user with the largest weighted rate given by

$$\begin{aligned} k &= \arg\max_{j \in \mathcal{K}} \theta_j \log_2 \left(1 + \frac{\max_{\mathbf{g}_j} \mathbf{g}_j^H \mathbf{\Sigma}_j \mathbf{g}_j}{N_0^{\text{down}}}\right) \\ &= \arg\max_{j \in \mathcal{K}} \theta_j \log_2 \left(1 + \frac{|\boldsymbol{\alpha}_j|^2 + \sigma_e^2}{N_0^{\text{down}}}\right) \end{aligned} \tag{4.39}$$

where we used the fact that $\mathbf{g}_j$ with unit norm is given by the vector associated to the maximum eigen value of the covariance $\mathbf{\Sigma}_j = \boldsymbol{\alpha}_j^* \boldsymbol{\alpha}_j^T + \sigma_e^2 \mathbf{I}$, namely a scaled $\boldsymbol{\alpha}_j^*$. The complexity of the total algorithm is as follows. Denote $N_{\text{iter}}$ the number of iterations for the LB algorithm. Since we need $(K_0 - m + 1)$ computations of the complexity $\mathcal{O}(N_{\text{iter}} m^3)$ to choose a subset of $m$ users for each $m = 2, \ldots, M$, the total complexity of $\mathcal{O}(N_{\text{iter}} M^3 K_0)$ is required. In practice, $K_0$ is always smaller than $K$ since the arrival rate of interest should be much smaller than the boundary point of the stability region. By setting the number of iterations to be $N_{\text{iter}} = 15$, the total complexity is reasonably low.

## 4.6 Numerical Results

**Simulation setting** We considered mutually independent arrival processes such that $A_k(t) = \sum_{j=1}^{M_k(t)} b_{k,j}(t)$, where $M_k(t)$ is an i.i.d. Poisson distributed sequence that counts the number of packets arrived to the $k$-th buffer at the

beginning of slot $t$ and $\{b_{k,j}(t)\}$ are i.i.d. exponentially distributed random variables expressing the number of bits per packet. We take $\mathbb{E}[b_{k,j}(t)] = T$ ($T = 2000$ in our simulations), so that $\lambda_k$ coincides with the average number of packets arrived in a slot ($T$ channel uses). We consider a first order Gauss-Markov process where the coefficient in (4.20) is matched to Jakes' autocorrelation model [92], i.e. $r = J_0(2\pi f_D T_{\text{slot}})$ where $f_D$ denotes the one-sided Doppler bandwidth normalized to the signal bandwidth and $T_{\text{slot}}$ denotes the duration of a slot. Inspired by the HDR system [2], we let $T_{\text{slot}} = 1.67$ msec, and the feedback delay $d = 2$ slots. Both the average SNR of the downlink and the feedback link is set equal to $1/N_0^{\text{down}} = 1/N_0^{\text{up}} = 10$ dB. Under this setting, the mobile speeds $v = 0$, 7.2, 12.0, 19.4, 31.5, 42.7, 56.5 km/h yield a channel prediction MMSE $\sigma_e^2 = 0.00$, 0.05, 0.10,0.20, 0.40, 0.60, 0.80 respectively.

**Maximum sum-rate vs. number of users** First, we evaluate the performance of our proposed algorithm for user selection and linear beamforming in terms of maximum sum-rate. By setting $\theta_k(t) = 1/K$ for all $k$ and $t$, we compute by simulation the maximum sum-rate given by

$$\mathbb{E}\left[\sum_{k=1}^{K} \rho^{\star}_{M,B} \hat{R}_k \mathbf{1}\{\rho^{\star}_{M,B} \hat{R}_k(\mathbf{G}, \boldsymbol{\alpha}) < \overline{R}_k(\mathbf{G}, \boldsymbol{\alpha})\}\right]$$

Figures 4.4,4.5 and 4.6 show the maximum sum-rate vs. the number of users $K$ for different prediction MMSE values $\sigma_e^2 = 0.0, 0.05, 0.40$ respectively. $M = 1$ refers to the performance of a single-antenna TDMA system. When CSIT is perfect or almost perfect, multiplexing gain, power gain and multiuser diversity gain are full exploited independently of the number of users in the system. Under an equal weight setting and $K >> M$, the algorithm finds the best set of $M$ users whose channels are quasi-orthogonal to maximize the sum-rate. As the quality of CSIT gets worse, multiplexing gain decreases significantly. With $M = 2$, not only multiplexing gain but also multiuser diversity gain decreases for $\sigma_e^2 = 0.40$. Figure 4.7 shows the averaged number of beams $\mathbb{E}[B(t)]$ assigned by the algorithm as a function of the number of users for $\sigma_e^2 = 0.05, 0.40$. When CSIT is almost perfect ($\sigma_e^2 = 0.05$), full multiplexing gain is achieved even for a small number of users. It appears clearly that the averaged number of beams depends on $K$ as well as $\sigma_e^2$ and that multiplexing gain can be recovered by a large number of users even for moderate quality of feedback ($\sigma_e^2 = 0.40$).

It is worth comparing the performance of our proposed scheme with STC and opportunistic beamforming considered in Chapter 3. Comparing with Figures 3.2,3.3, our proposed scheme outperforms both STC and opportunistic $M$-beamforming for any $K$ and for any quality of CSIT. Even under perfect CSIT and $K >> M$, our linear beamforming scheme can take more advantage than opportunistic $M$-beamforming by allocating high power to the user at the peak of its fading gain (power gain). On the other hand, for moderate quality of CSIT, our proposed scheme can still achieve multiuser diversity gain, while STC suffers from the channel-hardening effect.

**Two-user stability region** Figure 4.8 shows the two-user stability region achieved by the proposed linear beamforming scheme with $M = 2$ transmit antennas for $\sigma_e^2 = 0.0, 0.05$. $M = 1$ refers to a single-antenna TDMA system. As observed from the figure, our proposed scheme improves the stability region with respect to $M = 1$ thanks to accurate beamforming. On the contrary, opportunistic beamforming with two beams decreases the stability region with respect to $M = 1$ even under perfect CSIT as seen in Figure 3.5, because random beamforming is not efficient with such a small number of users.

**Maximum sum-rate vs. SNR** Figure 4.9 shows the sum-rate performance of different downlink strategies versus SNR for two mobile speeds $v = 0, 7.2$ km/h. We have a system with $K = 20$ users and $M = 4$ transmit antennas. $v = 0$ km/h is considered for the sake of comparison with dirty-paper coding (DPC) that requires perfect CSIT. Under perfect CSIT, the feedback link is noise- and delay-free and SNR on the figure denotes the downlink SNR given by $1/N_0^{\text{down}}$. On the other hand, for $v = 7.2$ km/h we let the downlink and the uplink SNR to be equal, i.e. $1/N_0^{\text{down}} = 1/N_0^{\text{up}}$. Notice that as the uplink SNR increases, the prediction MMSE $\sigma_e^2$ decreases and the quality of CSIT improves.

We briefly review DPC and TDMA downlink strategies. DPC is the capacity-achieving strategy in MIMO broadcast channels [27] and achieves the following maximum sum rate under perfect CSIT [32].

$$\mathbb{E}\left[\max_{\sum_k p_k \le 1} \log_2 \det\left(\mathbf{I}_M + \frac{1}{N_0^{\text{down}}}\sum_{k=1}^{K} p_k \mathbf{h}_k \mathbf{h}_k^H\right)\right] \tag{4.40}$$

TDMA allocates the whole power to the user with the largest eigenvalue associated to its covariance matrix. As seen in (4.39), this corresponds to serving the user with largest fading coefficient under an equal weight condition. The resulting maximum sum rate is given by

$$\mathbb{E}\left[\max_k \rho^{\star}_{M,1} \log_2\left(1+\frac{|\boldsymbol{\alpha}_k|^2+\sigma_e^2}{N_0^{\text{down}}}\right)\right] \tag{4.41}$$

where the rate margin optimized for a system of $M$ antennas and 1 beam is taken into account under non-perfect CSIT.

The sum-rate of our proposed scheme increases linearly in SNR as the optimum sum-rate achieved by DPC under perfect CSIT. Moreover, our proposed scheme performs only 2dB worse than DPC for a given sum rate. For $v = 7.2$ km/h, our proposed linear beamforming scheme achieves higher sum-rate than TDMA by exploiting multiplexing gain until $\mathsf{SNR} = 40$ dB. When the prediction MMSE almost disappears (CSIT becomes almost perfect), our scheme coincides with the TDMA strategy so that the sum-rate could grow linearly in SNR. At such a high SNR region, linear beamforming with multiple beams becomes interference-limited. Figure 4.10 illustrates the evolution of the averaged number of beams as a function of SNR. We observe that the averaged number of beams is four (full multiplexing gain) between 10dB and 40dB and it converges to 1 at 60dB.

**Average delay vs. sum arrival rate** We evaluate the averaged delay $\mathbb{E}[D_k]/K = \mathbb{E}[S_k]/(TK\lambda_k)$ expressed in slot for a 20-user system under symmetric traffic condition $\lambda_1 = \cdots = \lambda_{20}$. The mobile speed is set to $v = 7.2$ km/h ($\sigma_e^2 = 0.05$). Figure 4.11 shows the averaged delay performance of our proposed scheme for $M = 2, 4$. For $M = 1$, the performance of a single-antenna TDMA system is given as a reference. Interestingly, our proposed linear beamforming scheme seems to stabilize the system until the sum arrival rate achieves the boundary of the stability region, although our scheme can be seen as a practical version of the max-stability policy. Notice that under an equal weight setting, the boundary point of the stability region is the maximum sum rate point given in Figure 4.5. Our proposed scheme with $M = 4$ decreases the average delay significantly compared to $M = 1$ for a fixed arrival rate and can stabilize the system for the sum arrival rate twice as large as $M = 1$. By increasing $M$, the queues become better and better balanced thanks to multiplexing gain and power gain, i.e. high power

can be allocated to the user with a large queue. The additional power gain enables our proposed scheme to fully exploit the spatial degrees of freedom in a queued multiuser downlink.

**Average delay vs. mobile speed** We evaluate the merit of our proposed LB scheme compared to opportunistic $M$-beamforming and STC over the time-varying channel. We consider the symmetric arrival with 50 users and let the sum arrival rate to be 2.5 bit/channel use. Figures 4.12, 4.14, and 4.13 show the average delay vs. the mobile speed for our proposed scheme, STC, and opportunistic $M$-beamforming respectively. In all these figures, $M = 1$ refers to a single-antenna TDMA system.

For static to slow channels (up to $v = 20$ km/h), opportunistic $M$-beamforming and our proposed scheme outperform STC. The proposed scheme achieves the smallest average delay because accurate beams and powers are allocated to the users with large weighted rates by exploiting both multiplexing gain and power gain. On the other hand, the gain with opportunistic $M$-beamforming comes mainly from the fact that multiple antennas induce artificially i.i.d. varying fading channel from one slot to another.

As $v$ increases, opportunistic $M$-beamforming becomes strongly interference-limited because the users are assigned the wrong beams at large probability. Therefore, the system becomes unstable at much smaller speed than the other schemes. On the contrary, our proposed scheme allocates the whole power to the user with the largest weighted sum when $v$ is larger than 31.5 km/h, 42.7 km/h for $M = 2$, $M = 4$ respectively. Compared to STC that also operates in TDMA mode, our proposed scheme achieves better average delay because of higher outage rate.

## 4.7 Conclusions

We investigated the use of multiple transmit antennas based on linear beamforming in MIMO broadcast channels, by taking into account random packet arrivals and non-perfect CSIT due to a delay over a feedback link. We considered a quantization-wise optimal "analog feedback", namely, each user terminal sends back its estimated channel vector without quantizing and coding over the feedback link. Then, we derived the stability region under stationary policies based on linear beamforming and max-stability stationary policies that stabilize the system whenever the arrival rate is inside the

stability region. Since such policies are difficult to implement in practice, we proposed a novel heuristic algorithm for user selection and linear beamforming that maximizes the weighted sum-rate given CSIT.

Under perfect CSIT, we observed that our proposed scheme achieves near DPC performance by exploiting multiplexing gain, power gain and multiuser diversity gain. Furthermore, the sum rate of our proposed scheme increases linearly in SNR exactly as DPC. Under non-perfect CSIT, our proposed scheme adjusts the number of beams so that it should not be interference-limited.

We evaluated the relative merit of our linear beamforming scheme and the previously considered schemes, space-time coding and opportunistic $M$-beamforming, over the time-varying channel. At a given feedback delay, it was found that our proposed scheme achieves the best average delay over any range of mobile speed. For slow varying channels, our proposed scheme can exploit multiplexing gain and power gain to balance the buffer sizes. For faster channels, while the opportunistic $M$-beamforming becomes strongly interference-limited and then becomes unstable, the proposed scheme tends to allocate the whole power to the user with the largest weighted rate at each slot. Compared to STC that also operates in TDMA mode, our proposed scheme achieves better average delay because of the larger outage rate.

We can conclude that our scheme offers an attractive solution for a next generation of high-speed downlink system such as HSDPA Release 6 or 1xEV-DO Rev.A.

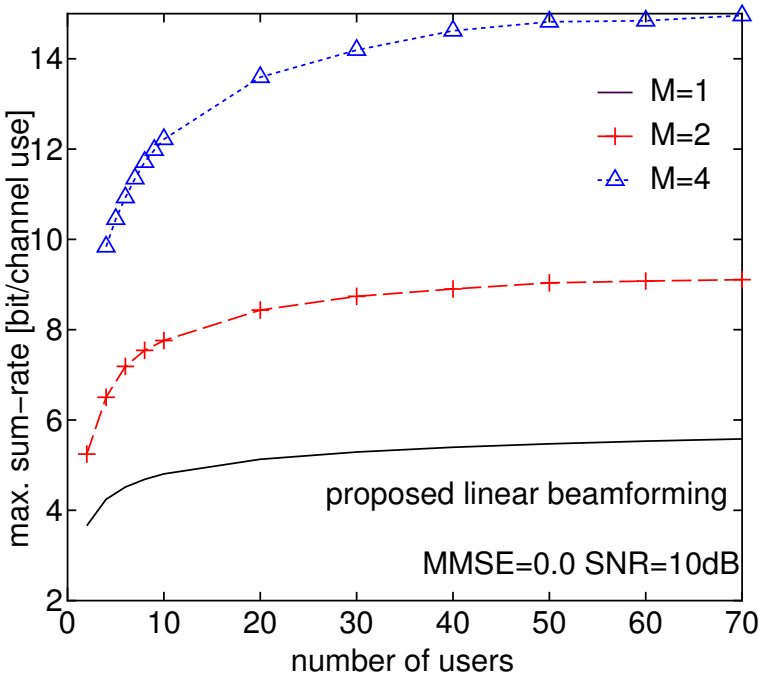


Figure 4.4: Sum rate vs. $K$ ($\sigma_e^2 = 0.0$)

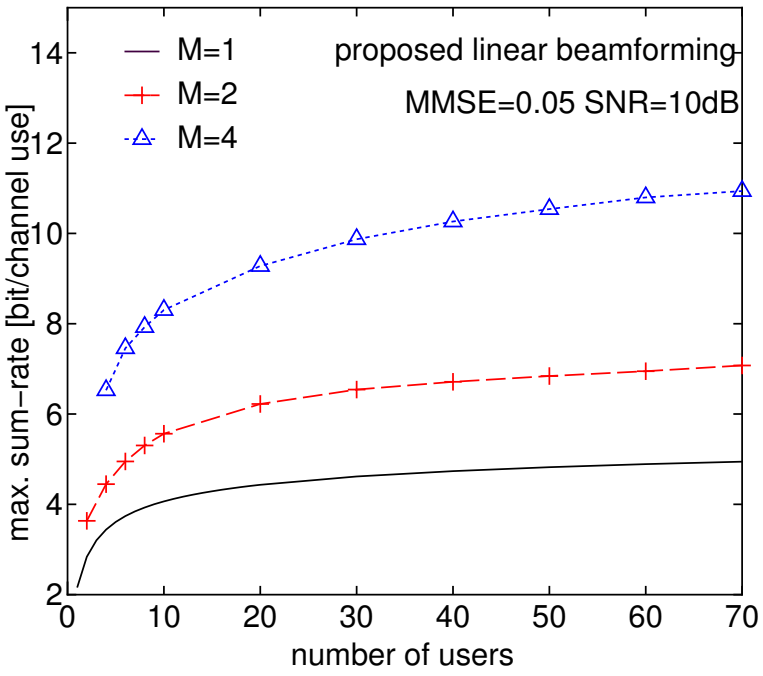


Figure 4.5: Sum rate vs. $K$ ($\sigma_e^2 = 0.05$)

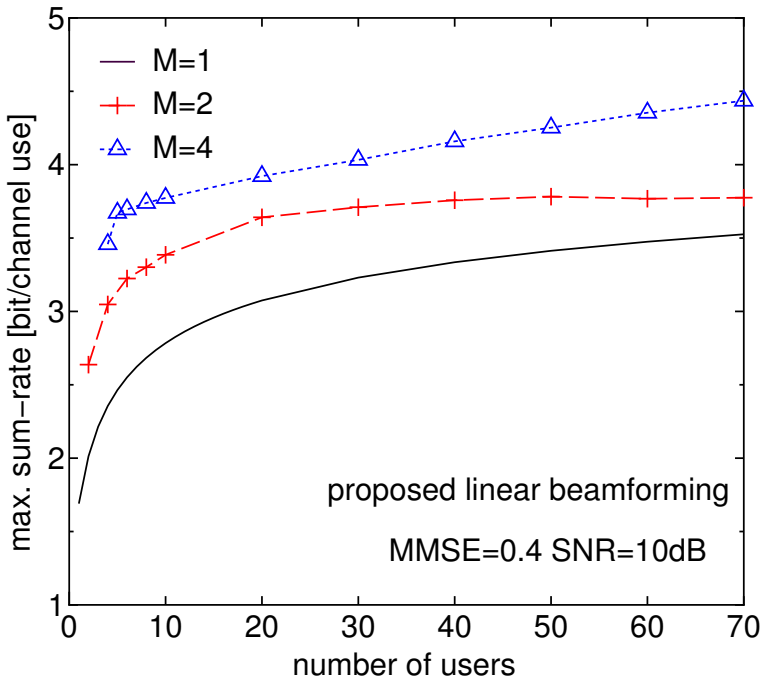


Figure 4.6: Sum rate vs. $K$ ($\sigma_e^2 = 0.4$)

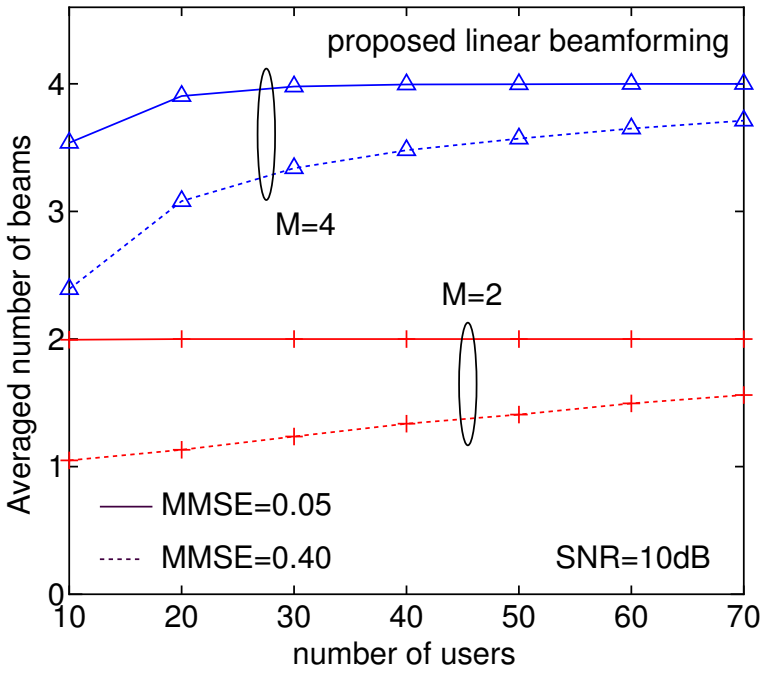


Figure 4.7: Averaged number of beams vs. $K$ ($\sigma_e^2 = 0.05,\ 0.4$)

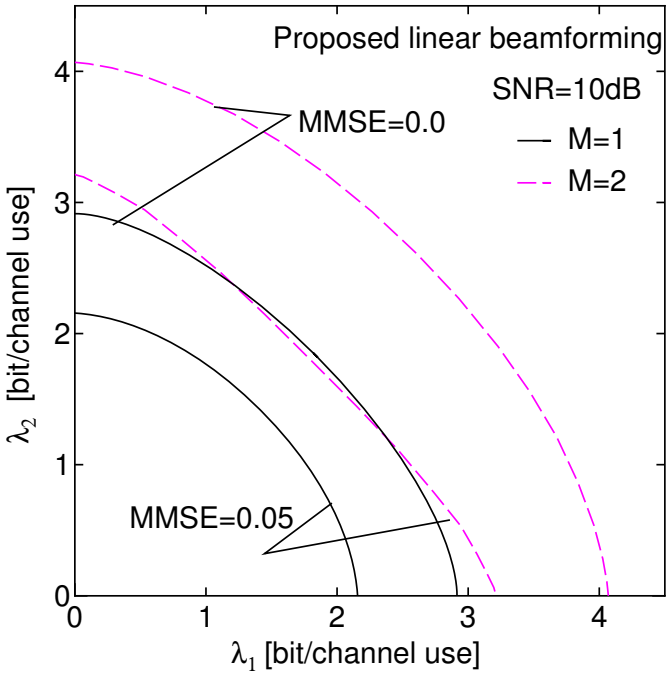


Figure 4.8: Two-user stability region

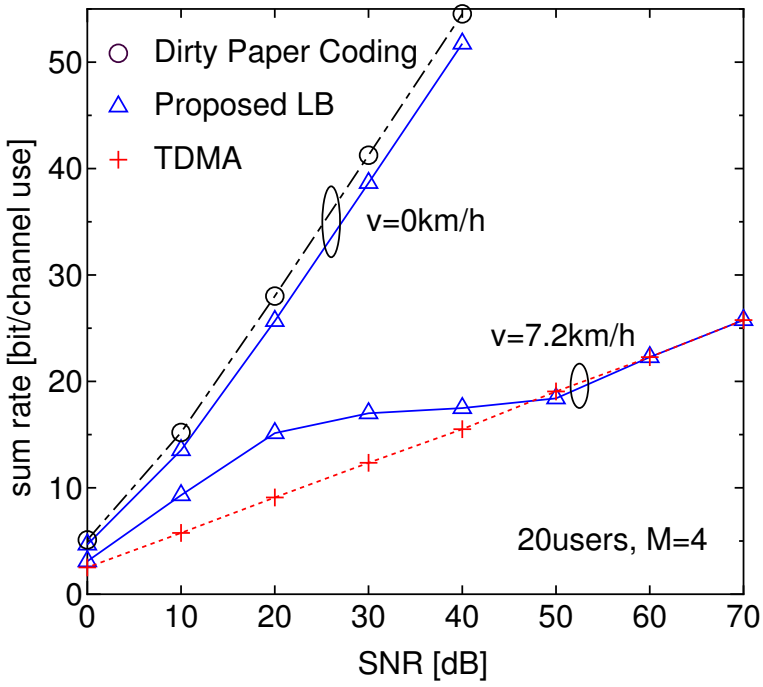


Figure 4.9: Sum rate vs. SNR

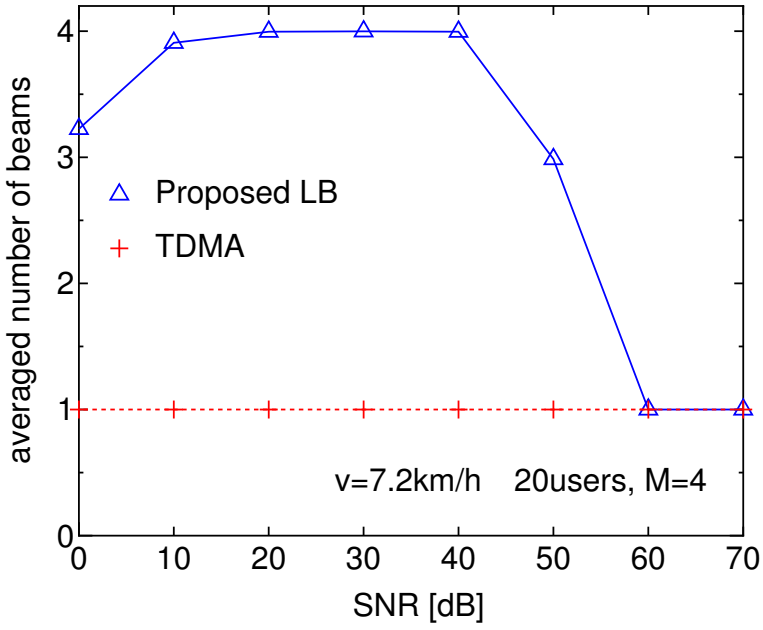


Figure 4.10: Number of beams vs. SNR

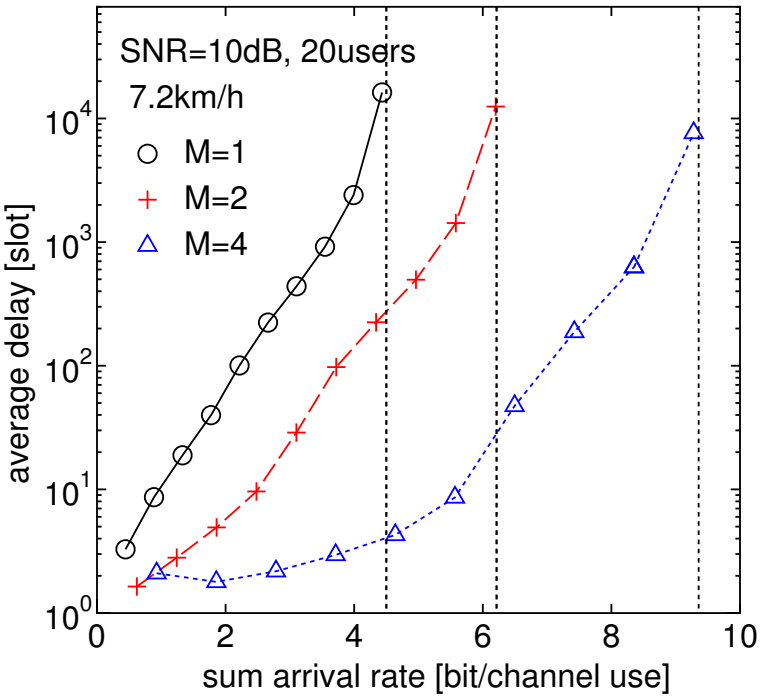


Figure 4.11: Average delay vs. arrival rate

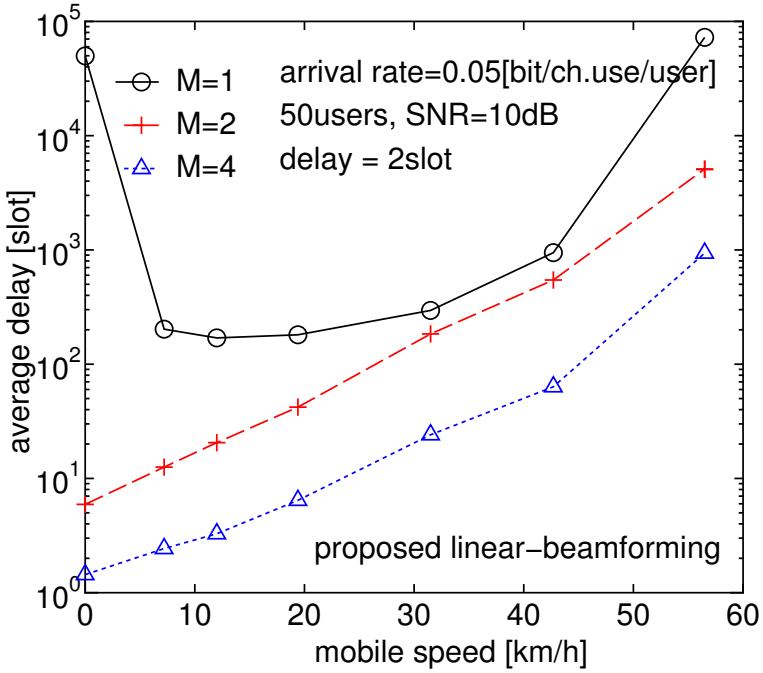


Figure 4.12: average delay vs. speed (proposed scheme)

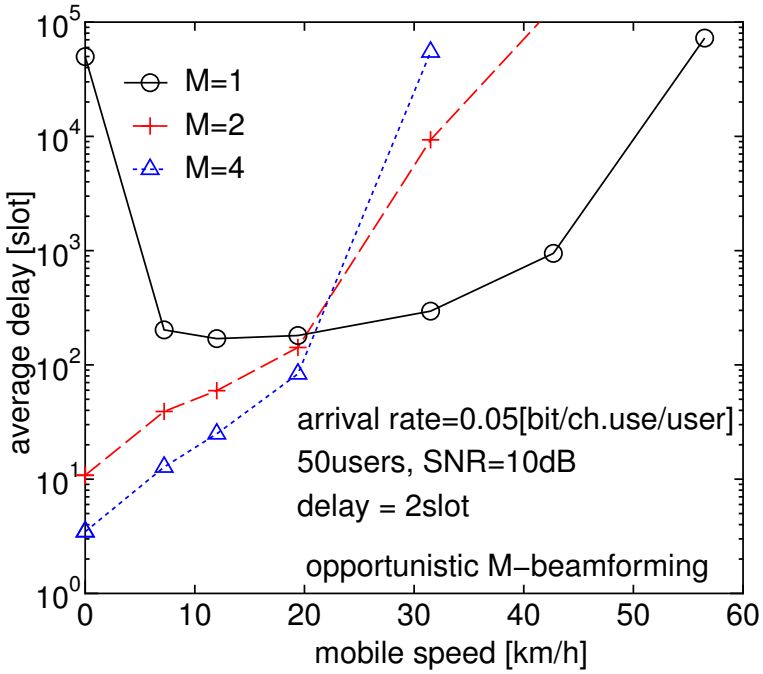


Figure 4.13: Average delay vs. speed (opportunistic $M$-beamforming)

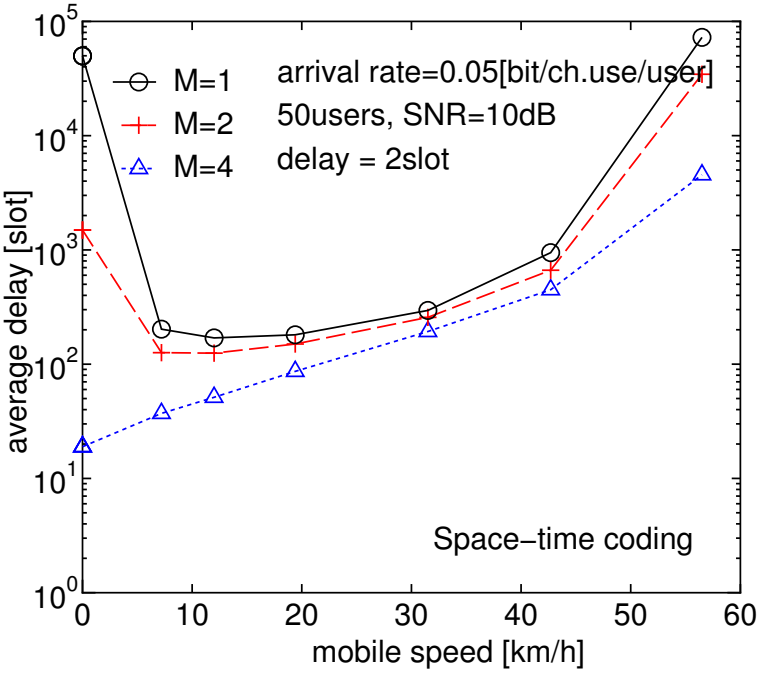


Figure 4.14: Average delay vs. speed (STC)

# APPENDIX

## 4.A Proof of Theorem 1

For a SDMA scheduling policy, the instantaneous service rate (information bits per slot) for user $k$ is given by $\mu_k(t) = TR_k(t)1\{R_k(t) < \overline{R}_k(t)\}$. The necessary condition (3.6) says that the system is stable if and only if

$$\lambda_k \leq \underline{\mu}_k = \liminf_{t\to\infty} \frac{1}{tT} \sum_{\tau=1}^{t} \mu_k(\tau), \quad k = 1, \ldots, K \tag{4.42}$$

Let $\widetilde{\Omega}$ denote the stability region of a new system with channel state $\boldsymbol{\alpha}(t)$ and feasible rate $\mathbf{R}^{\text{out}}(t)$. The instantaneous service rate (information bits per slot) of user $k$ in the new system is given by $\widetilde{\mu}_k(t) = TR_k^{\text{out}}(\mathbf{G}(t), \boldsymbol{\alpha}(t))$. Theorem 1 of [55] yields that the stability region $\widetilde{\Omega}$ of the new system is given by (4.14).

By restricting the resource allocation policies for the original system to the stationary policies $\{\mathbf{G}, \mathbf{R}\}$, $\mathbf{R}$ defined in (4.8), it is immediate to show, by ergodicity, that $\liminf_{t\to\infty} \frac{1}{tT} \sum_{\tau=1}^{t} \mu_k(\tau) = \mathbb{E}\left[R_k^{\text{out}}(\mathbf{G}, \boldsymbol{\alpha})\right]$. Hence, any point $\boldsymbol{\lambda} \in \widetilde{\Omega}$ is also in $\Omega$ and it is stabilized by a policy $\{\mathbf{G}, \mathbf{R}\}$. This shows that $\widetilde{\Omega} \subseteq \Omega$.

□

## 4.B Proof of Theorem 2

We demonstrate that max-stability adaptive policy stabilizes the system whenever the arrival rate vector is inside the stability region. For the sake of analysis simplicity, we assume that both the underlying channel process $\{\mathbf{h}(t)\}$ and the arrival process $\{\mathbf{A}(t)\}$ vary i.i.d. from one slot to another. Let's consider one-slot buffer dynamic under max-stability policy.

$$\begin{aligned} S_k^2(t+1) &= \{\max[0, S_k(t) - \mu_k(t)] + A_k(t)\}^2 \\ &\leq (S_k(t) - \mu_k(t))^2 + A_k(t)^2 + 2A_k(t)\max[0, S_k(t) - \mu_k(t)\}] \\ &\leq (S_k(t) - \mu_k(t))^2 + A_k(t)^2 + 2A_k(t)S_k(t) \\ &= S_k(t)^2 - 2TS_k(t)\left(\frac{\mu_k(t)}{T} - \frac{A_k(t)}{T}\right) + T^2\left\{\left(\frac{\mu_k(t)}{T}\right)^2 + \left(\frac{A_k(t)}{T}\right)^2\right\} \end{aligned}$$

where the instantaneous service rate under the max-stability policy is given by

$$\mu_k(t) = TR_k(t)\ 1\{R_k(t) < \overline{R}_k(t)\}$$

Taking expectations conditioned on $\mathbf{S}(t)$, scaling by weights $\theta_k$ and summing over all $k$, we have

$$\mathbb{E}\left[L(\mathbf{S}(t+1)) - L(\mathbf{S}(t))|\mathbf{S}(t)\right] \leq T^2\tilde{\mathcal{C}} - 2T\sum_{k=1}^{K}\theta_k S_k(t)\left(\mathbb{E}[R_k^{\text{out}}(t)|\mathbf{S}(t)] - \lambda_k\right) \quad (4.43)$$

where we used $\lambda_k = \mathbb{E}[\frac{A_k}{T}|\mathbf{S}_t]$ expressed in bit per channel use from i.i.d. arrival process and used $\mathbb{E}[\frac{\mu_k}{T}|\mathbf{S}(t)] = \mathbb{E}[R_k^{\text{out}}(t)|\mathbf{S}(t)]$ from the definition, and we have $\tilde{\mathcal{C}} = \sum_k \theta_k \mathbb{E}[(\mu_k(t)/T)^2|\mathbf{S}(t)] + \sum_k \theta_k \mathbb{E}[(A_k(t)/T)^2|\mathbf{S}(t)]$. $\tilde{\mathcal{C}}$ is upper bounded by a constant $\mathcal{C}$ given by

$$\mathcal{C} = \sum_{k=1}^{K}\theta_k \mathbb{E}\left[R_{\max}(t)^2\right] + \sum_{k=1}^{K}\theta_k \mathbb{E}\left[\left(\frac{A_k(t)}{T}\right)^2\right]$$

where we used the i.i.d. arrival assumption. In order to derive the first term of $\mathcal{C}$, we used the following

$$\begin{aligned} \sum_{k=1}^{K}\theta_k \mathbb{E}\left[\left(\frac{\mu_k(t)}{T}\right)^2\right] &= \sum_{k=1}^{K}\theta_k \mathbb{E}[\left(R_k^{\text{out}}(t)\right)^2|\mathbf{S}(t)] \\ &\leq \sum_{k=1}^{K}\theta_k \mathbb{E}[\hat{R}_{\max}(t)^2] \end{aligned}$$

where we define $\hat{R}_{\max}(t)$ as the upper bound of the rate achieved when the whole transmission power is allocated to the user with the largest eigen value

associated to the covariance matrix $\mathbf{\Sigma}_k(t)$. The rate allocated to that user is given by

$$\rho\hat{R}_{\max}(t) = \rho \max_k \log_2\left(1 + \frac{|\boldsymbol{\alpha}_k|^2 + \sigma_e^2}{N_0^{\text{down}}}\right) \tag{4.44}$$

where $\rho \in (0,1)$ is the rate margin. Letting $k^\star$ be the user achieving the maximum in (4.44), it follows immediately that for any $k$ we have

$$\begin{aligned} R_k^{\text{out}}(t) &\leq \rho\hat{R}_{\max}(t)(1 - \mathbb{P}(\overline{R}_{k^\star}(t) \leq \rho\hat{R}_{\max}(t))) \\ &\leq \rho\hat{R}_{\max}(t) \leq \hat{R}_{\max}(t) \end{aligned}$$

where the last two inequalities hold only for perfect CSIT. Notice that the term $\sum_{k=1}^K \theta_k S_k(t)\mathbb{E}[R_k^{\text{out}}(t)|\mathbf{S}(t)]$ in (4.43) maximizes the value of $\sum_{k=1}^K \theta_k S_k(t)\eta_k$ over all the vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K) \in \Omega$. Hence,

$$\sum_{k=1}^K \theta_k S_k(t)\mathbb{E}[R_k^{\text{out}}(t)|\mathbf{S}(t)] \geq \sum_{k=1}^K \theta_k S_k(t)\eta_k \tag{4.45}$$

For any $\boldsymbol{\lambda}$ inside the stability region $\Omega$, there exists $\boldsymbol{\delta} = (\delta, \ldots, \delta)$ such that $\boldsymbol{\lambda} + \boldsymbol{\delta} \in \Omega$. Then, the one step drift (4.43) is rewritten by

$$\mathbb{E}\left[L(\mathbf{S}(t+1)) - L(\mathbf{S}(t))|\mathbf{S}(t)\right] \leq T^2\mathcal{C} - 2T\delta\sum_{k=1}^K \theta_k S_k(t) \tag{4.46}$$

Choose any $\epsilon > 0$ and define the compact region $\Lambda$

$$\Lambda \triangleq \left\{\mathbf{S} \in \mathbb{R}^K : S_k > 0, \sum_{k=1}^K \theta_k S_k \leq \frac{T^2\mathcal{C} + \epsilon}{2T\delta}\right\}. \tag{4.47}$$

Then, for $\mathbf{S} \notin \Lambda$, the RHS of (4.46) is strictly less than $-\epsilon$. □

## 4.C Kalman filter to predict a GM process

The Discrete Kalman filter [96] can be used to obtain the CSIT given a delayed and noisy observation. Focusing without loss of generality on a scalar channel coefficient $h(t)$ where we drop the user and antenna index

because of the i.i.d. assumption, let define the state equation as well as the observation equation for the Kalman filter.

$$h(t) = rh(t-1) + \nu(t) \tag{4.48}$$
$$z(t) = h(t) + n(t) \tag{4.49}$$

where $\nu(t) \sim \mathcal{N}_{\mathbb{C}}(0, 1-r^2)$, $n(t) \sim \mathcal{N}_{\mathbb{C}}(0, N_0^{up})$ are both AWGN and uncorrelated each other. In the following, $\hat{h}(t|\tau)$ denotes the best linear estimate of $h(t)$ at time $t$ given the observation $z(t)$ for $t = 0, \ldots, \tau$ and $e(t|\tau) = h(t) - \hat{h}(t|\tau)$ denotes the estimation error of $h(t)$ given the linear estimate $\hat{h}(t|\tau)$ using the observation up to $t = 0, \ldots, \tau$. $P(t|\tau) = \mathbb{E}[|e(t|\tau)|^2]$ denotes the MMSE of $\hat{h}(t|\tau)$.

The Kalman filter consists of the following two steps. The first step is to find $\hat{h}(t|t-1)$ given $\hat{h}(t-1|t-1)$ without the observation $z(t)$. Then, the second step is to estimate $\hat{h}(t|t)$ given the observation $z(t)$ and $\hat{h}(t|t-1)$.

Suppose at $t = 0$, we have $h(0|0) = 0$ and $P_e(0|0) \triangleq \mathbb{E}[|e(0|0)|^2] = 1$. The first step can be done by using the system equation (4.48), i.e. we let

$$\hat{h}(t|t-1) = r\hat{h}(t-1|t-1)$$

which has an estimation error given by

$$\begin{aligned} e(t|t-1) &= h(t) - \hat{h}(t|t-1) \\ &= r(h(t-1) - \hat{h}(t-1|t-1)) + \nu(t) \\ &= re(t-1|t-1) + \nu(t) \end{aligned}$$

If $\hat{h}(t|t-1)$ is an unbiased estimate, since $\nu(t)$ has zero-mean we have $\mathbb{E}[e(t|t-1)] = 0$. Then, $\hat{h}(t|t-1)$ will be unbiased as well. The variance of the error $e(t|t-1)$ is given by

$$P(t|t-1) = r^2 P(t-1|t-1) + (1-r^2) \tag{4.50}$$

This completes the first step of the Kalman filter.

In the second step we incorporate the new observation $z(t)$ for estimating $h(t)$. By introducing the Kalman gain $K(t)$ and another variable $F(t)$, the new estimate is given as a linear combination of the estimate of the first step and the observation such that

$$\hat{h}(t|t) = F(t)\hat{h}(t|t-1) + K(t)z(t)$$

So that $\hat{h}(t|t)$ should be unbiased, $F(t) = 1 - K(t)$ should be satisfied. After a trivial development, we obtain

$$\hat{h}(t|t) = r\hat{h}(t|t-1) + K(t)[z(t) - r\hat{h}(t|t-1)] \tag{4.51}$$

and the estimation error is

$$e(t|t) = [1 - K(t)]e(t|t-1) - K(t)n(t) \tag{4.52}$$

The error variance is updated by

$$P(t|t) = [1 - K(t)]^2 P(t|t-1) + K(t)^2 N_0^{\text{up}}$$

The Kalman gain $K(t)$ that minimizes $P(t|t)$ is given by

$$K(t) = \frac{P(t|t-1)}{N_0^{\text{up}} + P(t|t-1)} \tag{4.53}$$

Plugging (4.53) into $P(t|t)$, we obtain the recursive expression for the error variance

$$P(t|t) = [1 - K(t)]P(t|t-1) \tag{4.54}$$

When $t \to \infty$, the estimation MMSE $P(t|t)$ approaches a steady state value $\xi_{mmse}$ given in (4.24).

Now, we are interested in predicting $h(t+d)$ at time $t$ given the delayed observation $z(t), \ldots, z(0)$ for the delay $d \geq 1$. Since the future observation $z(t+d), \ldots, z(t+1)$ are not available, we are in the condition of the first step. Using the system equation (4.48) with $d$ steps, we let the CSIT at time $t+d$ as follows.

$$\alpha(t+d) \stackrel{\Delta}{=} \hat{h}(t+d|t) = r^d \hat{h}(t|t) \tag{4.55}$$

which has the error given by

$$\begin{aligned} e(t+d|t) &= h(t+d) - \hat{h}(t+d|t) \\ &= \left( r^d h(t) + \sum_{l=0}^{d-1} r^l \nu(t+d-l) \right) - r^d \hat{h}(t|t) \\ &= r^d e(t|t) + \sum_{l=0}^{d-1} r^l \nu(t+d-l) \end{aligned}$$

Noticing that $e(t|t)$ and $\nu(t+d-l)$ for $l=0,\ldots,d-1$ is uncorrelated, the error variance is given by

$$P(t+d|t) = r^{2d}P(t|t) + (1-r^2)\sum_{l=0}^{d-1} r^{2l} \tag{4.56}$$

When $t \to \infty$, the prediction MMSE $P(t+d|t) = \mathbb{E}[|\alpha(t+d) - h(t+d|t)|^2]$ approaches a steady state value that we denote $\sigma_e^2$ given in (4.23).

# Chapter 5

# Conclusions

In this thesis, we investigated methods of using multiple antennas at the transmitter in the downlink of a cellular system. In Chapter 2, we focused on the use of multiple transmit antennas to improve the point-to-point link reliability with no CSIT. We proposed a concatenated TCM TR-STBC scheme for single-carrier transmission over frequency selective MISO channel. Thanks to a reduced-state joint equalization and decoding approach, our scheme achieved much lower complexity with similar/superior performance than previously proposed schemes. Moreover, since the receiver complexity is independent of the modulation constellation size and Ungerboeck TCM schemes incorporate variable constellation sizes, our scheme is suitable for implementing adaptive modulation with low complexity. This fact has particular relevance for the implementation of high-speed downlink schemes based on dynamic scheduling.

In Chapters 3 and 4, we studied the different uses of multiple transmit antennas in MIMO broadcast channels by exploiting the CSIT. In Chapter 3, we compared two practical schemes, STC and opportunistic beamforming, by restricting ourselves to a simple SNR feedback. In Chapter 4, we proposed a linear beamforming scheme by considering the more informative analog feedback. Table 5.1 summarizes these different schemes for the MIMO broadcast channel (BC) where (role) means the role of multiple transmit antennas. Opportunistic beamforming with a single beam is excluded from the

list because we found in Chapter 3 that its performance is dominated either by opportunistic $M$-beamforming for slow fading or by STC for fast fading. The amount of feedback per user in opportunistic $M$-beamforming is $M$ real values whereas it is 1 real value in STC. The proposed linear beamforming exploits the analog feedback where each user directly sends its estimated channel coefficients to the transmitter without quantizing and coding. The amount of the feedback per user is $M$ complex values, which is comparable to the other schemes. The number of simultaneously served users (beams) is fixed to 1 for STC and $M$ for opportunistic $M$-beamforming, although it can vary between 1 and $M$ in the proposed linear beamforming scheme. This flexibility of adjusting the number of beams makes our linear beamforming scheme much more robust than opportunistic $M$-beamforming under non-perfect CSIT. In the improved opportunistic beamforming where we decouple the problem of channel prediction from the variation of the beamformer, users should be synchronized so that the random beam matrices are known a priori. This additional complexity is not necessary in our linear beamforming scheme. Over a slow fading channel, our linear beamforming scheme is an attractive solution because it can exploit multiplexing gain, power gain, and multiuser diversity (MUD) gain. Power gain together with multiplexing gain makes the queues well balanced and thus enables our proposed scheme to achieve small average delay. Over a fast fading channel our linear beamforming scheme is still appealing, since it converges to a TDMA scheduling by allocating the whole power to the user with the largest weighted rate when the mobility increases (as reliable CSIT becomes unavailable).

Table 5.1: Different uses of multiple transmit antennas in MIMO-BC

| | Opportunistic $M$ beamforming | Space-time coding | Proposed Linear beamforming |
|---|---|---|---|
| Feedback | SINR | SNR | estimated channel |
| # required feedback/user | $M$ real | 1 real | $M$ complex |
| # users served/slot | $M$ | 1 | $1 \leq B \leq M$ |
| Slow fading performance (role) | multiplexing gain↓ + MUD↓ (induce faster fading) | MUD for small $K$ (channel-hardening) | multiplexing/power gain + MUD (accurate beamforming) |
| Fast fading performance (role) | interference-limited (no help) | MUD↓ (channel-hardening) | multiplexing/power gain↓ + MUD↓ (single beamforming) |

# Bibliography


[1] T.E.Kolding et al., "High Speed Downlink Packet Access : WCDMA Evolution," *IEEE Vehic. Tech. Society News*, pp. 4–10, February 2003.

[2] 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.

[3] "Cdma2000 high rate packet dataair interface specification," *TIA/EIA/3GPP2 Standard IS-856/3GPP2 C.S.0024, v3.0*, December 2001.

[4] "3gpp2 c.s0024, cdma2000 high rate packet data air interface specification," 2002.

[5] T.Bonald, "Flow-level performance analysis of some opportunistic scheduling algorithms," *Euro. Trans. on Telecomms.*, vol. 16, pp. 65–75, 2005.

[6] E.Biglieri and G.Taricco, "Transmission and reception with multiple antennas: Theoretical foundations," *Foundations and Trends in Communications and Information Theory*, 2004.

[7] A.Paulraj, R.Nabar, and D.Gore, "Introduction to space-time wireless communications," *Cambridge University Press*, 2003.

[8] D.Tse and P. Viswanath, "Fundamentals of Wireless Communications," *Cambridge University Press.*

[9] E.G.Larsson and P.Stoica, "Space-time block coding for wireless communications," *Cambridge University Press*, 2003.

[10] S.M.Alamouti, "A Simple Transmit Diversity Techniques for Wireless Communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, October 1998.

[11] V.Tarokh, N.Seshadri, and A.R.Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 744 – 765, March 1998.

[12] L.H.Ozarow, S.Shamai, and A.D.Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehic. Tech.*, vol. 43, pp. 359 – 378, May 1994.

[13] J.G.Holma and A.Toskala, "WCDMA for UMTS," *Wiley*, 2000.

[14] "Digital cellular communications system (Phase 2+)," *ETSI/3GPP, Sophia-Antipolis, France, Technical Specifications 3GPP TS 05.01-05*, 2001.

[15] E.Lindskog and A.Paulraj, "A transmit diversity scheme for channels with intersymbol interference," *Communications, 2000. ICC 2000. 2000 IEEE International Conference on*, vol. 1, pp. 307 –311, 2000.

[16] S.Zhou and G.B.Giannakis, "Single-Carrier Space-Time Block Coded Transmissions over Frequency-Selective Fading Channels," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 164–179, January 2003.

[17] H.El Gamal, A.R.Hammons, Y.Liu, M.P.Fitz, and O.Y.Takeshita, "On the Design of Space-Time and Space-Frequency codes for MIMO Frequency Selective Fading Channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 9, September 2003.

[18] W.Choi and J.M.Cioffi, "Multiple input/multiple output(MIMO) equalization for space-time block coding," *in Proc. IEEE Pacific Rim Conf. Commun., Computers and Sig. Proc.*, pp. 341–344, 1999.

[19] A.O.Berthet, R.Visoz, and J.J.Boutros, "Space-time BICM versus space-time trellis code for MIMO block fading multipath AWGN channel," *Proc. in ITW2003*, March 2003.

[20] Y.Liu, M.P.Fitz, and O.Y.Takeshita, "Space-Time Codes Performance Criteria and Design for Frequency Selective Fading Channels," *ICC'2001 Proceeding*, vol. 9, pp. 2800–2804, 2001.

[21] L.Li and A.J.Goldsmith, "Capacity and Optimal resource Allocation for Fading Broadcast Channels-Part I:Ergodic Capacity," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 1083–1102, March 2001.

[22] D.N.C.Tse, "Optimal Power Allocation over parallel Gaussian Broadcast Channels," *unpublished.*

[23] P.Viswanath, D.N.C.Tse, and R.Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, June 2002.

[24] R.Knopp and P.A.Humblet, "Information capacity and power control in single-cell multiuser communications," *Proc. IEEE ICC*, vol. 1, pp. 331–335, 1995.

[25] R.Knopp, "Coding and Multiple-Access over Fading Channels," *Ph.D dissertation, Ecole Polytechnique Fédérale de Lausanne*, 1997.

[26] M.Sharif and B.Hassibi, "On the Capacity of MIMO Broadcast Channel with Partial Side Information," *IEEE Trans. on Inform. Theory*, vol. 51, no. 2, pp. 506 – 522, February 2005.

[27] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *Proceeding of ISIT 2004, Chicago IL*, July 2004.

[28] P.Viswanath and D.N.C. Tse, "Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 1912 – 1921, August 2003.

[29] N. Jindal, S. Vishwanath, and A.J. Goldsmith, "Duality, achieving rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Trans. on Inform. Theory*, , no. 10, October 2003.

[30] W.Yu and J.Cioffi, "Sum capacity of vector gaussian broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 50, no. 9, September 2004.

[31] M.Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[32] G.Caire, S.Shamai, Y.Steinberg, and H.Weingarten, "Space-time wireless systems: From array processing to mimo communications, chapter 1 : Mimo broadcast channel," .

[33] M.Schubert and H.Boche, "Solution of Multiuser Downlink Beamforming Problem With Individual SINR Constraints," *IEEE Trans. on Vehic. Tech.*, vol. 53, January 2004.

[34] M.Stojnic, H.Vikalo, and B.H.Hassibi, "Rate maximization in multi-antenna broadcast channels with linear preprocessing," *Proceeding of IEEE Globecom '2004*, 2004.

[35] M.A.Meddah-Ali, M.ANsari, and A.K.Khandani, "An Efficient Algorithm for User Selection over MIMO Multiuser Systems," *Technical Report, University of Waterloo*, August 2003.

[36] A.Bayesteh and A.K.Khandani, "An Efficient Method for User Selection in MIMO Broadcast Channels," *Proceeding of CISS'2005*, March 2005.

[37] T.Yoo and A.Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," *To apper in Proceeding of ICC'2005*, May 2005.

[38] B.M.Hochwald, C.B.Peel, and A.L.Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication-Part ii:Perturbation," *submitted to* IEEE Trans. on Communications, June 2003.

[39] C.Windpassinger, R.F.H.Ficher, and J.B.Huber, "Lattice-Reduction-Aided Broadcast Precoding," *ITG Conference*, vol. Munich, Germany, January 2004.

[40] G.Caire and S.Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 1691 – 1706, July 2003.

[41] R.Gozali, R.M.Buehrer, and B.D.Woerner, "The Impact of Multiuser Diversity on Space-Time Block Coding," *IEEE Communications Letters*, vol. 7, no. 5, May 2003.

[42] B.M.Hochwald, T.L.Marzetta, and V.Tarokh, "Multiple-Antenna Channel Hardening and its Implications for Rate Feedback and Scheduling," *IEEE Trans. on Inform. Theory*, vol. 50, no. 9, September 2004.

[43] H.Viswanathan and S.Venkatesan, "The Impact of Antenna Diversity in Packet Data Systems with Scheduling," *IEEE Trans. on Commun.*, pp. 546–549, April 2004.

[44] D.Samardzija and N.Mandayam, "Unquantized and Uncoded Channel State Information Feedback on Wireless Channels," *Proceeding of IEEE WCNC'2005*, pp. New Orleans, LA, USA, March 2005.

[45] T.L.Marzetta and B.M.Hochwald, "Fast Transfer of Channel State Information in Wireless Systems," *Submitted to "IEEE Transactions on Signal Processing"*, June 2004.

[46] N.Jindal, "MIMO Broadcast Channels with Finite Rate Feedback," *Submitted to ISIT'2005.*

[47] M.Kobayashi and G.Caire, "A low-complexity approach to space-time coding for multipath fading channels," *Proceeding of WPMC'2003*, vol. Yokosuka, Japan, October 2003.

[48] M.Kobayashi and G.Caire, "A low-complexity approach to space-time coding for multipath fading channels," *to appear in EURASIP Jounal on Wireless Communications and Networking*, 2005.

[49] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets. I:Introduction," *IEEE Commun.Mag.*, vol. 25, no. 2, pp. 5–11, 1987.

[50] M.Kobayashi, G.Caire, and D.Gesbert, "Transmit diversity vs. opportunistic beamforming in data packet mobile downlink transmission," *submitted to IEEE Trans. on Commun.*, April 2005.

[51] M.Kobayashi, G.Caire, and D.Gesbert, "Antenna Diversity vs. Multiuser Diversity: Quantifying the Tradeoffs," *Proceeding of ISITA'2005*, pp. Parma, Italy, October 2004.

[52] M.Kobayashi, G.Caire, and D.Gesbert, "Impact of transmit antennas in a queued tdma/sdma downlink," *to appear in Proceeding of SPAWC'2005, New-York,USA*, June 2005.

[53] M.Kobayashi, G.Caire, and D.Gesbert, "Opportunistic beamforminv vs. space time coding in a queued downlink," *Proceeding of IST Summit 2005*, vol. Dresden, Germany, June 2005.

[54] M.Kobayashi and G.Caire, "A linear beamforming scheduling for the mimo broadcast channel with analog feedback," *in preparation.*

[55] M.J.Neely, E.Modiano, and C.E.Rohrs, "Power Allocation and Routing in Multibeam Satellites With Time-Varying Channels," *IEEE/ACM Transaction on Networking*, vol. 11, pp. 138–152, February 2003.

[56] E.M.Yeh and A.S.Cohen, "Information Theory, Queueing, and Resource Allocation in Multi-user Fading Communications," *Proceedings of the 2004 Conference on Information Sciences and Systems,Princeton, NJ*, March 2004.

[57] S.Borst and P.Whiting, "The Use of Diversity Antennas in High-Speed Wireless Systems: Capacity Gains, Fairness Issues," *Bell Laboratories Technical Memorandum, 2001*, 2001.

[58] S.Shakkottai and A.Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," *Proc. of the 17th International Teletraffic Congress (ITC-17)*, September 2001.

[59] A.Sang, X.Wang, M.Madihian, and R.D.Gitlin, "Downlink scheduling schemes in cellular packet data systems of multiple-input multiple-output antennas," *Proceeding of IEEE Globecom 2004*, pp. 421–427, December 2004.

[60] John G.Proakis, "Digital communications," *McGraw-Hill Book Company*, 1997.

[61] E. Biglieri, J. Proakis, , and S. Shamai, "Fading channels : information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 2619–2692, October 1998.

[62] L.Zheng and D.N.C.Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[63] A.Medles, "Coding and Advanced Signal Processing for MIMO Systems," *Ph.D thesis, Ecole Nationale Supérieure des Télécommunications*, April 2004.

[64] R.Schober, H.Z.B.Chen, and W.H.Gerstacker, "Decision Feedback Sequence Estimation for Time-Reversal Space-Time Block Coded Transmission," *IEEE Trans. on Vehic. Tech.*, vol. 53, pp. 1273–1278, July 2004.

[65] N.Al-Dhahir, "Decision Feedback Sequence Estimation for Time-Reversal Space-Time Block Coded Transmission," *IEEE Trans. on Vehic. Tech.*, vol. 53, pp. 1273–1278, July 2004.

[66] L.R.Bahl, J.Cocke, F.Jelinek, and J.Raviv, "Optimal decoding," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, October 1998.

[67] G.D.Forney and G.Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 2384–2415, 1998.

[68] G.Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. on Inform. Theory*, vol. 28, pp. 55–67, 1982.

[69] E.Biglieri, D.Divsalar, P.J.McLane, and M.K.Simon, "Introduction to Trellis-Coded Modulation with Applications," *Macmillan*, 1997.

[70] V.Tarokh, H.Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[71] O.Tirkkonen, "Square-matrix embeddable space-time block codes for complex signal constellations," *IEEE Trans. on Inform. Theory*, vol. 48, no. 2, pp. 384–395, February 2002.

[72] Xue-Bin Liang, "Orthogonal designs with maximal rates," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2468 – 2503, October 2003.

[73] G.K.Kaleh, "Channel Equalization for Block Transmission Systems," *IEEE J. Select. Areas Commun.*, vol. 13, no. 1, pp. 110–121, January 1995.

[74] R.Raheli, A.Polydoros, and C.-K.Tzou, "Per-Survivor Processing: A general approach to MLSE in uncertain environments," *IEEE Trans. on Commun.*, vol. 43, no. 2, pp. 354–364, February 1995.

[75] G.Caire and G.Colavolpe, "On space-time coding for quasi-static multiple-antenna channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 6, pp. 1400–1416, June 2003.

[76] C.Schlegel, "Trellis coding," *IEEE Press*, 1997.

[77] R.D.Wesel, "Reduced Complexity Trellis Code Transfer Function Computation," *Proc. of Commun. Theory Mini-Conference*, pp. 37–41, June 1999.

[78] E.Malkamaki and H.Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1643–1646, July 1999.

[79] J.M.Cioffi, G.P.Dudevoir, M.V.Eyuboglu, and G.D.Forney, "MMSE Decision-Feedback Equalizers and Coding-Part I: Equalization Results," *IEEE Trans. on Commun.*, vol. 43, no. 10, pp. 2582–2594, October 1995.

[80] T. Cover and J. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[81] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 1250–1276, June 2002.

[82] ITU, "Rec.ITU-RM 1225, Guidelines for evaluation of radio transmission technologies for IMT-2000," 1997.

[83] "TD-SCDMA," *http://www.tdscdma-forum.org*.

[84] L.Zheng and D.N.C.Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 1073 – 1096, May 2003.

[85] M.J.Neely, "Dynamic Power Allocation and Routing for Satellites and Wireless Networks with Time-Varying Channels," *Ph.D dissertation, MIT*, November 2003.

[86] H.Boche and M.Wiczanowski, "Stability Region of Arrival Rates and Optimal Scheduling for MIMO-MAC - A Cross-Layer Approach," *Proceeding of IZS,Zurich*, February 2004.

[87] N.Randolph, "Probability, stochastic process and queueing theory :the mathematics of computer performance modeling," *Springer-Verlag*, 1995.

[88] S.Sanayei and A.Nosratinia, "Opportunistic Beamforming with Limited Feedback," *submitted to IEEE International Symposium on Information Theory*, 2005.

[89] E.Biglieri, G.Caire, G.Taricco, and J.Ventura-Traveset, "Simple method for evaluating error probabilities," *Electron.Lett.*, vol. 32, pp. 191–192, February 1996.

[90] W.Turin and R.Nobelen, "Hidden Markov modeling of flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1809–1817, December 1998.

[91] K.E.Baddour and N.C.Beaulieu, "Autoregressive models for fading channel simulation," *Proceeding of IEEE Globecom 2001*, pp. 1187–1192, November 2001.

[92] William C Jakes, "Microwave Mobile Communications," *IEEE Press*, 1974.

[93] M.Schwartz, W.Bennet, and S.Stein, "Communication systems and techniques," *McGraw-Hill, New York*, 1996.

[94] M.Gastpar, B.Rimoldi, and M.Vetterli, "To code, or not to code : Lossy source-channel communication revisited," *IEEE Trans. on Inform. Theory*, vol. 49, May 2003.

[95] H.Boche and M.Wiczanowski, "Stability Optimal Transmission Policy for the Multiple Antenna Multiple Access Channel in the Geometric View," *Unpublished*, 2005.

[96] M.H.Hayes, "Statistical digital signal processing and modeling," *John Wiley & Sons, Inc.*, 1996.