POLITECNICO DI MILANO
Dipartimento di Elettronica e Informazione
DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE

LINEAR SIGNAL PROCESSING FOR SINGLE/MULTI-USER MIMO COMMUNICATION SYSTEMS OVER FREQUENCY-SELECTIVE CHANNELS

Doctoral Dissertation of:
Osvaldo Simeone

Advisor:
Prof. Umberto Spagnolini

Tutor:
Prof. Giuseppe Drufuca

Supervisor of the Doctoral Program:
Prof. Stefano Crespi Reghizzi

2005 - XVII

Ringraziamenti

Questa tesi tratta di sistemi di comunicazione in grado di trasmettere simultaneamente molteplici informazioni ad un medesimo destinatario. Come è lecito attendersi, la chiave del successo della comunicazione dipende dalla capacità dell'ascoltatore di distinguere i diversi stimoli benché questi siano ricevuti contemporaneamente. A tal fine, il destinatario ha bisogno essenzialmente di due cose: di ascoltare sufficientemente a lungo e da diverse posizioni. In altre parole, ha bisogno di tempo e spazio.

L'oggetto della tesi fornisce una facile metafora per descrivere l'esperienza del dottorato. Durante i tre anni di corso, gli stimoli sono molteplici e attinenti a campi diversi. Come disse un giorno il prof. D., il dottorato serve per imparare "a leggere e scrivere, ad ascoltare e comunicare, a corteggiare le ragazze, a organizzare il proprio tempo" (non necessariamente in questo ordine). Le idee, per entrare a far parte del proprio bagaglio, richiedono tempo e spazio. E le idee hanno sempre il viso, almeno quando nascono, delle persone che le trasmettono. Nel mio caso, i tempi, i luoghi e i visi di questa comunicazione (tuttora in corso) sono quelli che ricostruisco brevemente e in modo inevitabilmente incompleto nel seguito in forma di ringraziamenti.

Il primo e il più grande va al prof. Umberto Spagnolini: da lui ho imparato l'atteggiamento critico nei confronti dei problemi e l'importanza della comunicazione dei risultati prima, durante e dopo l'attività di ricerca. La sua passione e la sua competenza hanno fatto di questi tre anni un'esperienza gratificante, sia a livello personale che professionale.

Grazie a Monica (Nicoli) per la sua professionalità, disponibilità e simpatia.

Un ringraziamento speciale al prof. Giuseppe Drufuca per il supporto durante l'attività di ricerca relativa al tema minore e per le lunghe conversazioni.

I would like to gratefully acknowledge the supervision and hospitality of Dr. Yeheskel Bar-Ness during my stay at NJIT (Newark, USA). Thanks also to Marlene Toeroek and the entire CCSPR gang. A special acknowledgement goes to prof. Ali Abdi for the interesting discussions.

Grazie alla dott. Ada Bosisio (anche per i consigli musicali e cinematografici), al prof. Vittorio Rampa, al prof. Fabio Rocca, al dott. Maurizio Magarini e a tutto il Dipartimento.

Grazie a Beppe Primolevo per il suo inestimabile supporto, a Roberto Bosisio, a Giovanni Del Galdo e ai tesisti che mi hanno aiutato a portare avanti diverse attività di ricerca: Marcello Cicerone, Ivan Rizzo (il dott. Rizzo), Stefano Savazzi e Nicola Varanese.

L'ultimo ringraziamento è per i miei instancabili amici di sempre, per Bi e per la mia famiglia ("...se riesci incontrando il successo e la sconfitta a trattare questi due impostori allo stesso modo...").

Dissertation for the degree of Doctor of Philosophy
in Information Engineering at Politecnico di Milano, 2005.

ABSTRACT
O. Simeone, Linear signal processing for single/multi-user MIMO communication systems over frequency-selective channels

In recent years, with an ever increasing number of wireless services to be allocated within a limited frequency resource, the focus of both the Information Theory and Signal Processing communities has been the development of spectrally efficient transmission technologies at the physical layer. The most promising of such technologies is that of the MIMO systems, that consist in a radio link with multiple antennas at both ends. A crucial factor in determining the performance of a MIMO system.is the availability of some information about the propagation channel (Channel State Information, in short CSI) at the receiver and possibly the transmitter. The CSI is acquired by the receiver by estimating the channel through the observation of training symbols sent within each block. On the other hand, the transmitter can obtain some CSI by a feedback channel from the receiver and use this information in order to optimize the signal precoding. This thesis attempts to fill a gap in the present body of work on this subject by considering both the CSI acquisition phase and the design of linear precoding and equalization. The main contribution is the study of these aspects within a realistic frequency-selective channel model, suitable for different propagation environment and antenna geometries, that accounts for a fundamental property of multipath channel. In particular, the fact the time variability of the multipath channel is due to two classes of parameters that have different rates of variation is exploited. In this way, one is able to distinguish between long term channel features and the fast varying fading process. The treatment encompasses both time-domain and multicarrier transmission. In the first part of the thesis, the impact of this assumption is investigated for the CSI acquisition phase, at first from a theoretical standpoint, by evaluating a lower bound on the channel estimation error through the computation of the hybrid Cramér-Rao bound (HCRB). Then, linear channel estimators, based on the separate computation of long term features of the channel (through subspace tracking) and tracking of the fading process, that are able to attain the bound, are proposed. In the second part of the thesis, the study of linear precoding and equalization is performed from both an information theoretic and signal processing persepective in order to evaluate the system performance within the considered realistic propagation model. In particular, the problem of designing linear precoding algorithms based on long term CSI at the transmitter is addressed for a single user MIMO link, whereas the problem of joint precoding, equalization and scheduling is tackled for the downlink of a multiuser MIMO system with instantaneous CSI at the transmitter.

Keywords: MIMO systems, channel estimation, imperfect CSI, long-term CSI, information rate, precoding, channel aware scheduling.

Osvaldo Simeone, Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza L. da Vinci 32, I-20133 Milano, Italy. Email: simeone@elet.polimi.it.

RIASSUNTO
O. Simeone, Linear signal processing for single/multi-user MIMO communication systems
over frequency-selective channels

Recentemente, dato il sempre crescente numero di servizi radio da allocare in un spettro limitato di frequenze, l'interesse dei ricercatori sia nel campo della teoria dell'informazione che della teoria dei segnali, si è concentrato sullo sviluppo di tecniche di trasmissione al livello fisico che garantiscano alta efficienza spettrale. La più promettente tra queste tecnologie è quella dei sistemi MIMO, che prevede l'uso di antenne multiple ad entrambi gli estremi di un collegamento radio.

Un fattore cruciale nel determinare le prestazioni di un sistema MIMO è la disponibilità di qualche informazione sullo stato del canale (Channel State Information, CSI) al ricevitore ed eventualmente al trasmettitore. La CSI è ottenuta al ricevitore effettuando una stima del canale a partire dalla conoscenza dei simboli di training contenuti in ogni blocco trasmesso. Il trasmettitore, invece, può ottenere la CSI attraverso un canale di feedback dal ricevitore ed utilizzare questa informazione per ottimizzare la precodifica del segnale.

In questa tesi, si considera sia la fase di acquisizione della CSI che il progetto della precodifica e dell'equalizzazione lineare. Il contributo più rilevante è lo sudio di questi aspetti all'interno di un modello realistico di canale selettivo in frequenza, adatto a diversi scenari propagativi e geometrie delle schiere di antenne, che tiene conto di una proprietà fondamentale del canale radio. La trattazione comprende sia la trasmissione nel dominio del tempo che della frequenza (trasmissione multiportante). In particolare, si sfrutta il fatto che la variabilità temporale del canale multipercorso è dovuta a due classi di parametri che variano con diverse dinamiche. In questo modo, si è in grado di distinguere tra caratteristiche del canale di lungo periodo e il processo di fading velocemente variabile.

Nella prima parte della tesi, l'impatto della proprietà di cui sopra sul canale di propagazione è studiato per la fase di acquisizione della CSI, dapprima da un punto di vista teorico, valutando un limite inferiore sull'errore di stima di canale attraverso il calcolo del CRB ibrido. In seguito vengono proposti stimatori di canale lineari, basati sul calcolo separato delle caratteristiche di lungo periodo del canale (attraverso tecniche di inseguimentoi di sottospazi) e sul tracking del processo di fading, che sono in grado di raggiungere il limite inferiore. Nella seconda parte della tesi, si effettua lo studio della precodifica lineare ed equalizzazione sia dal punto di vista della teoria della informazione che dell'elaborazione dei segnali, in modo da valutare le prestazioni del sistema all'interno del modello di canale considerato. In particolare, si affronta il problema di progettare algoritmi di precodifica lineare basati sulla CSI di lungo periodo disponibile al trasmettitore per un collegamento MIMO singolo utente, mentre si considera il problema del progetto congiunto di precodifica, equalizzazione e scheduling per il downlink di un sistema MIMO multiutente con CSI istantanea al trasmettitore.

Contents

Notational conventions

Notations

$\mathbf{X}^T$	Transpose of the matrix $\mathbf{X}$
$\mathbf{X}^H$	Transpose conjugate of the matrix $\mathbf{X}$
$\mathbf{X}^*$	Conjugate of the matrix $\mathbf{X}$
$tr(\mathbf{X})$	Trace of the matrix $\mathbf{X}$
$\|\mathbf{X}\|$	Determinant of the matrix $\mathbf{X}$
$\mathbf{X} \geq 0$	$\mathbf{X}$ is semipositive definite ($>$ for positive definite)
$vec(\mathbf{X})$	Column vector formed by stacking the columns of matrix $\mathbf{X}$ on top of each other
$\mathbf{X}^{1/2}$	Cholesky factor of a positive definite matrix $\mathbf{X}$: $[\mathbf{X}^{1/2}]^H \mathbf{X}^{1/2} = \mathbf{X}$
$\mathbf{X}^{H/2}$	$[\mathbf{X}^{1/2}]^H$
$\mathbf{X}^{-H}$	$[\mathbf{X}^{-1}]^H$
$\mathbf{X}^\dagger$	Pseudo-inverse of $\mathbf{X}$: for a tall matrix it reads $\mathbf{X}^\dagger = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H$
$[\mathbf{X}]_{i,j}$	(i,j)th entry of matrix $\mathbf{X}$
$rank(\mathbf{X})$	Rank of matrix $\mathbf{X}$
$range(\mathbf{X})$	Subspace spanned by the columns of matrix $\mathbf{X}$ (range of $\mathbf{X}$)
$null(\mathbf{X})$	Orthogonal complement of $range(\mathbf{X}^H)$
$\mathbf{\Pi}_{\mathbf{X}}$	Orthogonal projector onto $range(\mathbf{X})$: $\mathbf{\Pi}_{\mathbf{X}} = \mathbf{X}\mathbf{X}^\dagger$
$\mathbf{\Pi}_{\mathbf{X}}^\perp$	Orthogonal projector onto $null(\mathbf{X}^H)$
$\mathbf{I}_N$	$N \times N$ identity matrix
$\|\|\mathbf{X}\|\|$	Frobenius norm of matrix $\mathbf{X}$
$\mathbf{0}_N$	$N \times N$ matrix with entries equal to 0
$diag(\mathbf{x})$	Diagonal matrix having the elements of the vector $\mathbf{x}$ as diagonal entries
$\mathrm{Re}(\cdot)$	Real part of
$\mathrm{Im}(\cdot)$	Imaginary part of
$\mathcal{F}\{\cdot\}$	Fourier transform

○   Hadamard product
⋆   Convolution
⊗   Kronecker product
⋄   Khatri-Rao product
≃   Approximated by
∝   Proportional to
$E[\cdot]$ Expected value of
$\mathcal{N}(\mu, \sigma^2)$ Normal distribution with mean μ and variance σ^2
$\mathcal{CN}(\mu, \sigma^2)$ Complex normal distribution with mean μ and variance σ^2

Abbreviations

AOA   Angle Of Arrival
AOD   Angle Of Departure
AGN   Additive Gaussian Noise
CDMA   Code Division Multiple Access
CRB   Cramér-Rao Bound
DFT   Discrete Fourier Transform
FDD   Frequency Division Duplex
FFT   Fast Fourier Transform
HCRB   Hybrid Cramér-Rao Bound
IDFT   Inverse Discrete Fourier Transform
IFFT   Inverse Fast Fourier Transform
LMS   Least Mean Squares
MA   Modal Analysis
MDL   Minimum Description Length
MIMO   Multiple Input Multiple Output
MISO   Multiple Input Single Output
ML   Maximum likelihood
MSE   Mean Square Error
MMSE   Minimum Mean Square Error
SIMO   Single Input Multiple Output
SISO   Single Input Single Output
SNR   Signal to Noise Ratio
TDD   Time Division Duplex
TDMA   Time Division Multiple Access
UML   Uncostrained Maximum Likelihood
ZF   Zero Forcing

Chapter 1

Introduction

WIRELESS systems are already ubiquitous in providing connection between people or devices, irrespective on their location. In recent years, the worldwide spread of digital cellular systems has modified the way people communicate, making it possible to access multimedia contents and to exchange information almost anywhere. The "anywhere, anytime" paradigm finds application not only for business and personal communication, but is also a promising means to build a telecommunication network in developing countries where no legacy infrastructure is present.

For the near future, it is expected that different wireless networks and systems will co-exist, furnishing the end user with the possibility to transparently connect to the service that best serves his needs. Various access technologies range from satellite systems that provide low bit rate but global coverage to hot spot access points with high bit rate and coverage limited to a few hundred meters. In such a scenario, with an ever increasing number of wireless services to be allocated within a limited frequency resource, it is of vital importance to develop spectrally efficient transmission technologies at the physical layer. The main impediment to this task is the nature of the propagation within the wireless medium, that is characterized by random power and phase fluctuations (referred to as fading).

Since the advent of digital wireless communication, there has been much debate within the standardization bodies on what is the best radio interface to be implemented for different propagation environments and services. For instance, second generation cellular systems in Europe have preferred the TDMA technology in order to grant access to the spectral resource to multiple users, whereas in the United States the preferred choice has been CDMA. More recently, most of the discussion has focused on whether time-domain or multicarrier transmission has the most desirable properties within an adverse propagation environment such as indoor or urban outdoor. Amidst these debates, in the mid 90's, the seminal works of Telatar [1] and Foschini [2] shed light on an novel technology that promised an increase of the bandwidth efficiency to a level that could not be achieved by any other known technique. According to this technology, multiple antennas are deployed at both the transmitter and receiver side, forming a

so called MIMO (that stands for Multiple Input Multiple Output) system.

From these early reports, MIMO systems have become a major research topic for both the Information Theory and Signal Processing communities. Ten years later, many questions have been answered but, as discussed below, there are still many important open issues. Addressing these aspects in a successful way is of utmost importance in making the promise of MIMO technology become reality.

1.1   State of the art and motivation

Before Telatar and Foschini, the best candidate technology in order to improve the performance of wireless systems was that of the so called *smart antennas*. This technique prescribes the use of an antenna array in order to communicate (either transmitting in downlink or receiving in uplink) with multiple single antenna terminals. The scenario is clearly of great importance in cellular systems where the base station can be quite easily provided with a more costly equipment given by a multielement antenna. The remarkable amount of research dedicated to this technology pointed out two main benefits to be gained from the deployment of an antenna array at one end of the radio link. The first such benefit is that related to transmit or receive *beamforming* techniques, the second to transmit or receive spatial *diversity* methods.

Beamforming techniques allow the antenna array to transmit (or receive) multiple data streams to (or from) single antenna terminals with controlled interference. This result is obtained by appropriate linear precoding of the data stream to be transmitted though the antenna array (transmit beamforming), or by linear filtering of the received signal by the array (receive beamforming). The interference rejection capability of a multiantenna system, and therefore the number of users that can be simultaneously served, is related to the number of antenna elements [3].

On the other hand, spatial diversity techniques are based on the reception and combination of different copies of the same trasmitted signal, arrived to the receiver through independent channels. These independent channels can be obtained either by transmitting the same signal from different antennas (transmit diversity) or by receiving the signal over multiple antennas (receive diversity). In this way, the probability of an outage event, i.e., of the received signal power being below a given threshold, is drastically reduced, and the performance of the system remarkably improved. The degree of improvement depends on the number of independent channels (i.e., antennas), also referred to as degreee of diversity [4].

In hindsight, the idea behind MIMO systems is essentially that of combining the benefits of beamforming and spatial diversity technologies at the receive and transmit side, by deploying multiple antennas at both ends of the link. In this way, the radio link can enhance the diversity degree up to the product of the number of transmit and receive antennas. Moreover, it can build multiple parallel spatial channels between transmitter and receiver, up to the minimum between the number of transmit and receive antennas [1]. In a recent work [5], Tse defined

the first effect as *diversity gain* and the latter as *multiplexing gain* and found the achievable trade-off between these two benefits.

Two factors are crucial in determining the performance of a MIMO system, i.e., its capability to exploit the diversity and multiplexing gain. The first is the propagation environment: one of the first widely recognized result in this regard showed that if the wireless channel does not present sufficiently rich scattering, the performance of a MIMO system can degrade up to that of a simple single antenna link. The other factor is the availability of some information about the propagation channel (*Channel State Information*, in short CSI) at the receiver and possibly the transmitter. Since the propagation environment is time-varying due to, e.g., the movement of the two ends of the link, this information may be related either to the state of the channel at a given time instant (instantaneous CSI) or to long-term features of the propagation, such as channel statistics (see discussion below).

The first and most commonly studied situation is that of perfect CSI available at the receiver.(and possibly at the transmitter) over a simple frequency-flat channel model. References [1] [2] [7] [9] investigated the MIMO link in this scenario from an information theoretic perspective, whereas linear precoding and equalization signal processing algorithms have been studied in [18]. However, in real world, the propagation environment may be more challenging than a simple frequency-flat scenario and the receiver and transmitter generally can not have a perfect knowledge of the CSI.

In fact, in many propagation environments, given the high data rate of current systems, the wireless channel has to modelled as frequency-selective and therefore inter-symbol interference has to be accounted for[1]. Moreover, the two ends of the link have to estimate the propagation channel based on the received signal. This estimate is inherently affected by an error that, though it can be alleviated by increasing the observation interval devoted to the channel estimation process or exploiting some prior information on the propagation, can not be eliminated. Information theoretical analyses of MIMO system with imperfect CSI at the receiver have been proposed in [10] [11].

The transmitter can either directly measure the CSI by exploiting the channel reciprocity in a Time Division Duplex (TDD) link or else obtain some CSI through a feedback channel by the receiver. Therefore, in this latter case, the CSI available at the transmitter is not only affected by an estimation error, but can also be outdated, as compared to the actual state of the propagation, because of the feedback delay. Information theoretic results related to this scenario have been presented in [12]. Moreover, toward the goal of reducing the spectral resources to be employed in order to guarantee the availability of CSI to the transmitter, signal processing algorithms at the transmit side based only on long-term CSI have been recently studied [14]-[17]. An extensive coverage of information theoretic results related to different CSI condition can be

---

[1]Or, for multicarrier transmission, the additional degree of freedom related to the frequency domain has to be taken into account.

found in [19].

The body of research illustrated above has the major limitation of constraining the analysis to simplified models for the propagation channel and consequently for the CSI available at the two ends of the link. The use of realistic channel models in information theoretic and signal processing investigations is exactly one of the most important issues that remain to be addressed in order to make the MIMO technology feasible. This effort has been attempted by a few authors. In particular, Shiu et al [7] introduced spatial correlation in frequency-flat channels while Raleigh and Cioffi [8] and Bolckei et al. [20] considered frequency-selective channels. What is still missing is a comprehensive and analytically tractable channel model, able to describe different environments. Based on this model, it would then possible to clearly define different degrees of (long-term or instantaneous) CSI and therefore develop of more realistic analyses from information theoretic and signal processing standpoints.

1.2   Overview of the work

This thesis is concerned with the investigation of single and multiuser MIMO systems from both an information theoretic and signal processing perspective by making use of a realistic and analytically tractable algebraic channel model. In particular, the treatment builds on a fundamental property of the fading channels, that is illustrated below. A general MIMO link is depicted in fig. 1.1. N_T antennas at the transmitter side communicate over a time-varying and frequency selective channel $\mathbf{H}(t, \tau)$ with N_R antennas at the receiver side. The transmitting antennas, and the same goes for the receiving antennas, may either belong to a single terminal or be partitioned among multiple terminals. In the first case, all the transmitting (or receiving) antennas cooperate when processing the signal to be transmitted (or the received signal). On the other hand, in the latter case, only the antennas belonging to the same terminal perform a cooperative signal processing. This thesis will be concerned with the single user case, and the multiuser case limited to the downlink and uplink scenarios. The author recalls that the downlink corresponds to the situation where multiple users are present at the receiving end, while the uplink accounts for the dual case (multiple users at the transmitting end).The time-varying and frequency-selective channel is described by a $N_R \times N_T$ matrix $\mathbf{H}(t, \tau)$, that depends on both the time-instant t and on the delay τ [6]. Accordingly, the signal received over the N_R receiving antenna $\mathbf{y}(t)$ is the convolution between the channel $\mathbf{H}(t, \tau)$ and the $N_T \times 1$ signal $\mathbf{x}(t)$ transmitted by the $N_T \times 1$ transmitting antennas:

$$\mathbf{y}(t) = \int \mathbf{H}(t, \tau)\mathbf{x}(t - \tau)d\tau + \mathbf{n}(t), \tag{1.1}$$

where $\mathbf{n}(t)$ is the additive gaussian noise accounting for both thermal noise and spatially distributed interferers.

In wireless systems, the channel $\mathbf{H}(t, \tau)$ accounts for scattering, reflection and diffraction of the radiated wave that propagates from the transmitter to the receiver. These propagation
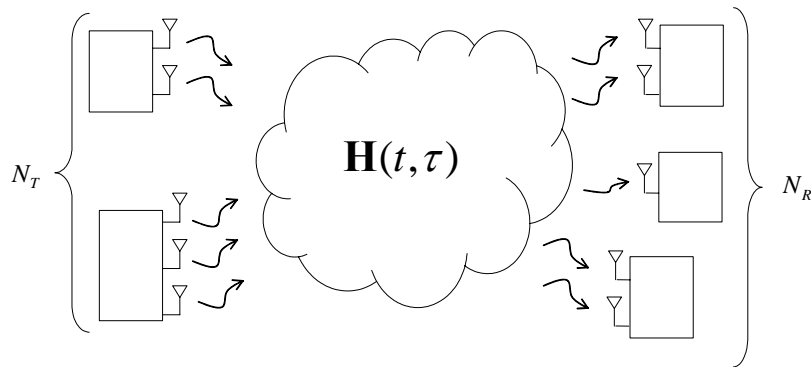
Figure 1.1: Illustration of a general MIMO system.

effects cause the signal to be received through multiple paths, collectively referred to as *multipath*. The characteristics of the multipath channel depend on both the propagation environment and the geometry of the antenna arrays at the ends of the link. For instance, in a given environment, it is expected that placing two antennas at a larger distance would entail a smaller degree of correlation between the channel gains relative to the two antennas.

1.2.1 Long-term channel features and fast varying fading process

The work presented in this thesis is mainly based on a fundamental property of the multipath channel, that holds for a wide range of assumptions regarding the propagation environment and the array geometries. In particular, the multipath channel can be conveniently studied by considering that the time-variability of the propagation is due to two classes of parameters with different rates of variations. To be specific, consider for instance the propagation delay for a given path. This parameter is related to large scale geometrical features of the environment, i.e., on the order of many wavelengths. On the other hand, the fading process, due to the constructive and destructive combination of multiple paths, is expected to vary for movements on the order of a fraction of a wavelength. Therefore, the propagation delay can be considered to have temporal variations much slower (at least one-two orders of magnitude) than the fading process. Other multipath parameters share this property with the propagation delays, such as for instance the power-delay profile, that is related to path loss and shadowing phenomena. The distinction between *long-term* (i.e., slow varying) features of the multipath channel and the fast varying fading process is depicted for illustration purposes in fig. 1.2.

1.2.2 Block fading assumption and algebraic structure of the channel

In this thesis, it is assumed that the transmission from each antenna is synchronized and organized into blocks, as shown in fig. 1.3. Each block can be either transmitted directly in the time-domain or else, for multicarrier transmission (MIMO-OFDM), in the frequency-domain.
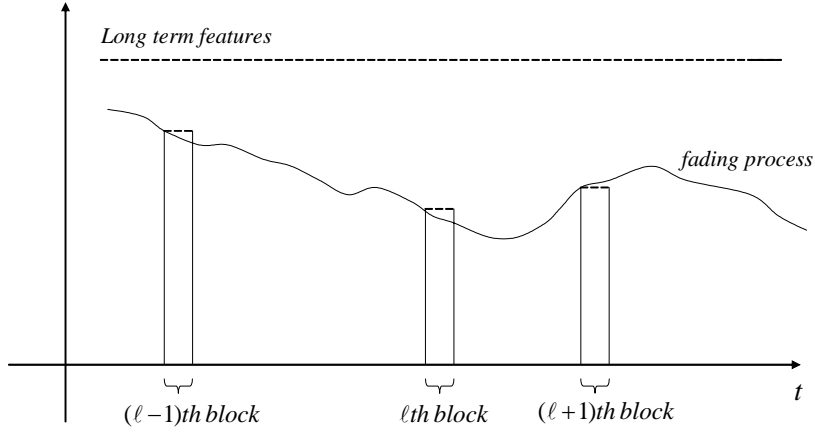
Figure 1.2: Illustration of the time variability of the long term features of the multipath channel and the fast varying fading process.

Moreover, each block contains both data symbols, that carry useful information, and training symbols, that are known to the receiver and used for channel estimation. Through the observation of the training part of the block, the receiver is able to acquire the CSI by estimating the propagation channel. This CSI is then used for detecting the data symbols and possibly fedback to the transmitter, as discussed below (see also fig. 1.4).

Accoding to fig. 1.2, the blocks are generally transmitted in a discontinuous way (i.e., packet transmission). Moreover, according to the *block-fading assumption,* the time variations of the channel, and specifically of the fading process, within each block are neglected. In other words, the channel matrix $\mathbf{H}(t, \tau)$ is modelled as constant within each block, i.e., for the ℓth block we have

$$\mathbf{H}(t, \tau) = \mathbf{H}_\ell(\tau) \; t \in \ell\text{th block}.$$

Furthermore, since this thesis is concerned with symbol (T)-spaced receivers, the propagation channel is completely described by the W matrices of size $N_R \times N_T$

$$\mathbf{H}_\ell[m] = \mathbf{H}_\ell(mT),$$

with $m = 0, ..., W - 1$, where we have assumed that the (causal) channel $\mathbf{H}_\ell(\tau)$ is supported within $0 \leq \tau \leq WT$.

Most of the work in this thesis builds on a fundamental result that capitalizes on the distinction between long-term channel features and fast varying fading process. In fact, the author proves that the channel vector $\mathbf{h}_\ell$, appropriately defined from the channel matrices $\mathbf{H}_\ell[m]$, has an algebraic structure that allows a simple decoupling of these two classes of parameters. Notice that $\mathbf{h}_\ell$ describes the the channel characteristics both in the spatial and in the time (delay)
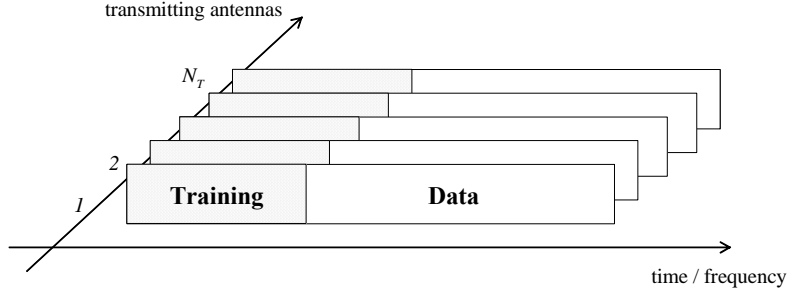
Figure 1.3: Block transmission either in time and frequency domain with training and data symbols.

domain. To be more specific, it is shown that, under wide assumptions about the propagation environment and the array geometries, the long-term features of the channel only determine the *space-time channel correlation matrix* $\mathbf{R}_H = E[\mathbf{h}_\ell \mathbf{h}_\ell^H]$. Moreover, this matrix is generally rank-deficient, due to the limited spatial and temporal resolution of the receiver. Therefore, it is completely defined by long-term *space-time modes* of the channel, that are identified by the principal eigenvectors of the channel correlation matrix $\mathbf{R}_H$. The fading process then only accounts for the fast-varying amplitudes to be assigned to each channel mode within each block. In other words, while the *space-time modes* (or equivalently the space-time correlation matrix) are slowly varying over the transmitted blocks since they depend only on the long-term channel features, the corresponding amplitudes vary from block to block according to the fading process.

1.2.3 System overview

The implications of the result discussed above are investigated throughout the thesis for both the CSI acquisition phase, carried out by processing of the training part of the block, and the detection phase, corresponding to the processing of the data part of the block. Notice that in this thesis the author does not consider joint processing of data and training symbols, that has been recently advocated within the context of iterative receivers (see, e.g., [21]). A block diagram of the transmitter and receiver is shown in fig. 1.4. Therein, for simplicity, a single user MIMO link (i.e., fully cooperative on both sides) is depicted. The extension to a multiuser scenario will be discussed later on.

This thesis is mainly concerned with *linear signal processing*. Therefore, except in a few cases, all the processing blocks that are considered consist in linear operators. As shown in fig. 1.4, training and data symbols are multiplexed to form a the N_T blocks (see fig. 1.3) that are mapped onto the N_T transmitting antennas[2]. Notice that the arrangement of training and

---

[2]If multicarrier transmission is employed, the mapping operation includes the IDFT operation needed to convert
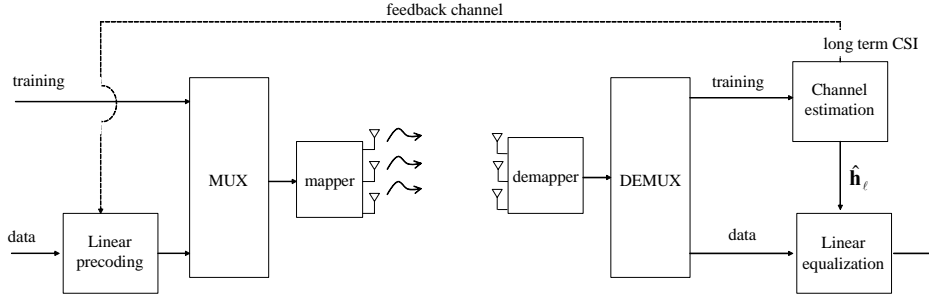
Figure 1.4: Block diagram of transmitter and receiver for a single user MIMO link.

data symbols in fig. 1.3 is only for illustration purposes since the training data could be placed according to different criteria in the (time or frequency-domain) block [22]. At the receiver side, after demultiplexing, the training symbols are used in order to obtain an estimate of the propagation channel $\hat{\mathbf{h}}_\ell$. Morever, they can be possibly exploited in order to compute the long-term features of the channel, i.e., according to the discussion above, the space-time channel modes. In other words, the channel estimation block perform the acquisition the *instantaneous CSI* $\hat{\mathbf{h}}_\ell$ and, possibly, of the *long-term CSI* $\mathbf{R}_H$ (or equivalently of the space-time modes). Notice that the CSI is inherently imperfect beacuse it is the result of an estimation process.

The instantaneous (imperfect) CSI is provided for the design of the linear equalizer that processes the data symbols. On the other hand, the long-term CSI can be possibly fedback to the transmitter, on a low rate feedback channel, in order to allow the design of a linear precoder. The latter can greatly improve the overall performance of the system by capitalizing on the side information at the transmitter. In principle, the receiver could direcly feedback the estimate $\hat{\mathbf{h}}_\ell$, but this would require to access the feedback channel on each block. On the other hand, by feeding back the long-term CSI, the feedback channel has to be accessed only when this information has significantly changed, which is expected to occur after a large number of blocks.

1.2.4   CSI acquisition based on modal analysis

This thesis is divided in two parts. The first part of the thesis is dedicated to the problem of CSI acquisition from the observation of the training part of received blocks. Building on the results obtained in the first part, the second is devoted to the analysis of precoding and equalization, i.e., to the processing of the data symbols within each block. In both parts, the focus is the investigation of the system performance when the realistic channel model discussed above, that assumes the distinction between long-term channel features and fast varying fading process, is considered.

---

the transmitted block from frequency to time domain.

Channel estimation for MIMO links is an active field of research. Most of the current literature relies on simple unconstrained ML estimate [10] [23]. Design of efficient channel estimators based on the exploitation of the the different varying rates of the channel features has been attempted by a few authors. In particular, such techniques have been developed for SIMO systems in [24] [25] [26]. These methods are based on a non-linear estimation of the long-term parameters and therefore suffer from typical impairments of these type of estimators (threshold effects). Moreover, they are quite sensitive to modelling mismatches, such as poor antenna calibration or residual timing offsets.

Here, the author starts by tackling this problem from a theoretical point of view. In particular, the impact of the assumption about the different varying rates on channel estimation is investigated by deriving a lower bound on the performance of any unbiased channel estimator. This goal is achieved through the computation of the hybrid CRB, a modification of the conventional CRB that is appropriate for the problem at hand [27]. In this way, a theoretical limit on the accuracy of the instantaneous CSI given channel and system parameters is determined.

The analysis suggests an asymptotically optimal strategy to perform channel estimation, that consists in a separate estimation of long-term channel features and fast varying fading amplitudes. In this thesis, a linear channel estimation algorithm that is able to attains the theoretical perfomance bound and that overcomes the limitations of the non-linear estimators discussed above is proposed. It is based on the direct estimation of the space-time modes (referred to as *modal analysis*) through a subspace tracking algorithm and the estimation of the fading amplitudes through least squares techniques.

1.2.5 Linear precoding and equalization

The second part of the thesis is devoted to the analysis of precoding and equalization, i.e., to the processing of the data symbols within each block. At first, the impact of imperfect CSI available at the trasmitter and receiver is studied from an information theoretic perspective within both single and multi-user systems with linear precoding. This part owes to the work of Medard [28] that derived a lower bound on the information rate for a SISO system with imperfect CSI.

Then, linear (and possibly non-linear) precoding and equalization based on long-term CSI at the transmitter is studied for a single user MIMO link. Similar works that did not consider the possibility to include non-linear blocks have been proposed in [15] [16]. Finally, linear precoding and equalization for multi-user MIMO systems is considered. Since for this scenario, precoding based on long-term CSI is still an open problem, the thesis concentrates on linear precoding based on instantaneous CSI and investigate the possibility to perform the joint design of precoding, decoding and scheduling algorithms for the downlink of a multiuser MIMO system. This study leverages on the work from Haardt et al. [29] that first tackled this problem by proposing a disjoint design of the quantities of interest.

1.3 Thesis contributions

The orignal contributions of this thesis can be listed as follows. Notice that the organization of this material in different Chapters and the related publications are discussed in the next Section.

- Definition of an analytically tractable channel model suitable for indoor and outdoor environments. The model is expressed in compact matrix notation and accounts for multipath with non-integer delays, space-time fading correlation, angles of departure/arrival and arbitrary power-delay profile. The analysis shows that the multipath channel can be parametrized as the combination of two matricial terms: the first contains the space-time modes of the channel and depends on long-terms parameters (delay, angles, correlation fuctions) whereas the second gathers the corresponding time-varying amplitudes and depends on the fading process. This model is used in order to account for both time-domain and multicarrier transmission so that all the results reported below hold for both scenarios.

- Computation of a lower bound on the channel (or CSI) estimation error for single and multiuser MIMO systems over frequency-selective channels through evaluation of the hybrid CRB. This analysis allows to determine from a theoretical point of view the impact of channel and system parameters on the perfomance of the CSI acquisition step, given the realistic channel model discussed above.

- Design of linear channel (or CSI) estimators based on the exploitation of the different varying rates of the channel features, for single and multiuser MIMO systems. The proposed estimators perform a separate computation of long-term characteristics and fast varying fading amplitudes. In particular, the long-term estimation is obtained by a direct evaluation of the space-time modes of the channel through subspace tracking (modal analysis) and the fading process is tracked by computing the corresponding amplitudes by least squares techniques. The performance of the proposed estimators is validated both analytically and through simulations for practical systems showing their ability to attain the perfomance bound discussed above.

- Information theoretic analysis of the performance of a single and multiuser MIMO systems with imperfect CSI at the transmitter and the receiver. The single user case is approached by assuming a frequency-selective channel and capitalizing on the results discussed above. For the multiuser case, downlink with linear precoding is considered over a frequency-flat channel in order to simplify the analysis. In this latter situation the impact of both channel estimation errors and feedback delay is investigated.

- Linear/non linear precoding and equalization based on long-term CSI at the transmitter for a single user MIMO system over a frequency-flat channel. This study generalizes well

studied transceivers, such as V-BLAST [30] and Tomlinson-Harashima Precoding [17], and evaluates the impact of the assumption of long-terms CSI available at the transmitter.

- Joint design of linear precoding, equalization and scheduling for the downlink of a multiuser MIMO system over a frequency-flat channel. The design is aimed at either maximizing the sum rate or ensuring some fairness constraints among the users. Herein, it is assumed that instantaneous CSI is available at both receiver and transmitter.

- Design of the adaptive placement of training symbols within each transmitter block based on long-term CSI at the transmitter. This study is presented for a single user multicarrier system.

1.4   Outline of the thesis and related publications

The organization of the thesis and the original publications where the contents of each Chapter have been presented are disclosed below.

Chapter 2: the basic result on the algebraic structure of the channel vector, decoupling long-term features and fast varying fading process, is presented. For the corresponding references, the reader is refer to Chapter 4.

Chapter 3: presents a review of the signal model for both time domain and multicarrier transmission. The treatment is aimed at showing the duality of the two approaches, that will be treated throughout the thesis with a unified notation. Both the single user and multi user cases are considered.

Part I: CSI acquisition

Chapter 4: derives the lower bound on the channel estimation error for a MIMO system over a frequency-selective channel. The performance limit is obtained by computing the hybrid CRB and taking into account the distinction between long-term channel features and the fast-varying fading process. This analysis allows to determine, from a theoretical standpoint, the effects of system and channel parameters on the accuracy of the CSI available for system optimization. Moreover, it provides, as a by-product, useful indication concerning the optimal channel estimation strategy. This result is capitalized upon in the next Chapter. Original references for the material presented in this Chapter are:

- *O. Simeone* and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels", *IEEE Trans. on Signal Processing,* vol. 52, no. 11, pp. 3265-3277, Nov. 2004.

- *O. Simeone* and U. Spagnolini, "Lower bounds on the channel estimation error for fast-varying frequency-selective Rayleigh MIMO channels", *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 5, pp. 69-72, April 6-10, 2003.

- *O. Simeone* and U. Spagnolini, "Channel estimation for block-fading frequency-selective Rayleigh MIMO channels: performance limits", *Proc. IEEE Vehicular Technology Conference*, May 17-19, Milan, 2004.

- *O. Simeone* and U. Spagnolini, "Hybrid CRB in linear regression for random parameters with unknown and rank deficient correlation matrix," submitted.

Chapter 5: discusses linear channel estimation algorithms that are based on a separate computation of long-term channel features and the fast-varying fading process. The long-term features are estimated by directly computing the space-time modes of the channel through subspace tracking (modal analysis) whereas the fading process is evaluated by tracking the fading amplitudes relative to different channel modes. The treatment encompasses single and multiuser systems. Performance of channel estimation based on modal analysis is studied both analytically and by simulations for practical systems, showing that, under appropriate conditions, it attains the theoretical bound derived in the previous Chapter. This material was originally disclosed in the following references:

- M. Nicoli, *O. Simeone* and U. Spagnolini, "Multi-slot estimation of fast-varying space-time channels in TD-CDMA systems", *IEEE Comm. Letters*, vol. 6, no. 9, pp. 376-378, Sept. 2002.

- M. Nicoli, *O. Simeone* and U. Spagnolini, "Multi-slot estimation of fast-varying space-time communication channels", *IEEE Trans. Signal Processing*, vol. 51, no. 5, pp. 1184-1195, May 2003.

- M. Nicoli, *O. Simeone* and U. Spagnolini, "Multi-slot estimation of fast-varying frequency-selective channels", *IEEE Trans. on Communications*, vol. 51, no. 8, pp. 1337-1347, Aug. 2003.

- *O. Simeone*, Y. Bar-Ness and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay subspace", *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 315-325, Jan. 2004.

- *O. Simeone* and U. Spagnolini, "Multi-slot estimation of space-time channels", *Proc. IEEE International Conference on Communications (ICC 2002)*, New York City, April 28-May 2, 2002.

- M. Nicoli, *O. Simeone* and U. Spagnolini, "A multi-slot method to estimate the fast-varying channels in TD-CDMA systems", *Proc. XI European Signal Processing Conference (EUSIPCO 2002)*, Toulouse, September 3-6, 2002.

- M. Aldrovrandi, M. Nicoli, *O. Simeone* and U. Spagnolini, "Subspace tracking methods for channel estimation in TD-SCDMA systems", *Proc. European Conference on Wireless Technology (ECWT 2002)*, Milan, September 26-27, 2002.

- *O. Simeone*, Y. Bar-Ness and U. Spagnolini, "Subspace based methods for channel estimation in OFDM systems", *Proc. European Conference on Wireless Technology (ECWT 2002)*, Milan, September 26-27, 2002.

- M. Cicerone , *O. Simeone*, N. Geng and U. Spagnolini, "Modal analysis/filtering to estimate time-varying MIMO-OFDM channels, *Proc. ITG Workshop on Smart Antennas (WSA)*, March 18-19, Munich, 2004.

- M. Aldrovrandi, M. B. Nicoli, *O. Simeone* and U. Spagnolini, "Method and device for channel estimation in digital radio communications", pending patent, no. EP 02425260.3, April 23, 2002.

- L. Moretti, M. Nicoli, *O. Simeone* and U. Spagnolini. "Method for the estimation and tracking of channel modes in multicarrier communication systems", pending patent, no. EP 03425721.2, November 5, 2003.

- M. Cicerone, *O. Simeone* and U. Spagnolini, "Channel estimation for MIMO-OFDM systems by modal analysis/filtering," submitted.

Part II: Precoding and equalization

Chapter 6: discloses a lower bound on the information rate for single and multiuser systems with imperfect CSI at the transmitter and the receiver. The single user case is approached by assuming a frequency-selective channel and capitalizing on the results discussed in Chapter 4. For the multiuser case, downlink with linear precoding is considered over a frequency-flat channel in order to simplify the analysis. In this latter situation the impact of both channel estimation errors and feedback delay is investigated. The single user case was originally presented in the first references listed above for Chapter 4, whereas the analysis of the downlink is discussed in:

- G. Primolevo, *O. Simeone* and U. Spagnolini, "Effects of imperfect channel state information on the capacity of broadcast OSDMA-MIMO systems", *Proc. IEEE Signal Processing Advances on wireless communications*, July 11-14, Lisbon, 2004.

Chapter 7: linear/non linear precoding and equalization are designed based on long-term CSI at the transmitter for a single user MIMO system over a frequency-flat channel. This study generalizes well studied transceivers, such as V-BLAST and Tomlinson-Harashima Precoding, and evaluates the benefits that can be gained by exploiting the side information provided by the long-term CSI at the transmitter. This material has been originally presented in the following references:

- *O. Simeone* and U. Spagnolini, "Combined linear pre-equalization and BLAST equalization with channel correlation feedback", *IEEE Communications Letters*, vol. 7, no. 10, pp. 487-489, Oct. 2003.

- *O. Simeone*, Y. Bar-Ness and U. Spagnolini, "Linear and non-linear precoding/decoding for MIMO systems with long-term channel state information at the transmitter", *IEEE Trans. on Wireless Communications*, vol. 3, no. 2, pp. 373-378, March 2004.

- *O. Simeone*, U. Spagnolini and Y. Bar-Ness, "Linear and non-linear precoding/decoding for MIMO systems using the fading correlation at the transmitter", *Proc. IEEE Signal Processing Advances on Wireless Communications (IEEE SPAWC 2003)*, Rome, June 15-18, 2003.

Chapter 8: discusses the joint design of linear precoding, equalization and scheduling for the downlink of a multiuser MIMO system over a frequency-flat channel. The design is aimed at either maximizing the sum rate or ensuring some fairness constraints among the users. Herein, it is assumed that instantaneous CSI is available at both receiver and transmitter. This work has been presented in the reference:

- G. Primolevo, *O. Simeone* and U.Spagnolini, "Channel aware scheduling for broadcast MIMO systems with orthogonal linear precoding and fairness constraints," *Proc. IEEE ICC 2005*, Seoul, May 16-19, 2005.

Chapter 9- Appendix: the appendix investigates a degree of freedom that was not considered throughout the thesis, i.e., the possibility to select a different number of training symbols in each transmitting block. This adaptive, block-by-block, selection is made possible since long-term CSI is assumed to be available at the transmitter. The latter, based on this side information, is able to predic the perfomance of channel estimation at the receiver and therefore adjust the number of training symbols according to the desired CSI accuracy at the receiver. This study is presented for a single user multicarrier system. This material was disclosed in:

- O. Simeone and U. Spagnolini, "Adaptive pilot pattern for OFDM systems", *Proc. IEEE International Conference on Communications*, June 20-24, Paris, 2004.

All the papers can be downloaded at: www.elet.polimi.it/upload/simeone.

Bibliography

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs internal report, June 1995

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Personal Communications, Vol. 6, No. 3, pp. 311–335, March 1998.

[3] F. Rashid-Farrokhi, K.J.R. Liu and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE Selected Areas in Commun.*, vol. 16, no. 8, pp. 1437-1450, Oct. 1998.

[4] J. H. Winters, "The diversity gain of transmit diversity in wireless systems with Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 47, no. 1, pp. 119-123, Feb. 1998.

[5] D. N. C. Tse, P. Viswanath and Lizhong Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1859-1874, Sept. 2004.

[6] P. A. Bello, "Characterization of randomly time-variant linear channel," *IEEE Trans. Circuits and Systems*, vol. 11, no. 4, pp.360-393, Dec. 1963.

[7] D. Shiu, G. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, Vol. 48, No. 3, pp. 502–513, March 2000.

[8] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, Vol. 46, No. 3, pp. 357 – 366, March 1998.

[9] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, Vol. 47, No. 4, pp. 1273–1289, May 2001.

[10] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-963, April 2003.

[11] H. Vikalo, B. Hassibi, B. Hochwald and T. Kailath, "On the capacity o frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2572-2583, Sept. 2004.

[12] D. Samardzija and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Trans. Signal Proc.*, vol. 51, no. 11 , pp. 2882-2890, Nov 2003.

[13] T. Marzetta and B. Hochwald, "Fast Transfer of Channel State Information in Wireless Systems," submitted to *IEEE Trans. Signal Proc*. [also available on http://mars.bell-labs.com/cm/ms/what/mars/papers/channel_estimation/]

[14] G. Jongren, M. Skoglund and B. Ottersten, "Combining beamforming and orthogonal space-time block coding ," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 611-627, Mar. 2002.

[15] H. Sampath and A. Paulraj, "Linear Precoding for Space-Time Coded Systems," *IEEE Comm. Letters*, vol. 6, pp. 239-241, June 2002.

[16] M. T. Ivrlac, T. P. Kurpjuhn, C. Brunner and W. Utschick, "Efficient use of fading correlations in MIMO systems," *Proc. VTC 2001*, vol.4, pp. 2763-2767, 2001.

[17] F. H. Fischer, C. Windpassinger, A. Lampe and J. B. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," *Proc. International Zurich Seminar on Broadband Communications*, pp. 7 1-7 6, 2002.

[18] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051-1064, May 2002.

[19] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684-702, June 2003.

[20] H. Bolckei, D. Gesbert and A.J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," IEEE Trans. Comm., vol. 50, no. 2 , pp. 225-234, Feb 2002.

[21] A. Kocian and B. H. Fleury, "EM-based joint data detection and channel estimation of DS-CDMA signals," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1709-1720, Oct. 2003.

[22] S. Adireddy, L. Tong and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2338-2353, Aug. 2002.

[23] T. L. Marzetta, "Blast training: estimating channel characteristics for high capacity space-time wireless," Proc. 37th Annual Allerton Conference on Communication, Control, and Computing, pp. 958–966, September 1999.

[24] A. Van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint Angle and Delay Estimation Using Shift-Invariance Techniques", *IEEE Trans. Signal Processing*, vol 46, pp 405-418, Feb. 1998.

[25] Yung-Yi Wang, Jiunn-Tsair Chen and Wen-Hsien Fang, "TST-MUSIC for Joint DOA-Delay Estimation," *IEEE Trans. Signal Processing*, vol. 49, pp. 721-729, April 2001.

[26] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing,* vol. 46, no. 2, pp. 449-462, Feb. 1998.

[27] H. L. Van Trees, *Optimum array processing*, Wiley, 2002.

[28] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933-946, May 2000.

[29] Q. H. Spencer, A. L. Swindlehurst and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Proc.*, vol. 52, no. 2, pp. 461-471.

[30] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

Chapter 2

Algebraic structure of time-varying frequency-selective MIMO channels

2.1 Introduction

IN wireless systems, the signal propagating from the transmitter to the receiver arrives at the destination through different paths, collectively referred to as multipath. The multiple paths are produced by scattering, reflection and diffraction of the radiated wave by the objects present in the environment. Each path is characterized by a set of parameters, such as propagation delay, mean power (caused by path loss and shadowing) and fast fading amplitude [1]. Moreover, in a multiantenna setting, the spatial features of each path need to be specified, such as angle of arrival (AOA) and departure (AOD) or spatial correlation of the fading amplitudes according to the propagation environment (see below). The reader is referred to the textbooks [1] [2] for details.

In a time-varying environment, the multipath parameters experience different rates of variations. In fact, while the fast fading amplitudes depend on small scale features of the propagation environment, on the order of a fraction of a wavelength, other parameters, such as delays, angles and spatial correlation, vary across distances of many wavelengths [3] [4]. The temporal variation of a typical outdoor channel for a SIMO system is shown in fig. 2.1 in terms of its power delay-angle diagram (received power versus propagation delay and AOA). As it can be seen, delays and angles relative to different paths vary much slowly than the power fluctuations caused by fast fading.

Characterization of multipath parameters in different frequency bands and scenarios has been the object of investigation of many measurement campaigns (see, e.g., [5] [6]). As an outcome of these efforts, several standardization groups have defined channel models to be used for performance evaluation through computer simulation [7] [8].

In this chapter, an analytical framework for the definition of MIMO channel models that encompasses multipath propagation with non-integer delays, space-time fading correlation,
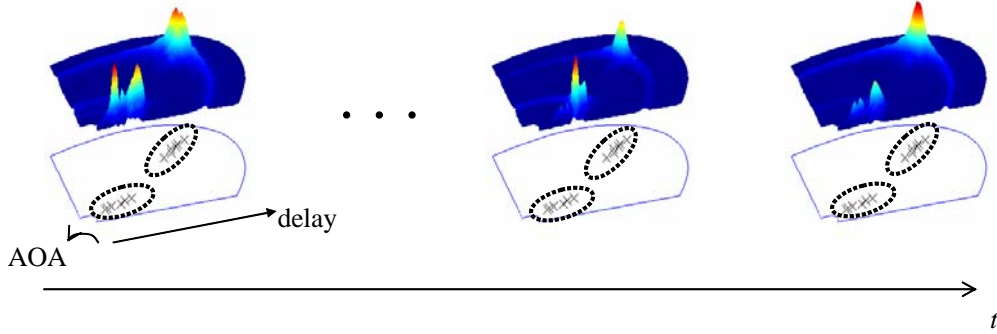
Figure 2.1: Temporal variation of the power delay-angle diagram of a typical outdoor channel for a SIMO system.

angles of departure/arrival and arbitrary power-delay profile is proposed. In particular, two scenarios are considered (see fig. 2.2):

1. *Beamforming scenario* [9]: the elements of both transmitting and receiving antenna arrays are co-located and the scatterers can be considered as point sources. Each path of the multipath channel is then characterized by a AOD and AOA that "steer" the array response at the transmitter and receiver respectively, a delay and a complex amplitude (fading). The latter is in turn modelled as a temporally correlated Gaussian stationary process. As a general rule, this model appears to be well suited for *outdoor* channels [5] and has been used in a SIMO context in, e.g., [3], [10].

2. *Diversity scenario* [11]: the elements of both the transmitting and receiving antenna arrays are not co-located and/or the different scatterers have to be modelled as distributed sources (see e.g., [12]). These assumptions are generally appropriate for an *indoor* environment [6]. For each delay of the multipath, the channel gains between different transmitting and receiving antennas can then be modeled as spatially and temporally correlated jointly Gaussian random variable with zero mean (Rayleigh fading). This model has been used (with some simplifications) by [13], and [14] in the context of MIMO-OFDM transmission.

The analysis of the algebraic properties of the channel vector $\mathbf{h}$, gathering the MIMO channel gains as rigorously defined below, shows that the contribution *long-term parameters*, such as delay, angles and correlation functions, and *fast-varying fading amplitudes* can be analytically decoupled. The main result will be that the channel vector can be written as the product of a slowly varying matrix $\mathbf{T}$, depending on the long-term parameters, and a vector $\boldsymbol{\beta}$ gathering the fading amplitudes:

$$\mathbf{h} = \mathbf{T}\boldsymbol{\beta}. \tag{2.1}$$
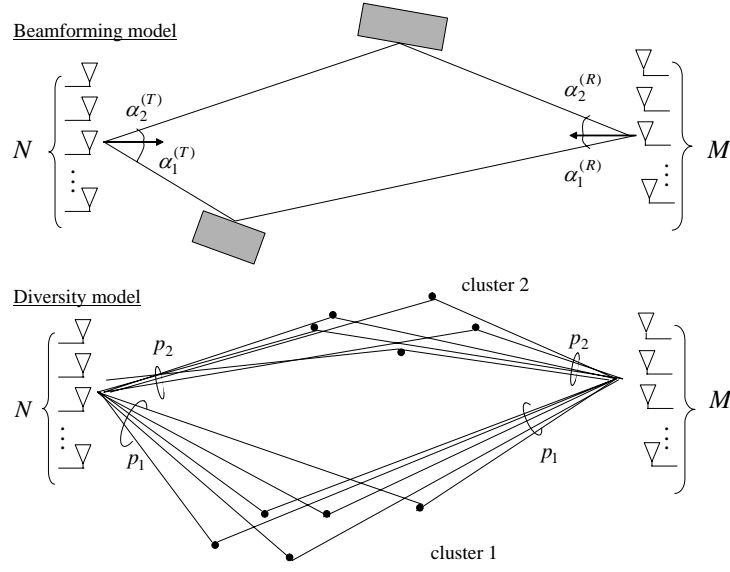
Figure 2.2: Beamforming and diversity channel scenarios.

This factorization not only yields insight on the structure of time-varying propagation but also sets the ground for the analysis and design of the channel estimation process, to be discussed in Chapters 4 and 5.

2.2 Multipath model of a MIMO frequency-selective channel

For the sake of illustration, herein we consider a single user MIMO link with N_T antennas at the transmitter and N_R at the receiver side (see fig. 2.2). In a multipath environment, communication over the link between the n_Tth transmitting antenna and and the n_Rth receiving antenna ($n_T = 1, ..., N_T$ and $n_R = 1, ..., N_R$) is impaired by the linear convolution of the transmitted signal with the channel impulse response $h_\ell^{(n_T, n_R)}(\tau)$. As explained in the previous Chapter, the channel coherence time is assumed to be larger than the time slot duration so that the temporal variation of $h_\ell^{(n_T, n_R)}(\tau)$ is accounted for by the subscript ℓ that runs over the blocks.

According to the convention adopted in most of the existing literature, the channel $h_\ell^{(n_T, n_R)}(\tau)$ is defined as the convolution of the physical propagation channel with the cascade of the transmitted baseband pulse and the receiving filter, denoted as $g(\tau)$. Therefore, the channel impulse response over the (n_T, n_R)th link within the ℓth block is described by the combination of d paths, each corresponding to a delayed replica of the waveform $g(\tau)$

$$h_\ell^{(n_T, n_R)}(\tau) = \sum_{i=1}^{d} \sqrt{\Omega_{i,\ell}} \cdot a_{i,\ell}^{(n_T, n_R)} \cdot g(\tau - \tau_{i,\ell}). \tag{2.2}$$
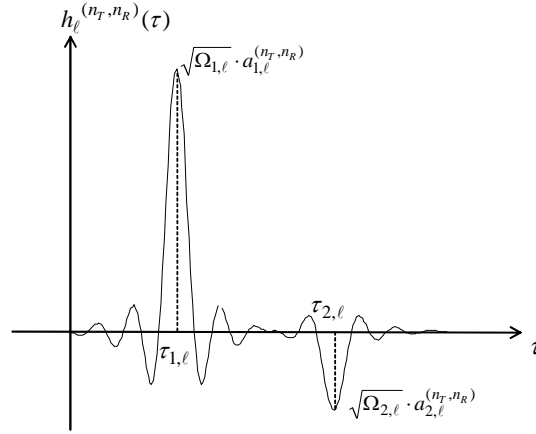
Figure 2.3: Example of the channel impulse $h_\ell^{(n_T,n_R)}(\tau)$ for $d = 2$ paths and a Nyquist wave-form $g(\tau)$ with roll-off 0.2.

Each path is characterized by delay $\tau_{i,\ell}$, mean power $\Omega_{i,\ell}$ (caused by path loss and shadow-ing) and fast fading amplitude $a_{i,\ell}^{(n_T,n_R)}$. Fig. 2.3 shows an example of the channel impulse $h_\ell^{(n_T,n_R)}(\tau)$ for $d = 2$ and a Nyquist waveform $g(\tau)$ with roll-off 0.2.

In this thesis, detection after sampling at symbol rate $1/T$ is considered. Therefore, the characterization of the discrete-time channel impulse response

$$h_\ell^{(n_T,n_R)}[m] = h_\ell^{(n_T,n_R)}(mT) \tag{2.3}$$

is required. Assuming that $h_\ell^{(n_T,n_R)}[m]$ is negligible for $m \geq W$ (and causal), the MIMO channel within the ℓth block is completely characterized by the $N_T N_R W$ samples

$$h_\ell^{(n_T,n_R)}[m] \qquad m = 0, ..., W-1; \; n_T = 1, ..., N_T; \; n_R = 1, ..., N_R. \tag{2.4}$$

For analysis, the channel coefficients (2.4) will be conveniently rearranged in different ma-trices in order to investigate different aspects of the channel structure or to relate transmitted and received signal in a compact form (see next Chapter). The reordering of the channel coef-ficient and the corresponding notation will be clarified within the context of interest whenever necessary.

Here the MIMO channel impulse response obtained by arranging the channel coefficients (2.4) versus the delay index m is considered, obtaining the $N_R \times N_T$ MIMO channel taps (recall (2.2))

$$\mathbf{H}_\ell[m] = \sum_{i=1}^{d} \sqrt{\Omega_{i,\ell}} \cdot g(mT - \tau_{i,\ell})\mathbf{A}_{i,\ell} \quad m = 0, ..., W-1 \tag{2.5}$$

where $[\mathbf{A}_{i,\ell}]_{n_T,n_R} = a_{i,\ell}^{(n_T,n_R)}$. Appropriate definitions of the matrices $\mathbf{A}_{i,\ell}$ specialize the MIMO channel structure (2.5) to beamforming and diversity scenarios as described below.

In order to ease the analysis of the interaction between long-term channel parameters and the fast fading amplitudes (to be discussed in Sections 2.3 and 2.4), the MIMO channel taps $\{\mathbf{H}_\ell[m]\}_{m=0}^{W-1}$ in (2.5) can be rearranged by stacking their columns into the $N_R N_T \times W$ matrix $\mathbf{H}_\ell$

$$\mathbf{H}_\ell = [\text{vec}\{\mathbf{H}_\ell[0]\} \cdots \text{vec}\{\mathbf{H}_\ell[W-1]\}] \tag{2.6}$$

obtaining from (2.5)

$$\mathbf{H}_\ell = \sum_{i=1}^{d} \sqrt{\Omega_{i,\ell}}\, \text{vec}\{\mathbf{A}_{i,\ell}\}\mathbf{g}(\tau_{i,\ell})^T = \boldsymbol{\mathcal{A}}_\ell \cdot \boldsymbol{\Omega}_\ell^{1/2} \cdot \mathbf{G}(\boldsymbol{\tau}_\ell)^T \tag{2.7}$$

where $\mathbf{g}(\tau)$ is the $W \times 1$ vector that gathers the T-spaced samples of the delayed waveform

$$\mathbf{g}(\tau) = [g(-\tau)\, g(T-\tau)\, \cdots g((W-1)T - \tau)]^T, \tag{2.8}$$

the $N_R N_T \times d$ matrix $\boldsymbol{\mathcal{A}}_\ell$ is

$$\boldsymbol{\mathcal{A}}_\ell = [\text{vec}\{\mathbf{A}_{1,\ell}\}, \text{vec}\{\mathbf{A}_{2,\ell}\}, \cdots, \text{vec}\{\mathbf{A}_{d,\ell}\}], \tag{2.9}$$

the $d \times d$ diagonal matrix containing the mean powers reads

$$\boldsymbol{\Omega}_\ell = \text{diag}\{\Omega_{1,\ell}\, \Omega_{2,\ell}, ...\Omega_{d,\ell}\} \tag{2.10}$$

and the $W \times d$ matrix $\mathbf{G}(\boldsymbol{\tau}_\ell) = [\mathbf{g}(\tau_{1,\ell}), ..., \mathbf{g}(\tau_{d,\ell})]$ collects all the delayed waveforms. Moreover, the $d \times 1$ vector of delays $\boldsymbol{\tau}_\ell = [\tau_{1,\ell}...\tau_{d,\ell}]^T$ is defined.

In Sections 2.3 and 2.4 the analytical model (2.7) will be specialized to the beamforming and diversity scenarios respectivelythrough specification of matrix $\mathbf{A}_{i,\ell}$ $i = 1, ..., d$.

2.3 Beamforming scenario

In the beamforming scenario (see fig. 2.2), the antennas are closely spaced apart (on the order of $\lambda/2$) and the angular spread of each path at both the transmitter and receiver side is smaller than the array resolutions [27]. Accordingly, the it path ($i = 1, .., d$) is spatially characterized by a AOD $\alpha_{i,\ell}^{(T)}$ and a AOA $\alpha_{i,\ell}^{(R)}$. Notice that each path, say the ith, is actually made of multiple micropaths (see for instance the standard model [7]), where all micropaths have AOD's and AOA's bounded within the array resolutions about the values $\alpha_{i,\ell}^{(T)}$ and $\alpha_{i,\ell}^{(R)}$ respectively. The constructive and destructive combination of different micropaths gives rise to the fast fading fluctuactions accounted for by the complex random amplitude $\beta_{i,\ell}$. Therefore, the matrix of array gains for the ith path $\mathbf{A}_{i,\ell}$ (2.5) reads

$$\mathbf{A}_{i,\ell} = \beta_{i,\ell}\mathbf{a}_R(\alpha_{i,\ell}^{(R)})\mathbf{a}_T(\alpha_{i,\ell}^{(T)})^T \tag{2.11}$$

where $\mathbf{a}_T(\alpha)$ (or $\mathbf{a}_R(\alpha)$) is the $N_T \times 1$ (or $N_R \times 1$) vector containing the array response to a plane wave transmitted (or received) with the angle α. The array response depends on the array

geometry, e.g., if the transmitter side is a uniform linear array with antenna spacing Δ, we get (λ is the wavelength of the radiated wave)

$$\mathbf{a}_T(\alpha) = [1 \ \exp(-j\frac{2\pi}{\lambda}\sin\alpha) \cdots \exp(-j\frac{2\pi}{\lambda}(N_T-1)\sin\alpha)]^T.$$

Notice that the two arrays can generally have different geometries and number of elements.

2.3.1 Long-term channel parameters

So far, the time-variability of all the channel parameters has been denoted by using the subscript ℓ that runs over the blocks. However, the quantities in parametrization (2.5), (2.11) vary over different time-scales (recall fig. 1.2). In fact, the geometry and the characteristics of the scatterers are known to vary slowly as compared to the coherence time of the fading processes. As a consequence, delays $\boldsymbol{\tau}_\ell$, AOAs $\boldsymbol{\alpha}_\ell^{(R)} = [\alpha_{1,\ell}^{(R)} \cdots \alpha_{d,\ell}^{(R)}]^T$, AODs $\boldsymbol{\alpha}_\ell^{(T)} = [\alpha_{1,\ell}^{(T)} \cdots \alpha_{d,\ell}^{(T)}]^T$, powers $\boldsymbol{\Omega}_\ell$ can be considered as long-term channel parameters with respect to the fast fading fluctuations. For the sake of the analysis, parameters $\boldsymbol{\tau}_\ell, \boldsymbol{\alpha}_\ell^{(R)}, \boldsymbol{\alpha}_\ell^{(T)}, \boldsymbol{\Omega}_\ell$ can be assumed to be constant over some interval of stationarity spanning multiple blocks. That is, it can be written:

$$\boldsymbol{\alpha}_\ell^{(T)} = \boldsymbol{\alpha}^{(T)}, \ \boldsymbol{\alpha}_\ell^{(R)} = \boldsymbol{\alpha}^{(R)}, \ \boldsymbol{\tau}_\ell = \boldsymbol{\tau}, \boldsymbol{\Omega}_\ell = \boldsymbol{\Omega} \tag{2.12}$$

$$\ell \in \{\text{interval of stationarity of long-terms parameters}\}.$$

The quasi-stationary model of temporal variations for the long-term parameters (2.12) is a first order approximation of the actual continuous (but slow) fluctuations over successive blocks. This assumption is only made so as to ease the presentation and will not be considered for design or performance evaluations throughout the thesis.

2.3.2 Fast-varying fading amplitudes

The $d \times 1$ fading amplitudes

$$\boldsymbol{\beta}_\ell = [\beta_{1,\ell} \cdots \beta_{d,\ell}]^T \tag{2.13}$$

in (2.11) model the fast fading fluctuations for the d paths. The main result presented in Sec. 2.3.3 does not depend on the specific statistic of the fading vector. However, this Section is devoted to the discussion of the main assumption that will be used in this regard throughout the following Chapters. For future reference, the number of fading amplitude is denoted by N_F, that for the beamforming scenario equals the number of paths: $N_F = d$.

Throughout the thesis, a Rayleigh probability density function for the envelope and a uniform phase distribution is assumed for each amplitude $\beta_{i,\ell}$ (*Rayleigh fading* [1]). Moreover, amplitudes corresponding to different paths are assumed to be statistically independent (*uncorrelated scattering* [15]). The temporal fluctuations of each amplitude $\beta_{i,\ell}$ are characterized statistically by means of the temporal correlation (over the blocks)

$$\varphi_i(\ell, m) = E\left[\beta_{d,\ell}\beta_{d,\ell-m}^H\right]. \tag{2.14}$$

From (2.14),in case of *packet* (or discontinuous) *transmission*, where the time interval between transmission of successive blocks is random, the non-uniform sampling makes the sampled fading process non-stationary. However, in case of *continuous transmission*, the assumption of *wide-sense sense stationarity* [15] can be assumed to hold leading to

$$\varphi_i(m) = E\left[\beta_{i,\ell}\beta_{i,\ell-m}^H\right]. \tag{2.15}$$

The Fourier transform of the correlation (2.15) defines the Doppler spectrum

$$S_{\varphi,i}(\omega) = \mathcal{F}\left\{\varphi_i(m)\right\}. \tag{2.16}$$

To simplify, the width of the Doppler spectrum (i.e., the Doppler spread) accounts for the amount of scattering each path goes through. Notice that different paths may have different Doppler spectra. For instance, in a typical outdoor scenario, a path with a larger delay corresponds to a larger Doppler spread since it is likely to have experienced more relevant scattering [16].

Example 1 *(Clarke's model): A model that has been widely employed for analysis and system simulation is the Clarke's model [2], that reads*

$$\varphi_i(m) = J_0(2\pi f_{D,i} m T_S), \tag{2.17}$$

$f_{D,i}$ *being the Doppler spread $[Hz]$ for the ith path and T_S the time interval between transmission of two successive blocks. The corresponding Doppler spectrum is*

$$S_{\varphi,i}(\omega) = \frac{1}{f_{D,i}\sqrt{1 - (\omega/(2\pi f_{D,i}))^2}}. \tag{2.18}$$

The assumptions made above can be summarized as follows. The fading amplitude vector $\boldsymbol{\beta}_\ell$ is a zero-mean circularly symmetric Gaussian process with correlation matrix

$$\boldsymbol{\Phi}(\ell,m) = E\left[\boldsymbol{\beta}_\ell\boldsymbol{\beta}_{\ell-m}^H\right] = \text{diag}\left([\varphi_1(\ell,m)\cdots\varphi_d(\ell,m)]\right), \tag{2.19}$$

that in case of continuous transmission becomes

$$\boldsymbol{\Phi}(m) = E\left[\boldsymbol{\beta}_\ell\boldsymbol{\beta}_{\ell-m}^H\right] = \text{diag}\left([\varphi_1(m)\cdots\varphi_d(m)]\right). \tag{2.20}$$

2.3.3 Decoupling the long-term parameters from the fast fading amplitudes

In this Section, the author evaluates the algebraic structure of the channel model (2.7) as it results from the assumption about the different rates of variation for the channel parameters reviewed in the previous Sections. Recalling the definition (2.11) and the properties of the Kronecker product[1], the matrix $\boldsymbol{\mathcal{A}}_\ell$ (2.9) can be factorized into the product of the $N_R N_T \times W$ stationary term $\boldsymbol{\mathcal{A}}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})$, that depends on the angles $\boldsymbol{\alpha}^{(T)}$ and $\boldsymbol{\alpha}^{(R)}$

$$\boldsymbol{\mathcal{A}}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) = [\mathbf{a}_T(\alpha_1^{(T)}) \otimes \mathbf{a}_R(\alpha_1^{(R)}), ..., \mathbf{a}_T(\alpha_d^{(T)}) \otimes \mathbf{a}_R(\alpha_d^{(R)})] \tag{2.21}$$

---

[1]Here the following relationship ($\mathbf{A}, \mathbf{B}, \mathbf{C}$ are matrices with appropriate dimensions) is employed: vec($\mathbf{ABC}$)= $(\mathbf{C}^T \otimes \mathbf{A})$ vec($\mathbf{B}$) .

and the $d \times 1$ time-varying random vector $\boldsymbol{\beta}_\ell$ as in

$$\boldsymbol{\mathcal{A}}_\ell = \boldsymbol{\mathcal{A}}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) \cdot \mathrm{diag}\{\boldsymbol{\beta}_\ell\}. \tag{2.22}$$

Notice that the stationary matrix (2.21) gathers by column the *spatial signatures* of the d paths $\mathbf{s}_{S,i} = \mathbf{a}_T(\alpha_i^{(T)}) \otimes \mathbf{a}_R(\alpha_i^{(R)})$, $i = 1, ..., d$. As explained in Sec. 2.3.1 equation (2.21) holds (as a first order approximation) within the interval of stationarity of the long-term channel parameters.

The analytical channel model (2.7) now reads

$$\mathbf{H}_\ell = \boldsymbol{\mathcal{A}}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) \cdot \mathrm{diag}\{\boldsymbol{\beta}_\ell\} \cdot \boldsymbol{\Omega}^{1/2} \cdot \mathbf{G}(\boldsymbol{\tau})^T. \tag{2.23}$$

In order to clearly decouple the long-term parameters from the fast fading, equation (2.23) is further rearranged into the $N_R N_T W \times 1$ channel vector

$$\mathbf{h}_\ell = \mathrm{vec}\{\mathbf{H}_\ell\} \tag{2.24}$$

obtaining the key result[2]

$$\mathbf{h}_\ell = \mathbf{T}\boldsymbol{\beta}_\ell,$$

where the stationary $N_R N_T W \times d$ matrix $\mathbf{T}$ contains by column the *space-time signatures* of each path $\mathbf{s}_{ST,i} = \Omega_i^{1/2} \mathbf{g}(\tau_i) \otimes \mathbf{s}_{S,i}$:

$$\begin{aligned}
\mathbf{T} &= (\mathbf{G}(\boldsymbol{\tau}) \diamond \boldsymbol{\mathcal{A}}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})) \cdot \boldsymbol{\Omega}^{1/2} = \\
&= [\Omega_1^{1/2} \mathbf{g}(\tau_1) \otimes \mathbf{a}_T(\alpha_1^{(T)}) \otimes \mathbf{a}_R(\alpha_1^{(R)}) \cdots \Omega_d^{1/2} \mathbf{g}(\tau_d) \otimes \mathbf{a}_T(\alpha_d^{(T)}) \otimes \mathbf{a}_R(\alpha_d^{(R)})]
\end{aligned} \tag{2.25}$$

2.4 Diversity scenario

In case the elements of the antenna arrays are sufficiently far apart and/or the propagation occurs in a rich scattering environment with large angular spread, the model (2.11) is not appropriate. To simplify and referring to fig. 2.2, let p paths be grouped into d subsets (or clusters) of $p_1, p_2, ..., p_d$ paths each such that $p = \sum_{i=1}^d p_i$. Each path within the ith cluster is characterized by the same delay $\tau_{i,\ell}$ but different amplitudes with overall power $\Omega_{i,\ell}$. Hence the matrix of array gains for the ith path $\mathbf{A}_{i,\ell}$ (2.5) is $\mathbf{A}_{i,\ell} = \sum_{\ell=1}^{p_i} \mathbf{a}_{i,\ell}[\ell]\mathbf{b}_{i,\ell}^T[\ell]$, where $\mathbf{a}_{i,\ell}[\ell]$ and $\mathbf{b}_{i,\ell}[\ell]$ contain the fading gains between different receiving and transmitting antennas, respectively. Here, we consider p_i very large ($p_i \to \infty$) and assume that scattering from different paths is uncorrelated so that, according to the central limit theorem,

$$\mathrm{vec}\{\mathbf{A}_{i,\ell}\} \sim \mathcal{CN}(0, \mathbf{R}_{i,\ell}) \tag{2.26}$$

where the $N_T N_R \times N_T N_R$ matrix $\mathbf{R}_{i,\ell}$ accounts for the spatial correlation of fading among the transmitting and receiving antennas. This is normalized so that along its main diagonal

---

[2]The following propertiy of the Khatri-Rao product (or columnwise Kronecker product) is used: $\mathrm{vec}\{\mathbf{A}\,\mathrm{diag}(\mathbf{b})\mathbf{C}^T\} = (\mathbf{C} \diamond \mathbf{A})\mathbf{b}$.

$[\mathbf{R}_{i,\ell}]_{mm} = 1$. The latter assumption implies that on average all the channels between each transmitting and receiving antenna are characterized by the same power gain. This exludes from our analysis macrodiversity systems in which different spatial channels may experience different shadowing and path loss [18].

In order to simplify the analysis (see, e.g., [11]) or to have a simple model to match the measurements [17], the spatial fading correlation $\mathbf{R}_{i,\ell}$ has been often assumed in the literature to be separable into a $N_T \times N_T$ spatial correlation at the transmitter side $\mathbf{R}_{i,\ell}^{(T)}$ and a $N_R \times N_R$ spatial correlation at the receiver side $\mathbf{R}_{i,\ell}^{(R)}$ as $\mathbf{R}_{i,\ell} = \mathbf{R}_{i,\ell}^{(T)} \otimes \mathbf{R}_{i,\ell}^{(R)}$. For a detailed discussion on the limits of the separable model for the spatial correlation see [19].

2.4.1 Long-term channel parameters

Similarly to Sec. 2.3.1, to which we refer for details, delays $\boldsymbol{\tau}_\ell$, powers $\boldsymbol{\Omega}_\ell$ and spatial correlation matrices $\mathbf{R}_{i,\ell}$ (that depend on the geometry of the propagation environment [19]) can be considered as long-term channel parameters with respect to the fast fading fluctuations. Hence, for the sake of the analysis, parameters $\boldsymbol{\tau}_\ell$, $\boldsymbol{\Omega}_\ell$ and $\mathbf{R}_{i,\ell}$ can be assumed to be constant over some interval of stationarity spanning multiple blocks. That is, we can write

$$\boldsymbol{\tau}_\ell = \boldsymbol{\tau}, \boldsymbol{\Omega}_\ell = \boldsymbol{\Omega}, \ \mathbf{R}_{i,\ell} = \mathbf{R}_i \tag{2.27}$$
$$\ell \in \{\text{interval of stationarity of long-term parameters}\}.$$

2.4.2 Fast-varying fading amplitudes

In the diversity scenario, each path is spatially characterized by $N_T N_R$ random fading amplitudes, modelling the gains between each transmitting and receiving antenna (see (2.26)). Therefore for each path, we define the $N_T N_R \times 1$ vector fading vector

$$\boldsymbol{\beta}_{i,\ell} = \text{vec}\{\mathbf{A}_{i,\ell}\} \tag{2.28}$$

and the overall $N_F \times 1$ fading vector (the number of fading amplitudes is $N_F = N_R N_T d$)

$$\boldsymbol{\beta}_\ell = [\beta_{1,\ell} \cdots \beta_{d,\ell}]^T. \tag{2.29}$$

As previously stated, the results presented in Sec. 2.4.3 does not depend on the specific statistics of the fading vector. However, the rest of this Section is devoted to the discussion of the main assumption that will be used in this regards throughout the thesis. Similarly to Sec. 2.3.2, Rayleigh fading with uncorrelated scattering is considered. Moreover, a space-time fading correlation that is separable in a spatial and a temporal term is assumed (see, e.g., [13])

$$E[\boldsymbol{\beta}_{i,\ell}\boldsymbol{\beta}_{i,\ell-m}^H] = \mathbf{R}_i \varphi_i(\ell, m). \tag{2.30}$$

Therefore, according to Sec. 2.3.2, the fading amplitude vector $\boldsymbol{\beta}_\ell$ is a zero-mean circularly symmetric Gaussian process with correlation matrix

$$\boldsymbol{\Phi}(\ell,m) = E\left[\boldsymbol{\beta}_\ell\boldsymbol{\beta}_{\ell-m}^H\right] = \begin{bmatrix} \mathbf{R}_1\varphi_1(\ell,m) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2\varphi_2(\ell,m) & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_d\varphi_d(\ell,m) \end{bmatrix}, \quad (2.31)$$

that in case of continuous transmission becomes

$$\boldsymbol{\Phi}(m) = E\left[\boldsymbol{\beta}_\ell\boldsymbol{\beta}_{\ell-m}^H\right] = \begin{bmatrix} \mathbf{R}_1\varphi_1(m) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2\varphi_2(m) & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_d\varphi_d(m) \end{bmatrix}. \quad (2.32)$$

2.4.3 Decoupling the long-term parameters from the fast fading amplitudes

In this Section, the author evaluates the algebraic structure of the channel model (2.7) as it results from the assumption about the different rate of variations of channel parameters for the diversity scenario. Recalling (2.26) and the definition (2.29), we have that matrix (2.9) satisfies

$$\text{vec}\{\boldsymbol{\mathcal{A}}_\ell\} = \boldsymbol{\beta}_\ell.$$

Similarly to Sec. 2.3.3, long-term parameters and fast fading amplitudes can be analytically decoupled by rearranging (2.7) into the $N_R N_T W \times 1$ vector $\mathbf{h}_\ell = \text{vec}\{\mathbf{H}_\ell\}$ obtaining again the key result

$$\mathbf{h}_\ell = \mathbf{T}\boldsymbol{\beta}_\ell, \quad (2.33)$$

where

$$\mathbf{T} = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\Omega}^{1/2} \otimes \mathbf{I}_{N_T N_R}. \quad (2.34)$$

The columns of matrix $\mathbf{G}(\boldsymbol{\tau})\boldsymbol{\Omega}^{1/2}$ in (2.34) are referred to as the *temporal signatures* of the d paths, $\mathbf{s}_{T,i} = \Omega_i^{1/2}\mathbf{g}(\tau_i)\ i = 1,...,d$.

2.5 Channel vector normalization

Throughout the thesis, for a single user link (see next Chapter for generalization to the multiuser case), the power delay profile $\boldsymbol{\Omega}$ is scaled so that the average (over the fading process) channel norm for each transmitting and receiving antenna is normalized to one:

$$E\left[\left|\sum_{m=0}^{W-1} h_\ell^{(n_T,n_R)}[m]\right|^2\right] = 1. \quad (2.35)$$

This implies that the channel vector norm reads (see next Chapter for a discussion on the definition of signal to noise ratio):

$$E[||\mathbf{h}_\ell||^2] = N_T N_R. \tag{2.36}$$

Notice that for symbol-spaced multipath delays, i.e., τ_i/T integer for $i = 1, ..., d$, we have from the assumption of uncorrelated scattering

$$E\left[\left|\sum_{m=0}^{W-1} h_\ell^{(n_T, n_R)}[m]\right|^2\right] = \sum_{m=0}^{W-1} E[|h_\ell^{(n_T, n_R)}[m]|^2] = 1, \tag{2.37}$$

that is satisfied if and only if $\sum_{i=1}^{d} \Omega_i = 1$.

Table 2.1: Summary of long-term features and number of fast varying fading amplitudes N_F for the beamforming and diversity scenarios.

Scenario	Long-terms parameters	N_F
Beamforming (outdoor)	AOA $\alpha_i^{(R)}$, AOD $\alpha_i^{(T)}$, delay τ_i, power-delay profile Ω_i, temporal fading correlation $\varphi_i(m)$	d
Diversity (indoor)	delay τ_i, power-delay profile Ω_i, space-time fading correlation $\mathbf{R}_i \varphi_i(m)$	$N_R N_T d$

2.6 Conclusion

In this chapter, an analytical framework for the definition of MIMO channel models that encompasses multipath propagation with non-integer delays, space-time fading correlation, angles of departure/arrival and arbitrary power-delay profile has been proposed. In particular, two scenarios have been considered: the beamforming and diversity scenarios, suited for outdoor and indoor applications respectively.

The study of the algebraic properties of the time-varying channel model showed that the contribution of long-term parameters, summarized in Table 2.1, and the fast-varying fading amplitudes can be analytically decoupled according to the matrix product $\mathbf{h}_\ell = \mathbf{T}\boldsymbol{\beta}_\ell$, where matrix $\mathbf{T}$ accounts for the long-term parameters and the $N_F \times 1$ vector $\boldsymbol{\beta}_\ell$ collects the fading amplitudes.

Remark 1 *Mixed beamforming/diversity scenarios for MIMO links that present asymmetric geometry of arrays and/or scatters at their ends (see, e.g., [21]) can be derived by generalizing the discussion above. This aspect will not be covered in this thesis.*

Bibliography

[1] G. L. Stuber, *Principles of Mobile Communication*, Kluwer Academic Publishers, 2001.

[2] W. C. Jakes, *Microwave Mobile Communications,* John Wiley & Sons Inc, 1975

[3] A. Van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint Angle and Delay Estimation Using Shift-Invariance Techniques", *IEEE Trans. Signal Processing*, vol 46, pp 405-418, Feb. 1998.

[4] Yung-Yi Wang, Jiunn-Tsair Chen and Wen-Hsien Fang, "TST-MUSIC for Joint AOA-Delay Estimation," *IEEE Trans. Signal Processing*, vol. 49, pp. 721-729, April 2001.

[5] A. Richter, D. Hampicke, G. Sommerkorn and R. Thoma, "MIMO measuremement and joint parameter estimation of mobile radio channels," in *Proc. VTC 2001*, vol. 1, pp. 214-218, 2001.

[6] P. Kyritsi, D. C. Cox, "Correlation properties of MIMO radio channels for indoor scenarios," in *Proc. Asilomar Conf. on Signals, Systems and Computers 2001*, pp. 994-998, 2001.

[7] 3GPP TR 25.996 V.6.0,"Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations (Release 6)", *3GPP Technical Specification Group Radio Access*, June 2003.

[8] H. Holma and A. Toskala, *WCDMA for UMTS,* John Wiley & Sons, 2000.

[9] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357-366, March 1998.

[10] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing,* vol. 46, no. 2, pp. 449-462, Feb. 1998.

[11] D. Shiu, G. J. Foschini and M. J. Gans, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502 - 513, March 2000.

[12] H. L. Van Trees, *Optimum array processing*, J. Wiley Ed., 2002.

[13] C. Komninakis, C. Fragouli, A. H. Sayed and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing,* vol. 50, no. 10, pp. 1065-1076, May 2002.

[14] Z. Liu, Y. Xin and G. B. Giannakis, "Space-time-frequency coded OFDM over frequency-selective fading channels," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 2465-2476, Oct. 2002.

[15] P. A. Bello, "Characterization of randomly time-variant linear channel", *IEEE Trans. Circuits and Systems*, vol. 11, no. 4, pp.360-393, Dec. 1963.

[16] R. Janaswamy, *Radiowave Propagation and Smart Antennas for Wireless Communications*, Kluwer Academic Publishers, 2000.

[17] K. I. Pedersen, J. B. Andersen, J. P. Kermoal and P. Mogensen, "A stochastic multiple-input-multiple-output radio channel model for evaluation of space-time coding algorithms," in *Proc. IEEE VTC*, Boston, pp. 893-897, 2000.

[18] S. Mukherjee and D. Avidor, "Effect of microdiversity and correlated macrodiversity on outages in a cellular system," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 50-58, Jan. 2003.

[19] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels", *IEEE J. Select. Areas Commun.*, vol. 20, no. 3, pp. 550-560, April 2002.

[20] T. L. Marzetta, "Blast training: estimating channel characteristics for high capacity space-time wireless," *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 958–966, September 1999.

[21] H. Bolckei, D. Gesbert, A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Comm.*, vol. 50, pp. 225-234, Feb. 2002.

[22] M. Patzold, *Mobile fading channels*, John Wiley & Sons, 2002.

Chapter 3

Signal model for MIMO systems over frequency-selective channels

3.1 Introduction

THIS Chapter is devoted to a brief review of the signal model used for the analysis of time and frequency (i.e., multicarrier) domain transmission and reception over single and multiuser MIMO links. As explained in 1, this thesis is concerned with systems that perform CSI acquisition through transmission of training sequences within each radiated block. This block can be obtained either according to time-domain or multicarrier transmission, as detailed in the following. In either case, it contains both training (or pilot) symbols, i.e., symbols known to the receiver, that are designed for channel estimation, and data symbols, that carry the useful information. Throughout the thesis, the number of training/pilot symbols and the number of data symbols are held fixed for the whole duration of the communication, i.e., for all blocks. However, the transmitter could adapt these parameters on a block-by-block basis according to the CSI available and the quality of service requirements coming from higher layers. This approach is proposed and investigated in Chapter 10.

The discrete-time signal received by the n_Rth receiving antenna at time instant m is the superposition of the signal radiated by the N_T transmitting antennas $\{x_\ell^{(n_T)}[m]\}_{n_T=1}^{N_T}$ after convolution with the channel impulse response $h_\ell^{(n_R,n_T)}[m]$

$$y_\ell^{(n_R)}[m] = \sum_{n_T=1}^{N_T} x_\ell^{(n_T)}[m] * h_\ell^{(n_R,n_T)}[m] + n_\ell^{(n_R)}[m], \qquad (3.1)$$

where $n_\ell^{(n_R)}[m]$ is the additive Gaussian noise, to be specified below. It is recall that subscript ℓ runs over the transmitted blocks. In this Chapter, it is shown that by rearranging the transmitted, received signal and the channel gains into appropriately defined matrices, the signal model (3.1) can be conveniently stated as anticipated below.

For the signal received over the training part of the block, it is useful to make explicit the

dependence on the channel vector $\mathbf{h}_\ell$ (2.24). Therefore, after the manipulations described in Sec. 3.2.1, the received signal $\mathbf{y}_\ell$ can be written as

$$\mathbf{y}_\ell = \mathbf{X}_\ell \mathbf{h}_\ell + \mathbf{n}_\ell, \qquad \text{(training or pilot phase)} \qquad (3.2)$$

where $\mathbf{X}_\ell$ is a block-convolution matrix built from the training sequences. As shown in Sec. 3.3.1, this model holds also for multicarrier transmission ($\mathbf{y}_\ell$ herein is to be intended as the signal received over different pilot subcarriers) but in this case $\mathbf{X}_\ell$ is a block-diagonal matrix, accounting for the absence of interference among the subcarriers.

On the other hand, it is convenient to write the signal received over the data part of the block so as to emphasize the dependence on the data vector $\mathbf{x}_\ell$. In particular, according to the discussion in Sec. 3.2.2, the received signal can be stated as

$$\mathbf{y}_\ell = \boldsymbol{\mathcal{H}}_\ell \mathbf{x}_\ell + \mathbf{n}_\ell, \qquad \text{(data phase)} \qquad (3.3)$$

where $\boldsymbol{\mathcal{H}}_\ell$ is a block-convolution matrix obtained from the channel gains. As proved in Sec. 3.3.2, this signal model holds for multicarrier transmission as well ($\mathbf{y}_\ell$ herein is to be intended as the signal received over different data subcarriers) but in this case $\boldsymbol{\mathcal{H}}_\ell$ is block-diagonal.

Notice that in (3.2)-(3.3) we are using the same notation for both time and multicarrier transmission in order to clarify the duality of the two approaches. Correct definition of the quantities involved is provided in the rest of this Chapter. Moreover, in Sec. 3.4 it will be shown how to extend the signal model to a multiuser system.

Remark 2 (*joint training and data processing*) *In this thesis we coinsider a separate processing of the training and data symbols. However, joint processing through, e.g., iterative techniques, has been recently advocated (see, e.g., [1]). This solution, though it complicates the structure of transmitter and receiver, holds the promise of achieving better spectral efficiency.*

3.2   Time-domain transmission/reception

In case of transmission in the time domain, the signal radiated by each antenna is organized in blocks containing a training sequence time-multiplexed with a data burst as in fig. 3.1. Notice that the arrangement shown in the figure is only for illustration purposes since in principle data field(s) and training sequence could be ordered in different ways, e.g., in the TDD-UMTS standard the training sequence is placed between two data bursts (midamble [2]). The shaded regions in fig. 3.1 correspond to the redundancy (e.g., zero padding or cyclic prefix [3]) needed at the beginning of each block in order to cancel the inter-block interference due to the temporal spread of the channel (W samples, see previous Chapter).

Within the training (or data) part of the ℓth block, the n_Tth transmitting antenna sends a training (or data) sequence $x_\ell^{(n_T)}[m]$ at symbol-rate $1/T$. For convenience, the same symbol
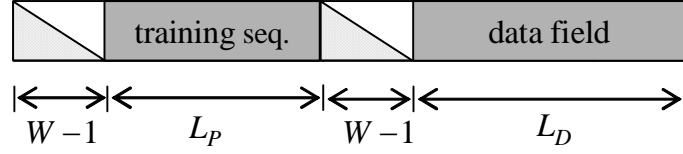
Figure 3.1: Time-slot for time-domain transmission.

denotes either the training and data sequences, the use will be made clear by the context. Training sequences are selected so as to ease the channel estimation process, as explained in Chapter 4. On the other hand, the data sequences are output by a MIMO encoder/modulator, designed with the general goal of minimizing the probability of error for detection of the input binary data b_m. The block diagram of the system for training and data transmission phases is shown in fig. 3.2.

Let L_P denote the length of the training (also referred to as pilot) sequences and L_D the number of samples corresponding to the data field. From fig. 3.1, the amount of redundancy needed in order to estimate the channel and ensure inter-block interference free transmission is $L_O = L_P + 2(W - 1)$. For efficient use of spectral resources, it is thus desirable that $L_D >> L_O$ under the condition that the fading coherence time is larger (at least equal) than the block duration $(L_D + L_O)T$. Notice that the the guard period right after the training block could be avoided by cancelling the estimated interference from the training symbols toward the data block using the estimated channel.

The discrete-time signal received by the n_Rth receiving antenna can be written as in (3.1). By rearranging the signal received over the N_R receiving antennas in the $N_R \times 1$ vector $\mathbf{y}_\ell[m] = [y_\ell^{(1)}[m] \cdots y_\ell^{(N_r)}[m]]^T$ we easily get

$$\mathbf{y}_\ell[m] = \sum_{i=0}^{W-1} \mathbf{H}_\ell[i]\mathbf{x}_\ell[m-i] + \mathbf{n}_\ell[m], \tag{3.4}$$

where $\mathbf{x}_\ell[i] = [x_\ell^{(1)}[i] \cdots x_\ell^{(N_T)}[i]]^T$ is $N_T \times 1$ and $\mathbf{H}_\ell[m]$ is the mth $N_R \times N_T$ MIMO tap of the channel impulse response defined in the previous Chapter ($[\mathbf{H}_k[m]]_{n_R,n_T} = h_k^{(n_R,n_T)}[m]$ with $m = 0, ..., W - 1$). The discrete-time AGN $\mathbf{n}_k[\ell]$ is assumed temporally uncorrelated but spatially correlated with covariance matrix $\mathbf{R}_n$:

$$E[\mathbf{n}_\ell[m]\mathbf{n}_\ell[m-i]^H] = \mathbf{R}_n\delta[i]. \tag{3.5}$$

In general, $\mathbf{R}_n$ is modelled as the sum of a diagonal matrix accounting for the thermal noise, that is always present in the receiving equipment, and a (positive definite) non-diagonal matrix that accounts for spatially distributed interfering signals [5]. For instance, in a cellular system, the latter term can model out-of-cell interferers. The gaussian assumption made in this regard can be justified, under appropriate conditions, according to the central limit theorem [6].

Training phase:
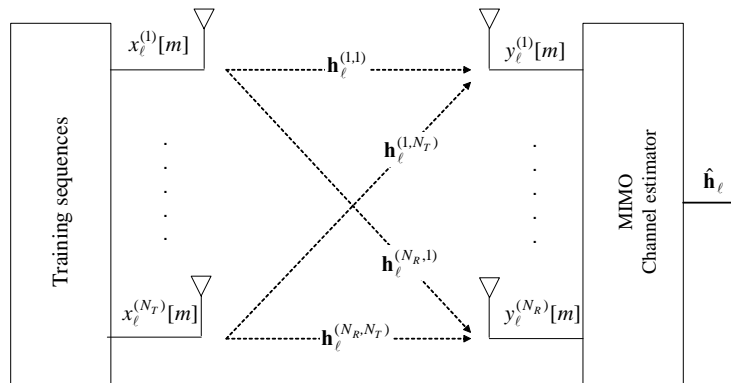
Data phase:
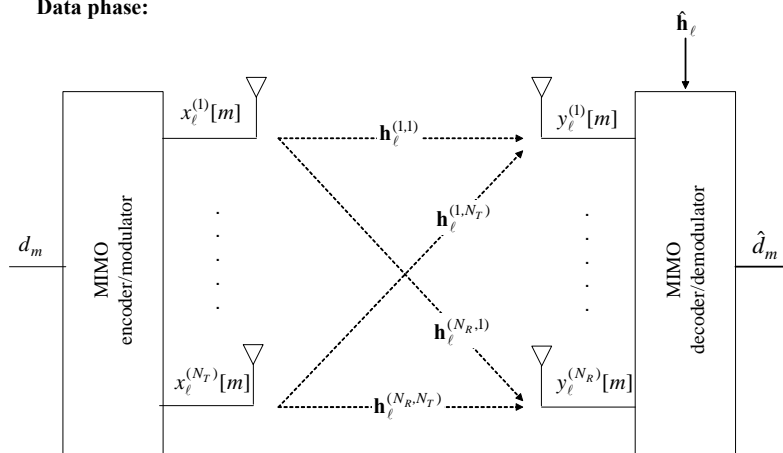
Figure 3.2: Block diagram for the training and data phase of a time-domain transmission system.

Discarding the preamble of $W - 1$ samples, the L_α (L_α stands for L_P for the pilot and L_D for the data sequence respectively) received samples can be arranged into the $N_R L_\alpha \times 1$ vector $\mathbf{y}_\ell = [\mathbf{y}_\ell[0]^T ... \mathbf{y}_\ell[L_\alpha - 1]^T]^T$. In this thesis, according to the principle of synchronized detection [4], training and data sequences are processed separately. For analysis, it is convenient to write the received vector $\mathbf{y}_\ell$ in different forms for the training and data phases.

3.2.1 Training phase

When studying the signal received over the training part of the block for the purposes of designing or analyzing the channel estimation process, it is convenient to write the received signal so as to make explicit the dependence on the channel vector $\mathbf{h}_\ell$ (2.24). Toward this goal, it is useful to first define the $N_R \times L_P$ matrix $\mathbf{Y}_\ell = [\mathbf{y}_\ell[0] \cdots \mathbf{y}_\ell[L_P - 1]]$, that can be written as follows

$$\mathbf{Y}_\ell = \check{\mathbf{H}}_\ell \boldsymbol{\mathcal{X}}_\ell^T + \mathbf{N}_\ell, \tag{3.6}$$

where

$$\check{\mathbf{H}}_\ell = [\mathbf{H}_\ell[0], \mathbf{H}_\ell[1], ..., \mathbf{H}_\ell[W - 1]] \tag{3.7}$$

is the $N_R \times N_T W$ MIMO-FIR channel matrix that gathers the MIMO channel taps; $\boldsymbol{\mathcal{X}}_\ell$ is the $L_P \times N_T W$ convolution matrix obtained from the N_T training sequences

$$\boldsymbol{\mathcal{X}}_\ell = \begin{bmatrix} \mathbf{x}_\ell^T[0] & \mathbf{x}_\ell^T[-1] & & \mathbf{x}_\ell^T[-W + 1] \\ \mathbf{x}_\ell^T[1] & \mathbf{x}_\ell^T[0] & & \mathbf{x}_\ell^T[-W + 2] \\ \vdots & \ddots & \cdots & \vdots \\ \mathbf{x}_\ell^T[L_P] & \mathbf{x}_\ell^T[L_P - 1] & & \mathbf{x}_\ell^T[L_P - W + 1] \end{bmatrix}; \tag{3.8}$$

$\mathbf{N}_\ell$ has the same structure as $\mathbf{Y}_\ell$ and from (3.5) $1/L_P \cdot E[\mathbf{N}_\ell \mathbf{N}_\ell^H] = \mathbf{R}_n$. Notice that in (3.8), the training symbols $\mathbf{x}_\ell[-m]$ with $m > 0$ correspond to the $W - 1$ samples transmitted from each antenna in the guard period (see fig. 3.1). In case a cyclic prefix is employed, $\mathbf{x}_\ell[-m] = \mathbf{x}_\ell[L_P - m]$ ($m > 0$) whereas for zero padding it is $\mathbf{x}_\ell[-m] = 0$. We remark that, as it will be clarified in the next Chapter, since each receiving antenna has to estimate $N_T W$ channel coefficients (i.e., the n_Rth receiving antenna needs to acquire $h_k^{(n_R, n_T)}[m]$ with $n_T = 1, ..., N_T$ and $m = 0, ..., W - 1$), the length of the training sequences L_P is required to satisfy $L_P \geq N_T W$.

Then, by stacking the L_P columns of the matrix $\mathbf{Y}_\ell$ into the $N_R L_P \times 1$ vector $\mathbf{y}_\ell = \text{vec}\{\mathbf{Y}_k\}$ we get

$$\mathbf{y}_\ell = (\boldsymbol{\mathcal{X}}_\ell \otimes \mathbf{I}_{N_R}) \mathbf{h}_\ell + \mathbf{n}_\ell = \mathbf{X}_\ell \mathbf{h}_\ell + \mathbf{n}_\ell, \tag{3.9}$$

that shows the desired dependence on the channel vector $\mathbf{h}_\ell$. The spatial covariance matrix for the AGN $\mathbf{n}_\ell = \text{vec}\{\mathbf{N}_\ell\}$ is $E[\mathbf{n}_\ell \mathbf{n}_\ell^H] = \mathbf{I}_{L_P} \otimes \mathbf{R}_n$.

As studied in Chapters 4 and 5, the performance of channel estimation depends on the correlation properties of the training sequences. On the one hand, it is clear that if the training sequences sent by two antennas are highly correlated, the channel estimator at the receiver is

not be able to clearly separate the contribution of the two sequences. Therefore, when sensing the channel relative to one antenna, its task is impaired by the interference due to the signal corresponding to the other antenna. This problem is solved if the training sequences assigned to different antennas are orthogonal (or approximate this optimal condition). On the other hand, in order to guarantee a channel estimate with a good temporal resolution, the training sequence sent by each antenna has to present an narrow temporal correlation [7]. Both the spatial (i.e., among transmitting antennas) and temporal correlation of the training sequences are accounted for by the $N_T W \times N_T W$ correlation matrix

$$\mathbf{R}_x = \boldsymbol{\mathcal{X}}_\ell^H \boldsymbol{\mathcal{X}}_\ell. \tag{3.10}$$

According to the discussion above, it is expected that channel estimation attains optimal performance $\mathbf{R}_x$ is diagonal (to be discussed in the following Chapters).

A useful model to investigate the correlation properties of the training sequences is to decouple the effect of spatial and temporal correlation by assuming for $\mathbf{R}_x$ the following separable model

$$\mathbf{R}_x = \mathbf{R}_{x,T} \otimes \mathbf{R}_{x,S}, \tag{3.11}$$

where the correlation matrix $\mathbf{R}_{x,S}$ is $N_T \times N_T$ and accounts for spatial correlation while the $W \times W$ matrix $\mathbf{R}_{x,T}$ models the temporal correlation (assumed to be the same for each sequence).

3.2.2   Data phase

The signal received over the data part of the block can be conveniently expressed as a function of the $N_T L_D \times 1$ vector of transmitted data $\mathbf{x}_\ell = [\mathbf{x}_\ell[0]^T \cdots \mathbf{x}_\ell[L_D - 1]^T]^T$ so as to simplify the analysis and design of the equalization/detection process. The received signal $\mathbf{y}_\ell$ can be easily written in terms of $\mathbf{x}_\ell$ as follows

$$\mathbf{y}_\ell = \boldsymbol{\mathcal{H}}_\ell \mathbf{x}_\ell + \mathbf{n}_\ell \tag{3.12}$$

where $\boldsymbol{\mathcal{H}}_\ell$ denotes the $N_R L_D \times N_T L_D$ convolution matrix of the MIMO channel

$$\boldsymbol{\mathcal{H}}_\ell = \begin{bmatrix} \mathbf{H}_\ell[0] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{H}_\ell[0] & \mathbf{0} & \mathbf{0} \\ \mathbf{H}_\ell[W-1] & \ddots & \ddots & \cdots & \vdots \\ \mathbf{0} & \mathbf{H}_\ell[W-1] & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{H}_\ell[0] \end{bmatrix}. \tag{3.13}$$

3.2.3   Signal to noise ratio definition

Throughout the thesis the signal to noise ratio SNR (averaged over the fading, noise and data/training sequence distributions) is defined per receiving antenna, i.e., from (3.1) and re-

calling the channel normalization (2.35)

$$SNR = \frac{P}{\sigma_n^2},\qquad(3.14)$$

where P is the total power radiated by the transmitting array, $P = \sum_{n_T=1}^{N_T} E[|x_\ell^{(n_T)}[m]|^2]$, and $E[|n_\ell^{(n_R)}[m]|^2] = \sigma_n^2$. Notice that in general the transmitted power over the training and data part of the block can be different in order to optimize the system performance (see Chapter 6 and [8] [9]).

3.3 Multicarrier transmission/reception (MIMO-OFDM)

In multicarrier systems, as exemplified in figures 3.3 and 3.4, the transmitters collect a block of L (data and/or training) symbols and performs an IDFT on the resulting vector before transmission. This operation transforms the original (data and/or training) signal from the frequency domain (i.e., each symbol is transmitted on a given subcarrier) to the temporal domain. Moreover, a cyclic prefix of $W - 1$ (see also discussion in Sec. 3.2.1) is appended to the transformed block in the temporal domain. This contains the last $W - 1$ samples of the block and not only allows interference-free reception of successive blocks but also avoids inter-carrier interference. In fact, as shown in Appendix, after the cyclic prefix has been removed from the the received block, the receivers perform a DFT on the remaining L samples. This operation transforms the received signal from the temporal domain to the frequency domain (i.e., each symbol of the transformed block correspond to a given subcarrier). If the temporal support of the channel is smaller (or equal to) the length of the cyclic prefix, the signal received over each subcarrier do not interfere with each other.

As shown in fig. 3.3, each OFDM symbol contains L subcarriers (recall that L is also the length of the block in the time-domain). Out of the available L subcarriers, L_P subcarriers contain pilot (or training) symbols used for channel estimation, L_D contains data and the remaining L_G are allocated as guard bands so as to simplify the system implementation (e.g., of analog shaping filters) [10]. Notice that the arrangement shown in fig. 3.3 is only for illustration purposes since in principle data subcarriers and training sequence can be ordered in different ways, see, e.g., the multicarrier standards for wireless LAN [10] or digital video broadcasting [11] or the theoretical analysis [13] [14]. Moreover, the fraction of pilot and data symbols could be different for different blocks. From fig. 3.3, the amount of redundancy needed in order to estimate the channel, ensure inter-block and intercarrier interference free transmission and ease the implementation of the analog shaping filters is $L_O = L_P + L_G + W - 1$. For efficient use of spectral resources, it is thus desirable that $L_D >> L_O$ under the condition that the fading coherence time is larger (at least equal) than the block duration $(L_D + L_O)T$.

Remark 3 *(intercarrier interference and Doppler spread): In the previous discussion referring to the results in Appendix, we have proved that in an ideal system the introduction of a*

Time-domain:
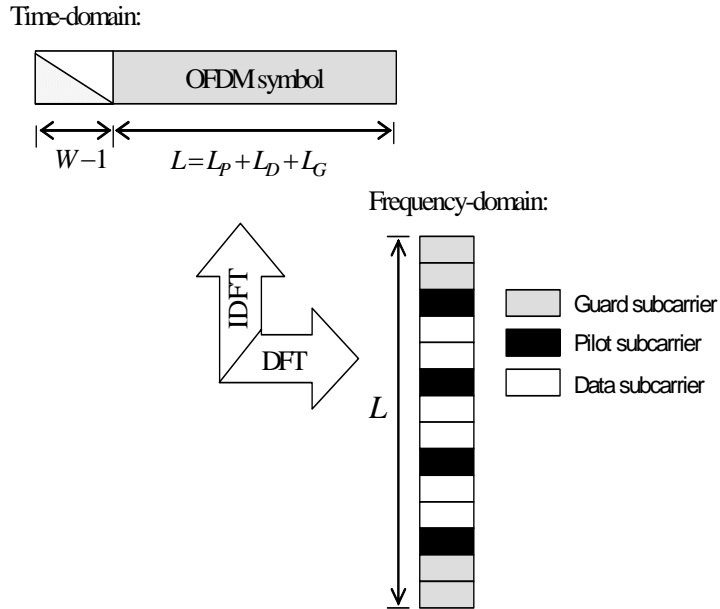
Frequency-domain:

Figure 3.3: OFDM symbol in time and frequency domain.
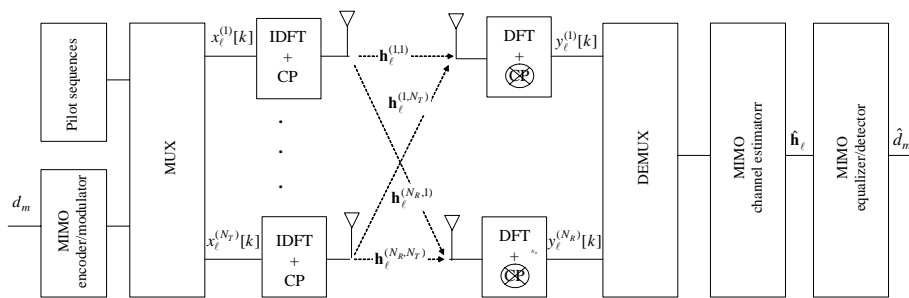
Figure 3.4: Block diagram of a frequency-domain transmission (MIMO-OFDM) system.

cyclic prefix nulls the interference among different subcarriers. However, intercarrier interference may arise because of possible residual frequency offset [15], phase noise [16] or large Doppler spread [12]. In practice, it is enough that the signal-to-intercarrier interference ratio is sufficiently larger than the operational signal to noise ratio. This condition can be tested for instance using the results in [12] and will be assumed throughout the thesis.

For the sake of simplicity, the same notation will be used in the following for dual variables in time and frequency domain (such as the received signal or the additive noise) whenever the use is clear from the context. In this way, the duality between the two approaches should be apparent by comparing dual equations.

As discussed in Appendix, the signal received by the n_Rth receiving antenna over the kth subcarrier is the superposition of the signal radiated by the N_T transmitting antennas after multiplication with the corresponding channel impulse coefficient in the frequency domain $f_\ell^{(n_R,n_T)}[k]$

$$y_\ell^{(n_R)}[k] = \sum_{n_T=1}^{N_T} x_\ell^{(n_T)}[k] f_\ell^{(n_R,n_T)}[k] + n_\ell^{(n_R)}[k], \tag{3.15}$$

where $n_\ell^{(n_R)}[m]$ is the additive Gaussian noise for the kth subcarrier, to be specified in the following. The channel coefficients in the frequency domain are obtained as the DFT of the channel in time domain:

$$f_\ell^{(n_R,n_T)}[k] = \frac{1}{\sqrt{L}} \sum_{m=0}^{W-1} h_\ell^{(n_R,n_T)}[m] \exp(-j\frac{2\pi}{L}km), \tag{3.16}$$

or equivalently letting $\mathbf{\Theta}$ be the $L \times W$ DFT matrix

$$\mathbf{\Theta} = \frac{1}{\sqrt{L}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \exp(-j\frac{2\pi}{L} \cdot 2) & \cdots & \exp(-j\frac{2\pi}{L} \cdot (W-1)) \\ \vdots & \vdots & & \vdots \\ 1 & \exp(-j\frac{2\pi}{L} \cdot (L-1)) & \cdots & \exp(-j\frac{2\pi}{L} \cdot (L-1)(W-1)) \end{bmatrix}, \tag{3.17}$$

as

$$\mathbf{f}_\ell^{(n_R,n_T)} = \mathbf{\Theta} \cdot \mathbf{h}_\ell^{(n_R,n_T)} \tag{3.18}$$

with definitions $\mathbf{f}_\ell^{(n_R,n_T)} = [f_\ell^{(n_R,n_T)}[0] \cdots f_\ell^{(n_R,n_T)}[L-1]]^T$ and $\mathbf{h}_\ell^{(n_R,n_T)} = h_\ell^{(n_R,n_T)}[0] \cdots h_\ell^{(n_R,n_T)}[W-1]]^T$.

By rearranging the signal received over the N_R receiving antennas in the $N_R \times 1$ vector $\mathbf{y}_\ell[k] = [y_\ell^{(1)}[k] \cdots y_\ell^{(N_T)}[k]]^T$ we easily get

$$\mathbf{y}_\ell[k] = \mathbf{F}_\ell[k]\mathbf{x}_\ell[k] + \mathbf{n}_\ell[k], \tag{3.19}$$

where $\mathbf{x}_\ell[k] = [x_\ell^{(1)}[k] \cdots x_\ell^{(N_T)}[k]]^T$ is $N_T \times 1$ and $\mathbf{F}_\ell[m]$ is the mth $N_R \times N_T$ MIMO channel impulse response for the kth frequency:

$$[\mathbf{F}_\ell[k]]_{n_R, n_T} = f_\ell^{(n_R, n_T)}[k]. \tag{3.20}$$

The discrete-time AGN $\mathbf{n}_\ell[k]$ is assumed temporally uncorrelated but spatially correlated with covariance matrix $\mathbf{R}_n$:

$$E[\mathbf{n}_\ell[m]\mathbf{n}_\ell[m-i]^H] = \mathbf{R}_n \delta[i]. \tag{3.21}$$

Notice that the spatial correlation of the noise in the frequency domain equals the same quantity in the time domain. Similarly to the time-domain case, in the following Sections, we cast the received vector $\mathbf{y}_\ell = [\mathbf{y}_\ell[0] \cdots \mathbf{y}_\ell[L_\alpha - 1]]^T$ (L_α stands for L_P for the pilot and L_D for the data sequence respectively) in algebraic forms suitable for analysis of the channel estimation phase (over the pilot subcarriers) and the equalization/detection phase (over the data subcarriers).

3.3.1 Pilot subcarriers

When studying the signal received over the pilot subcarriers for the purposes of designing or analyzing the channel estimation process, it is convenient to write the received signal so as to make explicit the dependence on the channel vector $\mathbf{h}_\ell$ (2.24). Here, to simplify the notation, index k in (3.19) is considered to run only over the pilot subcarriers, i.e., $k = 0, ..., L_P - 1$ indexes the kth pilot subcarrier. Now, toward the stated goal, it is useful to first define the $N_R \times L_P$ matrix $\mathbf{Y}_\ell = [\mathbf{y}_\ell[0] \cdots \mathbf{y}_\ell[L_P - 1]]$, that can be written as follows

$$\mathbf{Y}_\ell = \check{\mathbf{F}}_\ell \boldsymbol{\mathcal{X}}_\ell^T + \mathbf{N}_\ell, \tag{3.22}$$

where

$$\check{\mathbf{F}}_\ell = [\mathbf{F}_\ell[0] \ \mathbf{F}_\ell[1] ... \mathbf{F}_\ell[L_P - 1]] \tag{3.23}$$

is $N_R \times N_T L_P$; $\boldsymbol{\mathcal{X}}_\ell$ is the $L_P \times N_T W$ block-diagonal matrix obtained from the N_T pilot sequences

$$\boldsymbol{\mathcal{X}}_\ell = \begin{bmatrix} \mathbf{x}_\ell^T[0] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_\ell^T[1] & & \mathbf{0} \\ \vdots & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_\ell^T[L_P] \end{bmatrix}, \tag{3.24}$$

$\mathbf{N}_\ell$ has the same structure as $\mathbf{Y}_\ell$ and $1/L_P \cdot E[\mathbf{N}_\ell \mathbf{N}_\ell^H] = \mathbf{R}_n$.

Then, by stacking the L_P columns of the matrix $\mathbf{Y}_\ell$ into the $N_R L_P \times 1$ vector $\mathbf{y}_\ell = \text{vec}\{\mathbf{Y}_k\}$ we get

$$\mathbf{y}_\ell = (\boldsymbol{\mathcal{X}}_\ell \otimes \mathbf{I}_{N_R})\mathbf{f}_\ell + \mathbf{n}_\ell, \tag{3.25}$$

where $\mathbf{f}_\ell = \text{vec}\{\check{\mathbf{F}}_\ell\}$. But from (3.18), it easy to show that

$$\mathbf{f}_\ell = (\boldsymbol{\Theta}_P \otimes \mathbf{I}_{N_R N_T})\mathbf{h}_\ell,$$

with $\mathbf{\Theta}_P$ denoting the $L_P \times W$ DFT matrix obtained from $\mathbf{\Theta}$ by selecting the columns related to the pilot subcarriers. Finally, we get the desired result (compared with (3.9) for the time-domain):

$$\mathbf{y}_\ell = \mathbf{X}_\ell \mathbf{h}_\ell + \mathbf{n}_\ell \tag{3.26}$$

with

$$\mathbf{X}_\ell = (\boldsymbol{\mathcal{X}}_\ell^H \otimes \mathbf{I}_{N_R})(\mathbf{\Theta}_P \otimes \mathbf{I}_{N_R N_T}) = [\boldsymbol{\mathcal{X}}_\ell^H (\mathbf{\Theta}_P \otimes \mathbf{I}_{N_T})] \otimes \mathbf{I}_{N_R}. \tag{3.27}$$

The spatial covariance matrix for the AGN $\mathbf{n}_\ell = \text{vec}\{\mathbf{N}_\ell\}$ is $E[\mathbf{n}_\ell \mathbf{n}_\ell^H] = \mathbf{I}_{L_P} \otimes \mathbf{R}_n$.

As it will be clarified in the next Chapter, since each receiving antenna has to estimate $N_T W$ channel coefficients (i.e., the n_Rth receiving antenna needs to acquire $h_k^{(n_R, n_T)}[m]$ with $n_T = 1, ..., N_T$ and $m = 0, ..., W-1$), the number of the pilot subcarriers L_P is required to satisfy $L_P \geq N_T W$.

In the next Chapters, it will be shown that the performance of channel estimation depends on the correlation properties (either in time and space domain see discussion in Sec. 3.2.1) of the training sequences sent by different antennas and in particular on the $N_T W \times N_T W$ correlation matrix

$$\mathbf{R}_x = (\mathbf{\Theta}_P^H \otimes \mathbf{I}_{N_T}) \boldsymbol{\mathcal{X}}_\ell^H \boldsymbol{\mathcal{X}}_\ell (\mathbf{\Theta}_P \otimes \mathbf{I}_{N_T}). \tag{3.28}$$

Recalling the discussion in Sec. 3.2.1, we can conclude that optimum channel estimation performance can be attained if the pilot sequences are orthogonal, i.e., if the correlation matrix (3.28) is diagonal. Moreover, following the lines of Sec. 3.2.1, it is convenient to parametrize the correlation (3.28) by the separable model (3.11).

3.3.2 Data phase

The signal received over the data subcarriers of the block can be conveniently expressed as a function of the $N_T L_D \times 1$ vector of transmitted data $\mathbf{x}_\ell = [\mathbf{x}_\ell[0]^T \cdots \mathbf{x}_\ell[L_D - 1]^T]^T$ so as to simplify the analysis and design of the equalization/detection process. Notice that, similarly to the previous Section, here we are considering the index k to run over the data subcarriers, i.e., $k = 0, ..., L_D - 1$ indexes the kth data subcarrier. The received signal $\mathbf{y}_\ell$ over the pilot subcarriers can be easily written as

$$\mathbf{y}_\ell = \boldsymbol{\mathcal{H}}_\ell \mathbf{x}_\ell + \mathbf{n}_\ell \tag{3.29}$$

where $\boldsymbol{\mathcal{H}}_\ell$ denotes the $N_R L_D \times N_T L_D$ block-diagonal matrix

$$\boldsymbol{\mathcal{H}}_\ell = \begin{bmatrix} \mathbf{F}_\ell[0] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_\ell[1] & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}_\ell[L_D - 1] \end{bmatrix}. \tag{3.30}$$

3.3.3   Signal to noise ratio definition

The signal to noise ratio SNR (averaged over the fading, noise and data/training sequence distributions) defined in (3.14) for time-domain transmission corresponds for multicarrier transmission to the SNR per receiving antenna and subcarrier as it can be easily shown by using the channel normalization (2.35) and the relationship (3.46). Notice that in general the transmitted power over the pilot and data subcarrier of the OFDM symbol could be different (see Chapter 6). Moreover, as a final remark, power loading could be employed over the data subcarriers in order to maximize the spectral efficiency of the system or minimize the bit error rate [17].

3.4   Extension to multiuser system

So far the signal model has been discussed for a multiantenna link where all the antennas at both the transmitter and receiver side cooperate when processing the transmitted/received signal. This is usually referred to as single user MIMO link. Here, the signal model is extended to a multiuser system, and in particular to the uplink and downlink of a multiantenna system. The extension is straightforward since, as compared to the single user setting, it is enough to partition the antenna array at the transmitter (for uplink) or receiver (for downlink) into smaller non-cooperative arrays corresponding to different users. In other words the multiple arrays, or users, transmit (for uplink) or receive (for downlink) without performing a cooperative processing on the transmitted/received signal as it was the case for a single user MIMO system.

3.4.1   Uplink of a MIMO system (multiaccess channel)

In the uplink of a MIMO system, the N_T transmitting antennas are partitioned among K (non-cooperating) users, each having $n_T^{(k)}$ antennas ($k = 1, ..., K$), $\sum_{k=1}^{K} n_T^{(k)} = N_T$. The receiver, modelling the base station or access point of the multiaccess channel, is equipped with N_R receiving antennas.

Training phase (or pilot subcarriers)

For time-domain transmission, the signal received over the N_R antennas of the base station within the training block (3.6) becomes

$$\mathbf{Y}_\ell = \sum_{k=1}^{K} \breve{\mathbf{H}}_\ell^{(k)} \boldsymbol{\mathcal{X}}_\ell^{(k)} + \mathbf{N}_\ell, \tag{3.31}$$

where $\breve{\mathbf{H}}_\ell^{(k)}$ is the channel matrix built from the columns of $\breve{\mathbf{H}}_\ell$ corresponding to the kth user and $\boldsymbol{\mathcal{X}}_\ell^{(k)}$ is the convolution matrix (3.8) that contains the signal radiated by the $n_T^{(k)}$ antennas of the kth user. Similarly the received vector (3.9) reads with congruent definitions

$$\mathbf{y}_\ell = \sum_{k=1}^{K} \mathbf{X}_\ell^{(k)} \mathbf{h}_\ell^{(k)} + \mathbf{n}_\ell, \tag{3.32}$$

in particular, $\mathbf{h}_\ell^{(k)}$ is the $n_T^{(k)} N_R W \times 1$ channel vector of the $n_T^{(k)}$ user.

For multicarrier transmission, according to the definitions of Sec. 3.3.1, the signal received over the pilot subcarriers can be written as (3.32).

Data phase (or data subcarriers)

According to the discussion above, the signal received over the data part of the block is for both time-domain and multicarrier transmission:

$$\mathbf{y}_\ell = \sum_{k=1}^{K} \boldsymbol{\mathcal{H}}_\ell^{(k)} \mathbf{x}_\ell^{(k)} + \mathbf{n}_\ell. \tag{3.33}$$

3.4.2 Downlink of a MIMO system (broadcast channel)

In a broadcast channel, the base station (transmitter) is equipped with N_T cooperating antennas whereas the receiving array is partitioned into K (non-cooperating) users, each having $n_R^{(k)}$ antennas such that $\sum_{k=1}^{K} n_R^{(k)} = N_R$.

Training phase (or pilot subcarriers)

For time-domain transmission, the signal received by the $n_R^{(k)}$ antennas of the kth user within the training part of the block can be written as

$$\mathbf{Y}_\ell^{(k)} = \breve{\mathbf{H}}_\ell^{(k)} \boldsymbol{\mathcal{X}}_\ell + \mathbf{N}_\ell^{(k)}, \tag{3.34}$$

where $\breve{\mathbf{H}}_\ell^{(k)}$ is obtained by selecting the rows of $\breve{\mathbf{H}}_\ell$ corresponding to the receiving antennas of the kth user and $\boldsymbol{\mathcal{X}}_\ell$ is the convolution matrix built from the training symbols transmitted in broadcast to all users. The received vector (3.9) reads with congruent definitions

$$\mathbf{y}_\ell^{(k)} = (\boldsymbol{\mathcal{X}}_\ell \otimes \mathbf{I}_{n_R^{(k)}}) \mathbf{h}_\ell^{(k)} + \mathbf{n}_\ell, \tag{3.35}$$

vector $\mathbf{h}_\ell^{(k)}$ is the $N_T n_R^{(k)} W \times 1$ channel vector of the $n_T^{(k)}$ user.

For MIMO-OFDM, as explained above, the signal model is still (3.35).

Data phase (or data subcarriers)

The signal received by the kth user over the data part of the block (time-domain or multicarrier transmission) is

$$\mathbf{y}_\ell^{(k)} = \boldsymbol{\mathcal{H}}_\ell^{(k)} \sum_{i=1}^{K} \mathbf{x}_\ell^{(i)} + \mathbf{n}_\ell. \tag{3.36}$$

3.4.3   Signal to noise ratio definition

In a multiuser setting, the fading channel of each user may have different average power in order to account for different propagation conditions (path loss, shadowing). Let the users be ordered for decreasing average power. Accordingly, the channel norm of the kth user will be scaled as compared to the first user by an attenuation $\alpha^{(k)} \leq 1$ (i.e., users are order for decreasing power, $\alpha^{(1)} = 1$). Moreover, the power $P^{(k)}$ radiated by each user in uplink or the power destined to each user in downlink can be different (due to power control policies). However, we define for simplicity a SNR that accounts for the total power radiated P (by the base station in downlink or by all the user in uplink) as

$$SNR = \frac{P}{\sigma_n^2}, \tag{3.37}$$

where σ_n^2 is the noise power either at each user (downlink) or by the base station (uplink).

3.5   Conclusion

In this Chapter, a review of the signal model used for the analysis of time and frequency (i.e., multicarrier) domain transmission and reception over a single and multiuser MIMO link has been presented. The presentation was aimed at showing the duality of the two approaches. In particular, it was proved that for a single user MIMO link, the signal received over the L_P samples/subcarriers of the training part of the block can be written as the $N_R L_P \times 1$ vector $\mathbf{y}_\ell$ as

$$\mathbf{y}_\ell = \mathbf{X}_\ell \mathbf{h}_\ell + \mathbf{n}_\ell, \tag{3.38}$$

where $\mathbf{X}_\ell$ is the block convolution matrix defined in (3.9) for time-domain transmission and the block diagonal matrix defined in (3.27) for multicarrier transmission. It contains the training sequences sent by different antennas. On the other hand, the signal received over the L_D samples/subcarriers of the data part of the block can be arranged into the $N_R L_D \times 1$ vector $\mathbf{y}_\ell$ as

$$\mathbf{y}_\ell = \boldsymbol{\mathcal{H}}_\ell \mathbf{x}_\ell + \mathbf{n}_\ell, \tag{3.39}$$

where $\boldsymbol{\mathcal{H}}_\ell$ is the block convolution matrix defined in (3.13) for time-domain transmission and the block diagonal matrix in (3.30) for multicarrier transmission. We remark that the ability of the multicarrier approach to turn the block convolution matrices $\mathbf{X}_\ell$ and $\boldsymbol{\mathcal{H}}_\ell$ into block diagonal matrix is due to the transmission of the cyclic prefix and is commonly referred to by stating that frequency-domain transmission is able to turn a frequency-selective channel into a set of parallel frequency-flat channels.

3.6 Appendix: derivation of the signal model in the frequency domain (3.15)

The received signal (in the temporal domain) within the ℓth OFDM symbol can be written as in (3.9), here recalled for reference

$$\mathbf{y}_\ell = (\boldsymbol{\mathcal{X}}_\ell \otimes \mathbf{I}_{N_R})\mathbf{h}_\ell + \mathbf{n}_\ell = \mathbf{X}_\ell \mathbf{h}_\ell + \mathbf{n}_\ell, \tag{3.40}$$

where $\mathbf{y}_\ell$ is $N_R L \times 1$ and the channel vector $\mathbf{h}_\ell$ is redefined here as a $N_R N_T L \times 1$ vector by zero padding the channel vector, i.e., with a slight abuse of notation $\mathbf{h}_\ell = [\mathbf{h}_\ell^T \ \mathbf{0}^T]^T$. In this way, matrix $\boldsymbol{\mathcal{X}}_\ell$ in (3.8) is $L \times N_T L$ and is block circulant because of the cyclic prefix. In particular, matrix $\boldsymbol{\mathcal{X}}_\ell$ is obtained by interlacing circulant $L \times L$ square convolution matrices $\boldsymbol{\mathcal{X}}_\ell^{(n_T)}$, $n_T = 1, ..., N_T$, corresponding to the blocks radiated by each transmitting antenna as in $[\boldsymbol{\mathcal{X}}_\ell]_{n_T + N_T(i-1), j} = [\boldsymbol{\mathcal{X}}_\ell^{(n_T)}]_{i,j}$. But since matrices $\boldsymbol{\mathcal{X}}_\ell^{(n_T)}$ are circulant, their eigenvectors correspond to the complex exponential vectors in matrix (3.17), here redefined as full $L \times L$ DFT matrices, as

$$\boldsymbol{\mathcal{X}}_\ell^{(n_T)} = \boldsymbol{\Theta}^H \widetilde{\boldsymbol{\mathcal{X}}}_\ell^{(n_T)} \boldsymbol{\Theta}, \tag{3.41}$$

where $\widetilde{\boldsymbol{\mathcal{X}}}_\ell^{(n_T)}$ is a $L \times L$ diagonal matrix containing the eigenvalues of $\boldsymbol{\mathcal{X}}_\ell^{(n_T)}$, $\tilde{x}_\ell^{(n_T)}[k]$ $k = 0, ..., L-1$:

$$\widetilde{\boldsymbol{\mathcal{X}}}_\ell^{(n_T)} = \begin{bmatrix} \tilde{x}_\ell^{(n_T)}[0] & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{x}_\ell^{(n_T)}[L-1] \end{bmatrix} \tag{3.42}$$

It is then straightforward to show that

$$\boldsymbol{\mathcal{X}}_\ell = \boldsymbol{\Theta}^H \widetilde{\boldsymbol{\mathcal{X}}}_\ell (\boldsymbol{\Theta} \otimes \mathbf{I}_{N_T}), \tag{3.43}$$

where $\widetilde{\boldsymbol{\mathcal{X}}}_\ell$ defined as the $L \times N_T L$ matrix (see (3.24))

$$\widetilde{\boldsymbol{\mathcal{X}}}_\ell = \begin{bmatrix} \tilde{\mathbf{x}}_\ell[0]^T & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{\mathbf{x}}_\ell[L-1]^T \end{bmatrix} \tag{3.44}$$

with $\tilde{\mathbf{x}}_\ell[k] = [\tilde{x}_\ell^{(n_T)}[0] \cdots \tilde{x}_\ell^{(n_T)}[L-1]]^T$.

Therefore, by applying the DFT transform $\boldsymbol{\Theta}$ on the signal received by each antenna (i..e., transforming the signal from the time to the frequency domain), we easily get the $N_R L \times 1$ signal in the frequency domain:

$$\tilde{\mathbf{y}}_\ell = (\boldsymbol{\Theta} \otimes \mathbf{I}_{N_R})\mathbf{y}_\ell = (\widetilde{\boldsymbol{\mathcal{X}}}_\ell \otimes \mathbf{I}_{N_T N_R})\mathbf{f}_\ell + \tilde{\mathbf{n}}_\ell, \tag{3.45}$$

with $\mathbf{f}_\ell$ defined as in (3.18) and $\tilde{\mathbf{n}}_\ell = (\boldsymbol{\Theta} \otimes \mathbf{I}_{N_R})\mathbf{n}_\ell$ that has the same statistics as $\mathbf{n}_\ell$ ($\boldsymbol{\Theta}$ is a unitary matrix). Equation (3.45) implies the signal model (3.15), where the signal transmitted on each subcarrier $\tilde{\mathbf{x}}_\ell[\bar{k}]$ is received with no interference from other sucarriers $k \neq \bar{k}$. Notice

that outside this Section, as anticipated in the remark in Sec. 3.3, the tilde decoration has been dropped in defining the quantities in the signal domain for simplicity of notation.

As a final step, it is interesting to investigate the relationship between the signal transmitted on each subcarrier by the n_Tth transmitting antenna and the corresponding signal in the time domain, i.e., to expand (3.41)-(3.42), obtaining

$$x_\ell^{(n_T)}[m] = \frac{1}{L} \sum_{k=0}^{L-1} \tilde{x}_\ell^{(n_T)}[k] \exp(j\frac{2\pi}{L}mk), \tag{3.46}$$

that is, the signal in the temporal domain is computed as in fig. 3.4 through the IDFT transform with an appropriate scaling factor.

Bibliography

[1] A. Kocian and B. H. Fleury, "EM-based joint data detection and channel estimation of DS-CDMA signals," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1709-1720, Oct. 2003.

[2] H. Holma and A. Toskala, *WCDMA for UMTS,* John Wiley & Sons, 2000.

[3] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville, P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2136-2148, Dec. 2002.

[4] H. Meyr, M. Moeneclaey and S. A. Fechtel, *Digital communication receivers*, John Wiley & Sons Inc., 1998.

[5] M. C. Bromberg, "Optimizing MIMO multipoint wireless networks assuming Gaussian other-user interference," IEEE Inform. Theory, vol. 49, no. 10, pp. 2352-2362, Oct. 2003.

[6] S. Verdu, *Multiuser detection*, Cambridge University Press, 1998.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, 1993.

[8] B. Hassibi, B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-963, April 2003.

[9] H. Vikalo, B. Hassibi, B. Hochwald and T. Kailath, "On the capacity o frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2572-2583, Sept. 2004.

[10] IEEE Std 802.11a-1999, "Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5Ghz Band".

[11] H. Sari, G. Karam and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, no. 7, pp. 100-109, Feb. 1995.

[12] Y. Li, and L. J. Cimini, "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Comm.*, vol. 49, no. 3, pp. 401-404, March 2001.

[13] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electronics*, vol. 44, pp. 1122-1128, Aug. 1998.

[14] S. Adireddy, L. Tong and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2338-2353, Aug. 2002.

[15] J. Armstrong, "Analysis of new and existing methods of reducing intercarrier interference due to carrier frequency offset in OFDM," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 365-369, March 1999.

[16] Songping Wu and Y. Bar-Ness, "Performance analysis on the effect of phase noise in OFDM systems," in *Proc. IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications 2002*, vol 1, pp. 133-138, 2002.

[17] L. Goldfeld, V. Lyandres and D. Wulich, "Minimum BER power loading for OFDM in fading channel," *IEEE Trans. Commun.*, vol. 50, no. 11, pp. 1729-1733, Nov. 2002.

Chapter 4

Lower bound on the channel estimation error for frequency-selective MIMO channels

4.1 Introduction and problem formulation

IN the context of training-based transmission, as a preliminary operation before data detection, CSI acquisition is carried out. This is performed by processing the training part of the block (be it in time or frequency domain, see Chapter 3) is processed so as to yield an estimate of the channel vector $\mathbf{h}_\ell$. The performance of the data detector strongly depends on the quality of the channel estimate available [1]. Therefore, it is of interest to investigate theoretical limits on the accuracy of channel estimation for the propagation scenarios described in Chapter 2.

In this Chapter, a lower bound on the channel estimation error for any unbiased estimator is derived by means of the hybrid CRB (HCRB), a modification of the classical CRB for the case where the unknown parameter vector depends on both deterministic variables and random variables. This analytical tool is well suited for our application since, according to the discussion in Chapter 2, the channel vector can be parametrized by decoupling long term (deterministic) parameters and fast varying (random) fading amplitudes.

To formulate the problem within an analytical framework, we can proceed by recalling for reference the main relationships derived in the previous Chapters. The L_P samples received by the N_R receiving antennas within the training block are gathered in the $L_P N_R \times 1$ vector $\mathbf{y}_\ell$:

$$\mathbf{y}_\ell = \mathbf{X}\mathbf{h}_\ell + \mathbf{n}_\ell, \tag{4.1}$$

where definition of the training matrix $\mathbf{X}$ depends on the transmission strategy (see (3.8) for time-domain transmission and (3.24) for multicarrier transmission). In order to simplify the analysis, in (4.1) we made the reasonable assumption, that is met in many communication standards, see, e.g., [2], that the training matrix $\mathbf{X}$ is the same for each block[1]. The situation

---

[1]However, it can be shown that the results presented in the following hold whenever the correlation properties

where the training sequences are selected adaptively on a block-by-block basis is studied in Appendix 10. Moreover, we have $E[\mathbf{n}_\ell \mathbf{n}_\ell^H] = \mathbf{I}_{L_P} \otimes \mathbf{R}_n$ with $\mathbf{R}_n$ denoting the $N_R \times N_R$ spatial correlation of AGN. The $N_R N_T W \times 1$ channel vector $\mathbf{h}_\ell$ can be parametrized as

$$\mathbf{h}_\ell = \mathbf{T}\boldsymbol{\beta}_\ell, \tag{4.2}$$

where the $N_R N_T W \times N_F$ long term matrix $\mathbf{T}$ is defined according to (2.25) for the beamforming and (2.34) for the diversity scenario, whereas the $N_F \times 1$ fast fading vector $\boldsymbol{\beta}_\ell$ is a zero mean circular Gaussian random variable with correlation $E[\boldsymbol{\beta}_\ell \boldsymbol{\beta}_{\ell-m}^H] = \boldsymbol{\Phi}(\ell,m)$, specified in (2.19) and (2.31) for the beamforming and diversity scenario respectively.

Toward the goal of simplifying the analysis and obtaining a lower bound on the channel estimation error, the quasi-static model of temporal variations of the long term features of the channel is assumed, as exemplified by (4.2). In particular, we will consider model (4.2) to hold for $\ell = 1, ..., N_B$ blocks with N_B denoting the interval of stationarity of the long term channel parameters. In the next Chapter, when discussing practical channel estimator this condition will be dropped.

Let $\tilde{\mathbf{h}}_\ell$ be an estimate of the MIMO channel. The purpose is to evaluate a lower bound $\mathbf{Q}_{\hat{\mathbf{h}}_\ell}$ on the error correlation matrix

$$\mathbf{Q}_{\tilde{\mathbf{h}}_\ell} = E[(\tilde{\mathbf{h}}_\ell - \mathbf{h}_\ell)(\tilde{\mathbf{h}}_\ell - \mathbf{h}_\ell)^H] \geq \mathbf{Q}_{\hat{\mathbf{h}}_\ell} \tag{4.3}$$

where expectation in (4.3) is with respect to noise and fading. The corresponding bound on the mean square error (MSE) is

$$MSE_{\tilde{\mathbf{h}}_\ell} = \mathrm{tr}\{\mathbf{Q}_{\tilde{\mathbf{h}}_\ell}\} \geq MSE_{\hat{\mathbf{h}}_\ell} = \mathrm{tr}\{\mathbf{Q}_{\hat{\mathbf{h}}_\ell}\}. \tag{4.4}$$

4.2 Uncostrained ML channel estimation (UML)

Before going into the details of the derivation of a lower bound (4.3) on the performance of a channel estimator based on the parametrization (4.2), it is convenient to review the conventional ML channel estimator [4] [5] [6] . The latter performs an unconstrained estimate of the channel vector starting from the received signal model (4.1). In other words, it does not assume any prior information on the structure of the channel vector. It can be easily shown (see, e.g., [6] [4]) that the Uncostrained ML (UML) channel estimator reads

$$\mathbf{h}_{UML,\ell} = (\mathbf{R}_x^{-1} \otimes \mathbf{I}_{N_R})\mathbf{X}^H \mathbf{y}_\ell \tag{4.5}$$

where we used the fact that $\mathbf{X}^H \mathbf{X} = \mathbf{R}_x \otimes \mathbf{I}_{N_R}$. Notice that the estimator requires that the training sequence length L_P to satisfiy the condition $L_P \geq N_T W$ (see also Sec. 3.2.1 and

---

of $\mathbf{X}_\ell$, i.e, matrix $\mathbf{R}_x$, are unchanged over the blocks.

Sec. 3.3.1). Moreover, it is straightforward to prove that the estimator is unbiased and the error correlation matrix

$$
\begin{aligned}
\mathbf{Q}_{UML} &= E[(\mathbf{h}_{UML,\ell} - \mathbf{h}_\ell)(\mathbf{h}_{UML,\ell} - \mathbf{h}_\ell)^H] = \\
&= \mathbf{R}_x^{-1} \otimes \mathbf{R}_n
\end{aligned}
\tag{4.6}
$$

depends on the spatial correlation of AGN $\mathbf{R}_n$ and on the correlation properties of the training sequences (see also Sec. 4.7.1).

4.3 Overview of this Chapter

As stated above, this Chapter is concerned with the derivation of a lower bound on the channel estimation error (4.3)-(4.4) by taking into account the algebraic structure of the channel presented in the previous Chapter and reviewed in Sec. 4.4. As a consequence of this analysis, we will:

- derive an optimal channel estimation strategy (see Sec. 4.6), which is proved to promise remarkable performance improvement as compared to conventional unstructured estimators, such as UML (see also fig. 4.3). Practical implementation of this optimal strategy will be discussed in the next Chapter;

- quantify the impact of system and channel parameters on the channel estimation error (see Sec. 4.7.2);

- address a trade-off between computational complexity of channel estimation and performance as a function of the channel characteristics (see Sec. 4.10).

4.4 Algebraic structure of the MIMO channel revisited

As recalled above, Chapter 2 showed that the channel vector $\mathbf{h}_\ell$ has an algebraic structure that allows a simple decoupling of the long term and short term features of the channel, see (4.2). A channel estimator designed so as to be able to exploit this property is expected to achieve large perfomance gains as compared to unconstrained channel estimator (such as UML). However, in designing such an estimator, further insight on the properties of identifiability of model (4.2) is needed.

To elaborate on this point and in order to simplify the analysis, where not stated otherwise, we will assume *continuous transmission* and the *same Doppler spectrum for each path* (see Chapter 2) so that, denoting by $\varphi(m)$ the corresponding temporal correlation of all paths, for both scenarios the $N_F \times N_F$ fading correlation reads

$$
\mathbf{\Phi}(m) = \mathbf{\Phi}(0)\varphi(m),
\tag{4.7}
$$

where from (2.20) and (2.32)

$$\mathbf{\Phi}(0) = \mathbf{I}_d \qquad \text{(beamforming scenario)} \qquad (4.8a)$$

$$\mathbf{\Phi}(0) = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_d \end{bmatrix} \qquad \text{(diversity scenario)} \qquad (4.8b)$$

According to this assumption, the channel model (4.2) can be equivalently stated by saying that $\mathbf{h}_\ell$ is a complex circular Gaussian vector with correlation matrix

$$E[\mathbf{h}_\ell \mathbf{h}_{\ell-m}^H] = \mathbf{R}_h \varphi(m), \qquad (4.9)$$

where for reference the correlation matrix $\mathbf{R}_h = \mathbf{T}\mathbf{\Phi}(0)\mathbf{T}^H$ is defined. It follows that the unknown vector $\mathbf{h}_\ell$ can be written as

$$\mathbf{h}_\ell = \mathbf{T}\mathbf{\Phi}(0)^{1/2}\mathbf{b}_\ell, \qquad (4.10)$$

where $\mathbf{\Phi}(0)^{1/2}$ is a $N_F \times N_F$ full rank square-root matrix of $\mathbf{\Phi}(0)$, and the $N_F \times 1$ vector $\mathbf{b}_\ell$ is a stationary Gaussian process with

$$E[\mathbf{b}_\ell \mathbf{b}_{\ell-m}] = \mathbf{I}_{N_F} \varphi(m). \qquad (4.11)$$

4.4.1 Beamforming scenario: space-time modes

For the beamforming scenario, the long term matrix $\mathbf{T}$ depends on the space-time signatures of different paths, $\mathbf{s}_{ST,i} = \Omega_i^{1/2}\mathbf{g}(\tau_i) \otimes \mathbf{a}_T(\alpha_i^{(T)}) \otimes \mathbf{a}_R(\alpha_i^{(R)})$, $i = 1, ..., d$ as (recall (2.25))

$$\mathbf{T} = \mathbf{S}_{ST} = [\mathbf{s}_{ST,1} \cdots \mathbf{s}_{ST,d}] \qquad (4.12)$$

where we have defined the $N_R N_T W \times d$ matrix $\mathbf{S}_{ST}$ collecting the path signatures. Different paths having similar propagation parameters $(\tau_i, \alpha_i^{(T)}, \alpha_i^{(R)})$ may yield linearly dependent space-time signatures (see [26] for a discussion in the context of SIMO systems). In particular, when the separation between delays (and/or angles) is below the temporal (and spatial) resolution of the receiver, the number of resolvable space-time signatures reads

$$r_{ST} = \text{rank}\{\mathbf{S}_{ST}\} \le d. \qquad (4.13)$$

As a consequence, *matrix $\mathbf{T}$ is rank-deficient:*

$$r = \text{rank}\{\mathbf{T}\} = r_{ST} \le \min\{N_R N_T W, N_F\}, \qquad (4.14)$$

where $N_F = d$ is the number of fading amplitudes for the beamforming scenario. Therefore, model (4.2) (or equivalently (4.10)) is not a minimal parametrization for the channel vector [8].

The inequality (4.14) is generally strictly satisfied, i.e., $r_{ST} \ll \min\{N_R N_T W, N_F\}$, see, e.g., [7] for an analysis in the context of standardized channel model.

Based on the discussion above, the study of the channel estimation process requires consideration of an alternative minimal channel parametrization. Since $\mathbf{T} = \mathbf{S}_{ST}$ is generally rank-deficient, it can be conveniently parametrized as the product of two full rank matrices $\mathbf{U}$ ($N_R N_T W \times r$) and $\mathbf{C}$ ($N_F \times r$): $\mathbf{T} = \mathbf{U}\mathbf{C}^H$ [9]. Therefore, the model (4.10) becomes (recall (4.8a))

$$\mathbf{h}_\ell = \mathbf{U}\mathbf{C}^H \mathbf{b}_\ell \qquad (4.15)$$

The r columns of matrix $\mathbf{U}$ will be referred to as *space-time modes* of the channel since they span the subspace described by the space-time signatures of different paths, i.e., $\mathrm{span}\{\mathbf{S}_{ST}\} = \mathrm{span}\{\mathbf{U}\}$. Without limiting the generality, each $N_R N_T W \times 1$ space-time mode is assumed to have unit norm. The definition of *modes* is justified by the fact that the subspace $\mathrm{span}\{\mathbf{U}\}$ is a stationary feature of the channel according to model (4.2).

The parametrization (4.15) is not unique since decomposition of $\mathbf{T}$ into its factors $\mathbf{U}$ and $\mathbf{C}$ can be calculated in different ways [9]. For instance, orthonormal space-time modes can be easily defined by performing a singular value decomposition of the modal matrix $\mathbf{S}_{ST} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H$ and then setting $\mathbf{C} = \mathbf{V}\boldsymbol{\Lambda}^H$

Remark 4 (*on the minimal parametrization*): *one could object that a minimal parametrization of the channel vector should imply the exploitation of the functional dependence of matrix* $\mathbf{T}$ *on the path parameters* $(\tau_i, \alpha_i^{(T)}, \alpha_i^{(R)})$. *However, following the lines of [10] it could be shown that the asymptotic result derived in the following hold even for this kind of parametrization. Moreover, the approach proposed here has the advantage of disclosing useful guidelines for the implementation of channel estimators that are i) able to avoid the impairments of non-linear estimation related to the direct computation of* $(\tau_i, \alpha_i^{(T)}, \alpha_i^{(R)})$; *ii) able to cope with channels with non-resolved paths (see discussion above).*

4.4.2 Diversity scenario: temporal modes

In the context of a diversity scenario, the long term matrix $\mathbf{T}$ depends on the temporal signatures of different paths $\mathbf{s}_{T,i} = \Omega_i^{1/2} \mathbf{g}(\tau_i)$, $i = 1, ..., d$ through the $W \times d$ matrix $\mathbf{S}_T = [\mathbf{s}_{ST,1} \cdots \mathbf{s}_{ST,d}]$ as (recall (2.34))

$$\mathbf{T} = \mathbf{S}_T \otimes \mathbf{I}_{N_R N_T}. \qquad (4.16)$$

According to the discussion above, different paths having delays τ_i separated by less than the temporal resolution of the receiver do not contribute to matrix $\mathbf{S}_T$ with linearly independent signatures, so that the number of resolvable paths in the time domain is

$$r_T = \mathrm{rank}\{\mathbf{S}_T\} \le d \qquad (4.17)$$

and the rank of matrix $\mathbf{T}$ is

$$r = \text{rank}\{\mathbf{T}\} = N_R N_T r_T \leq \min\{N_R N_T W, N_F\}, \tag{4.18}$$

where $N_F = N_R N_T d$ is the number of fading amplitudes for the diversity scenario. Therefore, model (4.2) (or equivalently (4.10)) is not a minimal parametrization for the channel vector, even though the attainable reduction in model complexity is generally less relevant than in a beamforming scenario (compare (4.18) with (4.14)).

Since matrix $\mathbf{S}_T$ is reduced-rank, we can use the following parametrization: $\mathbf{S}_T = \mathbf{U}_T \mathbf{L}_T$, where the two full rank matrices $\mathbf{U}_T$ and $\mathbf{L}_T$ are $W \times r_T$ and $N_F \times r_T$. Similarly to the beamforming scenario, the colums of $\mathbf{U}_T$ will be referred to as *temporal modes* of the channel since they span the subspace described by the temporal signatures of different paths, i.e., $\text{span}\{\mathbf{S}_T\} = \text{span}\{\mathbf{U}_T\}$. Again, the definition of *modes* is justified by the fact that the subspace $\text{span}\{\mathbf{U}_T\}$ is a stationary feature of the channel according to model (4.2). Without limiting the generality of our approach, each $W \times 1$ temporal mode is assumed to have unit norm. Therefore, model (4.10) becomes $\mathbf{h}_\ell = (\mathbf{U}_T \otimes \mathbf{I}_{N_R N_T})(\mathbf{L}_T \otimes \mathbf{I}_{N_R N_T})\boldsymbol{\Phi}(0)^{1/2}\mathbf{b}_\ell$ and defining the $N_F \times r$ matrix $\mathbf{C} = \boldsymbol{\Phi}(0)^{H/2}(\mathbf{L}_T^H \otimes \mathbf{I}_{N_R N_T})$ we get

$$\mathbf{h}_\ell = \mathbf{U}\mathbf{C}^H \mathbf{b}_\ell \tag{4.19}$$

with

$$\mathbf{U} = \mathbf{U}_T \otimes \mathbf{I}_{N_R N_T}. \tag{4.20}$$

Comparing (4.19) with (4.15), we can conclude that with appropriate definitions the minimal parametrization for the beamforming and diversity scenarios coincide. Recall that parametrization (4.19) is not unique as explained in the previous Section. For instance, orthonormal temporal modes can be easily defined by computing the singular value decomposition of the temporal modal matrix $\mathbf{S}_T = \mathbf{U}_T \boldsymbol{\Lambda}_T \mathbf{V}_T^H$ and then setting $\mathbf{L}_T = \mathbf{V}_T \boldsymbol{\Lambda}_T^H$.

4.5   Hybrid CRB (HCRB)

As stated in the Introduction, the channel vector $\mathbf{h}_\ell$ depends on both deterministic (i.e., long term) and random (i.e., fading amplitudes) quantities. Therefore, a lower bound on the correlation matrix of the estimation error for any unbiased estimator $\left\{\tilde{\mathbf{h}}_\ell\right\}_{\ell=1}^{N_B}$ can be obtained by means of the HCRB [11]. Recall that N_B is the interval of stationarity (in blocks) of the long term channel features. To elaborate, according to the minimal parametrization (4.15) the long term parameters are accounted for by the full rank matrices $\mathbf{U}$ and $\mathbf{C}$ whereas the fading amplitudes are represented by vectors $\{\mathbf{b}_\ell\}_{\ell=1}^{N_B}$.

Gathering the channel vectors within the interval of stationarity N_B into the vector $\mathbf{h} = [\mathbf{h}_1^T \cdots \mathbf{h}_{N_B}^T]^T$ and the corresponding estimates into $\tilde{\mathbf{h}} = [\tilde{\mathbf{h}}_1^T \cdots \tilde{\mathbf{h}}_{N_B}^T]^T$, the HCRB provides the desired bound $\mathbf{Q}_{\hat{\mathbf{h}}}$

$$\mathbf{Q}_{\tilde{\mathbf{h}}} = E[(\tilde{\mathbf{h}} - \mathbf{h})(\tilde{\mathbf{h}} - \mathbf{h})^H] \geq \mathbf{Q}_{\hat{\mathbf{h}}}. \tag{4.21}$$

Computation of the HCRB requires at first the evaluation of the Fisher Information Matrix $\mathbf{J}_B$ relative to the unknown parameters $\mathbf{U}$, $\mathbf{C}$ and $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_{N_B}]$ averaged over the distribution of $\mathbf{B}$. Then, the lower bound (4.21) is computed as [11]

$$\mathbf{Q}_{\hat{\mathbf{h}}} = E_{\mathbf{b}}\left[\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{c}^T \mathbf{b}^T]}\right] \cdot \mathbf{J}_B^{\dagger} \cdot E_{\mathbf{b}}\left[\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{c}^T \mathbf{b}^T]}\right]^H, \tag{4.22}$$

where we defined $\mathbf{u} = \mathrm{vec}(\mathbf{U})$, $\mathbf{c} = \mathrm{vec}(\mathbf{C}^H)$ and $\mathbf{b} = \mathrm{vec}(\mathbf{B})$. A proof of (4.21)-(4.22) adapted from [12], can be found in Appendix-A. Computing (4.22) we get (see Appendix-B)

$$\mathbf{Q}_{\hat{\mathbf{h}}} = \mathbf{R}_t \otimes \mathbf{R}_h - (\mathbf{R}_t \otimes \mathbf{R}_h)(\mathbf{R}_t \otimes \mathbf{R}_h + \mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML})^{-1}(\mathbf{R}_t \otimes \mathbf{R}_h), \tag{4.23}$$

with definition

$$\mathbf{R}_t = \begin{bmatrix} \varphi(0) & \varphi(1) & \cdots & \varphi(N_B - 1) \\ \varphi(-1) & \varphi(0) & & \\ \vdots & & \ddots & \\ \varphi(-N_B + 1) & & & \varphi(0) \end{bmatrix}, \tag{4.24}$$

and $\mathbf{Q}_{UML}$ is the error correlation matrix of the UML estimator (4.6). Notice that if the fading amplitudes are temporally uncorrelated, i.e., $\varphi(m) = \delta(m)$, $\mathbf{R}_t = \mathbf{I}_{N_B}$ and the error correlation matrix (4.23) becomes block diagonal, yielding the bound on each block (4.3):

$$\mathbf{Q}_{\hat{\mathbf{h}}_\ell} = \mathbf{R}_h - \mathbf{R}_h(\mathbf{R}_h + \mathbf{Q}_{UML})^{-1}\mathbf{R}_h. \tag{4.25}$$

Moroever, as expected from the discussion in Sec. 4.4, the HCRB (4.23) depends only on $\mathbf{R}_h$ (and not on the specific choice of the factors $\mathbf{U}$ and $\mathbf{C}$).

4.6 Asymptotically optimal channel estimation strategy

Inspection of the HCRB (4.23) provides insight into the (asymptotically) optimal strategy for designing a channel estimator. In fact, let us consider a channel estimator based on a *separate estimation of the long term and short term channel parameters*. More specifically, let the estimator perform:

1. a consistent estimate of the long term features of the channel, i.e., the channel correlation matrix $\mathbf{R}_h$ (or any two factors $\{\mathbf{U}, \mathbf{C}\}$ up to their ambiguity) is estimated with an arbitrarily small error as N_B increases;

2. MMSE filtering of the channel vectors based on the estimate of $\mathbf{R}_h$ previously obtained (the temporal correlation $\varphi(m)$ is assumed to be known).

It can be easily shown from standard results [13] that the asymptotic ($N_B \to \infty$) error correlation matrix of this class of estimators coincides with the HCRB (4.23). We can conclude that the estimation strategy is optimal when the interval of stationarity of the long term features N_B is large enough.

4.7   Asymptotic HCRB

The result of the previous Section not only yields useful indications on the channel estimation design (as discussed in the next Chapter) but can also be used in order to evaluate the asymptotic ($N_B \rightarrow \infty$) expression of the HCRB (4.23).

Toward this goal and in order to ease the presentation of practical channel estimators in the next Chapter, it is convenient to define from (4.15) an equivalent parametrization featuring a reduced set of $r \times 1$ random amplitudes $\mathbf{d}_\ell$

$$\mathbf{h}_\ell = \mathbf{U}\mathbf{d}_\ell, \tag{4.26}$$

with $\mathbf{d}_\ell = \mathbf{C}^H\mathbf{b}_\ell$. Accordingly, vector $\mathbf{d}_\ell$ is zero mean random circular Gaussian with correlation

$$E[\mathbf{d}_\ell\mathbf{d}_{\ell-m}^H] = \mathbf{C}^H\mathbf{C}\varphi(m) = \mathbf{R}_d\varphi(m). \tag{4.27}$$

The asymptotic bound is now evaluated in a constructive way by computing the asymptotic performance of the (asymptotically) optimal channel estimation strategy presented in the previous Section. Accordingly, for $N_B \rightarrow \infty$ the long term features of the channel, accounted for in (4.26)-(4.27) by matrices $\mathbf{U}$ and $\mathbf{R}_d$, are assumed to be reliably estimated. Notice that similar assumptions have been exploited to evaluate the performance of a multichannel MLSE with channel estimation in [14]. The signal model (4.1) can now be restated in terms of the (asymptotically) known matrix $\mathbf{F} = \mathbf{X}\mathbf{U}$ as

$$\mathbf{y}_\ell = \mathbf{F}\mathbf{d}_\ell + \mathbf{n}_\ell. \tag{4.28}$$

The optimum channel estimation strategy presented in the previous Section for $N_B \rightarrow \infty$ reduces to the MMSE estimation of the amplitudes $\mathbf{d}_\ell$. This can be obtained within the infinite temporal horizon of our framework, by a Wiener filter that estimates the amplitudes in the frequency domain:

$$\mathcal{F}\{\hat{\mathbf{d}}_\ell\} = \mathbf{S}_{dy}(\omega)\mathbf{S}_{yy}(\omega)^{-1}\mathcal{F}\{\mathbf{y}_\ell\} \tag{4.29}$$

where $\mathbf{S}_{dy}(\omega) = \mathcal{F}\{E[\mathbf{d}_\ell\mathbf{y}_{\ell-m}^H]\}$ denotes the discrete-time Fourier transform of the crosscorrelation matrix between $\{\mathbf{d}_\ell\}$ and $\{\mathbf{z}_\ell\}$, $\mathbf{S}_{yy}(\omega)$ is similarly defined. Since the spatio-temporal correlation of the fading is separable (4.27), the MMSE estimate (4.29) depends on the power spectral density of the fading variations $S_\varphi(\omega) = \mathcal{F}\{\varphi(m)\}$ as

$$\mathbf{S}_{dy}(\omega)\mathbf{S}_{yy}(\omega)^{-1} = S_\varphi(\omega)\mathbf{R}_d\mathbf{F}^H\left(S_\varphi(\omega)\mathbf{F}\mathbf{R}_d\mathbf{F}^H + (\mathbf{I}_{N_B}\otimes\mathbf{R}_n)\right)^{-1}, \tag{4.30}$$

where we used the following equalities: $\mathbf{S}_{dy}(\omega) = \mathbf{S}_{dd}(\omega)\mathbf{F}^H$, $\mathbf{S}_{yy}(\omega) = \mathbf{F}\mathbf{S}_{dd}(\omega)\mathbf{F}^H + (\mathbf{I}_{L_P}\otimes\mathbf{R}_n)$ and $\mathbf{S}_{dd}(\omega) = \mathbf{R}_dS_\varphi(\omega)$. The error correlation matrix depends on the estimate of the amplitudes $\mathbf{d}_\ell$ as $\mathbf{Q}_{\hat{\mathbf{h}}_\ell} = \mathbf{U}\mathbf{Q}_{\hat{\mathbf{d}}_\ell}\mathbf{U}^H$, where $\mathbf{Q}_{\hat{\mathbf{d}}_\ell} = E[(\hat{\mathbf{d}}_\ell - \mathbf{d}_\ell)(\hat{\mathbf{d}}_\ell - \mathbf{d}_\ell)^H]$. In the

frequency domain, the error correlation matrix of the amplitudes reads

$$
\begin{aligned}
\mathbf{S}_{ee}(\omega) &= \mathcal{F}\{E[(\hat{\mathbf{d}}_\ell - \mathbf{d}_\ell)(\hat{\mathbf{d}}_{\ell-n} - \mathbf{d}_{\ell-m})^H]\} = \\
&= \mathbf{S}_{dd}(\omega) - \mathbf{S}_{dy}(\omega)\mathbf{S}_{yy}(\omega)^{-1}\mathbf{S}_{yd}(\omega) = \\
&= S_\varphi(\omega)(\mathbf{R}_d^{-1} + S_\varphi(\omega)\mathbf{F}^H(\mathbf{I}_{N_B}\otimes\mathbf{R}_n)\mathbf{F}) = \\
&= S_\varphi(\omega)(\mathbf{R}_d^{-1} + S_\varphi(\omega)\mathbf{R}_w)^{-1},
\end{aligned}
\tag{4.31}
$$

where for the second equality we used the matrix inversion lemma[2] and defined

$$
\mathbf{R}_w = \mathbf{U}^H(\mathbf{R}_x \otimes \mathbf{R}_n^{-1})\mathbf{U}.
\tag{4.32}
$$

Notice that $\mathbf{R}_w$ can be written in terms of the error correlation matrix of the UML estimate (4.6) as $\mathbf{R}_w = \mathbf{U}^H\mathbf{Q}_{UML}^{-1}\mathbf{U}$ if $L_P \geq N_T W$. By using the Parseval theorem, we have

$$
\begin{aligned}
\mathbf{Q}_{\hat{\mathbf{h}}_\ell} &= \mathbf{U}\int_{-\pi}^{\pi}\mathbf{S}_{ee}(\omega)\frac{d\omega}{2\pi}\mathbf{U}^H = \\
&= \mathbf{U}\int_{-\pi}^{\pi}S_\varphi(\omega)(\mathbf{R}_d^{-1} + S_\varphi(\omega)\mathbf{R}_w)^{-1}\frac{d\omega}{2\pi}\mathbf{U}^H.
\end{aligned}
\tag{4.33}
$$

Computation of the bound (4.33) requires integration over the Doppler spectrum $S_\varphi(\omega)$. In order to ease the analysis and allow to gain insight into the bound (4.33) in the Sec. 4.7.2 we consider a uniform Doppler spectrum.

4.7.1   Selection of training sequences: a review

As it is clear from (4.32)-(4.33), the channel estimation error depends on the correlation function of the training sequences $\mathbf{R}_x$. This is defined in (3.10) for time-domain transmission and in (3.28) for MIMO-OFDM and accounts for both the temporal correlation of the training sequences transmitted by each antenna and the mutual correlation of the training sequences of different antennas. As it is well known, the error is minimized if the training sequences are orthogonal both in time and across antennas, i.e.,

$$
\mathbf{R}_x = PL_P\mathbf{I}_{N_T W}
\tag{4.34}
$$

where P is the power radiated by the N_T antennas for each training symbol [5] [15].

For time-domain transmission, this result can be closely approximated by careful design of the training sequences, see, e.g., the UMTS standard [2]. On the other hand, for MIMO-OFDM if no guard band is allocated optimal training sequences can be easily designed either

- by selecting L_P equispaced (in the frequency domain), equipowered pilot subcarriers (comb pilot pattern) and letting all the transmitting antennas use all the L_P pilot subcarriers with phase shift orthogonal sequences [5] [15], or

---
[2] $(\mathbf{A}^{-1} + \mathbf{V}\mathbf{C}^{-1}\mathbf{V}^H)^{-1} = \mathbf{A} - \mathbf{A}\mathbf{V}(\mathbf{C} + \mathbf{V}^H\mathbf{A}\mathbf{V})^{-1}\mathbf{V}^H\mathbf{A}^H$

- by letting each antenna transmit over different (shifted) comb pilot patterns with equipow-
  ered symbols, arbitrarily selected [16].

However, in MIMO-OFDM systems, implementation issues call for the allocation of guard subcarriers. As a consequence, in practice, the optimality condition (4.34) can only be approx-imated.

4.7.2   Asymptotic HCRB for a uniform Doppler spectrum

The error correlation matrix bound $\mathbf{Q}_{\hat{\mathbf{h}}}$ can be easily evaluated in closed form for a uniform Doppler spectrum $S_\varphi(\omega) = 1/(2f_D T_S)$ over the support $\omega \in [-2\pi f_D T_s, +2\pi f_D T_s]$ ($0 \leq f_D T_s \leq 1/2$) with $f_D T_s$ denoting the normalized Doppler spread (see Chapter 2). In this case, the bound (4.33) simplifies as

$$\mathbf{Q}_{\hat{\mathbf{h}}_\ell} = 2f_D T_s \cdot \mathbf{U}(2f_D T_s \mathbf{R}_d^{-1} + \mathbf{R}_w)^{-1}\mathbf{U}^H \qquad (4.35)$$

and the corresponding MSE (4.4)

$$\begin{aligned} MSE_{\hat{\mathbf{h}}_\ell} &= 2f_D T_s \cdot \operatorname{tr}\{(2f_D T_s \mathbf{R}_d^{-1} + \mathbf{R}_w)^{-1}\} = \\ &= 2f_D T_s \cdot \sum_{i=1}^{r} \frac{1}{\mu_i[2f_D T_s \mathbf{R}_d^{-1} + \mathbf{R}_w]} \end{aligned} \qquad (4.36)$$

depends on the r eigenvalues of the $r \times r$ matrix $2f_D T_s \mathbf{R}_d^{-1} + \mathbf{R}_w$.

The bound (4.35)-(4.36) generalizes some known results on the performance of MMSE or ML channel estimation and some of these connections are discussed below under appropriate settings.

Static channel

In a static channel, fading is not varying across blocks so that $f_D = 0$ and from (4.35) it is $\mathbf{Q}_{\hat{\mathbf{h}}} = \mathbf{0}$. Indeed, in this case the channel vector is constant and can be consistently estimated with covariance $O(1/N_B)$.

Optimal training sequences and uncorrelated noise

Let us consider spatially white noise ($\mathbf{R}_n = \sigma_n^2 \mathbf{I}_{N_R}$) and ideal training sequences (i.e., or-thogonal between any two transmitting antennas and temporally uncorrelated), in this case $\mathbf{R}_w = PL_P/\sigma_n^2 \mathbf{I}_r$ and the MSE (4.36) can be evaluated for the two scenarios from the eigen-values $2f_D T_s \mu_i[\mathbf{R}_d^{-1}] + PL_P/\sigma_n^2$. In order to ease the analysis and without limiting the gen-erality of our results, we consider orthonormal channel (either space-time or temporal) modes as discussed in Sections 4.4.1 and 4.4.2.

As a reference, consider that under the stated assumptions the MSE of the UML estimator reads (easily proved from (4.6)):

$$MSE_{UML} = tr\{\mathbf{Q}_{UML}\} = \frac{\sigma_n^2}{L_P P} N_R N_T W. \tag{4.37}$$

Beamforming scenario   For a beamforming scenario we have $\mathbf{R}_d = \mathbf{\Lambda}^2$, hence $\mu_i[\mathbf{R}_d^{-1}] = 1/[\mathbf{\Lambda}]_{ii}^2$ and the asymptotic HCRB (4.36) is

$$MSE_{\hat{\mathbf{h}}_\ell} = 2f_D T_s \cdot \sum_{i=1}^{r} \frac{[\mathbf{\Lambda}]_{ii}^2}{2f_D + [\mathbf{\Lambda}]_{ii}^2 \frac{L_P P}{P}}. \tag{4.38}$$

For low SNR (or $L_P P/\sigma_n^2 \ll 2f_D T_s/[\mathbf{\Lambda}]_{ii}^2 \leq 1/[\mathbf{\Lambda}]_{ii}^2$) the MSE (4.38) is

$$MSE_{\hat{\mathbf{h}}_\ell} \simeq \sum_{i=1}^{r} [\mathbf{\Lambda}]_{ii}^2 = E[||\mathbf{h}_\ell||^2], \tag{4.39}$$

according to the standard behavior of the MMSE estimation [13], whereas for high SNR or small Doppler frequency (i.e., $L_P P/\sigma_n^2 \gg 2f_D T_s/[\mathbf{\Lambda}]_{ii}^2$) the MSE

$$MSE_{\hat{\mathbf{h}}_\ell} \simeq 2f_D T_s \frac{\sigma_n^2}{L_P P} r \tag{4.40}$$

is proportional to $r = \text{rank}\{\mathbf{T}\}$. Comparing (4.40) with (4.37), it is clear that *the (asymptotic) gain to be expected from the design of a channel estimator based on the knowledge of the channel model as compared to an uncostrained estimator can be quantified by the ratio* $N_R N_T W/(2f_D T_s r)$. Therefore, the gain decreases for increasing Doppler spread $f_D T_s$ and for increasingly dense multipath (i.e., larger r). This result can be easily interpreted by considering that the advantages in exploiting the channel model when designing an estimator are related to the possibility to: *i*) reduce the number of parameters to be estimated (which is only possible for $r < N_R N_T W$) and *ii*) use the prior information on the channel statistics to smooth the channel estimates (which is possible if $f_D T_s < 1/2$). See Sec. 4.8 for further analysis on this point.

Diversity scenario   Let us consider at first a frequency-flat channel in a diversity scenario, which constitutes the most investigated setting for MIMO systems in the open literature. Frequency-flatness occurs if the delays are not temporally resolvable (compared to the system bandwidth) so that $d = 1$ and $W = 1$. In this case $\mathbf{U} = \mathbf{I}_{N_R N_T}$ and the MSE bounds (4.35)-(4.36) depend on the spatial properties of fading, i.e., on the spatial correlation $\mathbf{R}$ as

$$\mathbf{Q}_{\hat{\mathbf{h}}_\ell} = 2f_D T_s \cdot \left(2f_D T_s \mathbf{R}^{-1} + \frac{L_P P}{\sigma_n^2} \mathbf{I}_{N_R N_T} \right)^{-1} \tag{4.41a}$$

$$MSE_{\hat{\mathbf{h}}_\ell} = 2f_D T_S \sum_{i=1}^{N_R N_T} \frac{\mu_i[\mathbf{R}]}{2f_D T_S + \frac{L_P P}{\sigma_n^2}\mu_i[\mathbf{R}]}. \tag{4.41b}$$

For spatially uncorrelated fading at both ends of the link ($\mathbf{R} = \mathbf{I}_{N_R N_T}$) the bounds (4.41a-4.41b) reduce to the results in [17]:

$$\mathbf{Q}_{\hat{\mathbf{h}}_\ell} = \frac{2f_D T_S}{2f_D T_S + \frac{L_P P}{\sigma_n^2}} \mathbf{I}_{N_R N_T} \tag{4.42a}$$

$$MSE_{\hat{\mathbf{h}}_\ell} = \frac{2f_D T_S N_R N_T}{2f_D T_S + \frac{L_P P}{\sigma_n^2}}. \tag{4.42b}$$

Moreover, if the faded amplitudes are uncorrelated across bursts ($f_D T_S = 1/2$) and for large SNR ($L_P P / \sigma_n^2 \gg 1$) the MSE (4.42b) becomes

$$MSE_{\hat{\mathbf{h}}_\ell} = \frac{\sigma_n^2 N_R N_T}{L_P P}, \tag{4.43}$$

which coincides with the (conventional) CRB computed in [6] and with the MSE of the UML estimator (4.37). Therefore, similarly to discussion above, we can conclude that for a frequency-flat channel with spatially and temporally uncorrelated fading amplitudes the UML estimator attains the performance bound.

Let us now turn to the analysis of a frequency-selective channel. As a further simplifying assumption we consider here that all the paths have the same spatial correlation, i.e., $\mathbf{R}_i = \mathbf{R}$. Moreover, we assume that the spatial correlation channel is separable (see Chapter 2)

$$\mathbf{R} = \mathbf{R}^{(T)} \otimes \mathbf{R}^{(R)}. \tag{4.44}$$

Therefore, it is easy to show that $\mathbf{R}_d = \mathbf{\Lambda}_T^2 \otimes \mathbf{R}^{(T)} \otimes \mathbf{R}^{(R)}$, hence $\mu_i[\mathbf{R}_d^{-1}] = 1/[\mathbf{\Lambda}_T]_{ii}^2 \cdot \mu_i[\mathbf{R}_T^{-1}] \cdot \mu_i[\mathbf{R}_R^{-1}]$. It follows that the MSE bound (4.36) becomes:

$$MSE_{\hat{\mathbf{h}}_\ell} = 2f_D \sum_{i=1}^{r_T} \sum_{n=1}^{N_T} \sum_{m=1}^{N_R} \frac{[\mathbf{\Lambda}_T]_{ii}^2 \mu_n[\mathbf{R}_T] \mu_m[\mathbf{R}_R]}{2f_D T_s + \frac{L_P P}{\sigma^2} [\mathbf{\Lambda}_T]_{ii}^2 \mu_n[\mathbf{R}_T] \mu_m[\mathbf{R}_R]}. \tag{4.45}$$

In spite of its complexity, bound (4.45) can be simplified in some useful cases. For low SNR (i.e., $L_P P / \sigma_n^2 \ll 2f_D T_s / ([\mathbf{\Lambda}_T]_{ii}^2 \mu_n[\mathbf{R}_T] \mu_m[\mathbf{R}_R])$) the MSE (4.45) coincides with (4.39) derived for the beamforming scenario. Similarly, for high SNR, we have (4.40). Therefore, we can conclude that *the gain to be expected from the design of a channel estimator based on the knowledge of the channel model as compared to an uncostrained estimator can be quantified by the ratio $N_R N_T W / (2f_D T_s r)$, where for a diversity scenario $r = N_T N_R r_T$ so that the gain reads $W / (2f_D T_s r_T)$.* The gain is thus reduced as compared to a beamforming scenario since the model complexity reduction that was therein accounted for by the ratio $(N_T N_R W / r)$ is here described by W/r. For further details we refer to the discussion above in the context of a beamforming scenario and to Sec. 4.8.

4.7.3 Asymptotic HCRB with different Doppler spectra for different paths

The asymptotic HCRB derived in Sec. 4.7 can be easily generalized to the case where different paths have different Doppler spectra $S_{\varphi,i}(\omega)$, gathered for convenience in the $N_F \times N_F$

diagonal matrices

$$\mathbf{S}_\Phi(\omega) = diag\left[S_{\varphi,1}(\omega)\cdots S_{\varphi,d}(\omega)\right] \qquad \text{(beamforming scenario)}$$

$$\mathbf{S}_\Phi(\omega) = diag\left[S_{\varphi,1}(\omega)\cdots S_{\varphi,d}(\omega)\right] \otimes \mathbf{I}_{N_R N_T} \quad \text{(diversity scenario)}$$

The computation follows the steps of Sec. 4.7 and will not be reported here (see [7]). The HCRB for this case reads

$$\mathbf{Q}_{\hat{\mathbf{h}}_\ell} = \mathbf{U} \int_{-\pi}^{\pi} \left(\mathbf{S}_{dd}(\omega) - \mathbf{S}_{dd}(\omega)\mathbf{F}^H\left(\mathbf{F}\mathbf{S}_{dd}(\omega)\mathbf{F}^H + \mathbf{R}_n \otimes \mathbf{I}_{L_P}\right)^{-1}\mathbf{F}\mathbf{S}_{dd}(\omega)^H\right)\frac{d\omega}{2\pi}\mathbf{U}^H,$$

$$(4.46)$$

with

$$\mathbf{S}_{dd}(\omega) = \mathbf{C}^H \mathbf{S}_\Phi(\omega)\mathbf{C}. \qquad (4.47)$$

Analytical simplifications of the bound (4.46)

The bound (4.46) needs to be evaluated numerically given the Doppler spectra of different paths. However, further analytical insight of (4.46) can be obtained by dividing the integration range into J non-overlapping subbands of support $-\pi < \Delta\omega_j \leq \pi$ so that d_j out of d paths have the related spectrum $S_{\varphi,i}(\omega)$ non zero for ω ranging over the jth subband (see fig. 4.1 for an example with $J = 6$). This operation allows the application of the matrix inversion lemma. Accordingly, the bound (4.46) can be written as

$$\mathbf{Q}_{\hat{\mathbf{h}}_\ell} = \sum_{j=1}^{J} \mathbf{U}_j \int_{-\pi}^{\pi} \left(\mathbf{S}_{dd,j}(\omega)^{-1} + \mathbf{R}_{w,j}\right)^{-1}\frac{d\omega}{2\pi}\mathbf{U}_j^H, \qquad (4.48)$$

with

$$\mathbf{R}_{w,j} = \mathbf{U}_j^H(\mathbf{R}_x \otimes \mathbf{R}_n^{-1})\mathbf{U}_j \qquad (4.49a)$$

$$\mathbf{S}_{dd,j}(\omega) = \mathbf{C}_j^H \mathbf{S}_{\Phi,j}(\omega)\mathbf{C}_j \qquad (4.49b)$$

where the $N_R N_T W \times r_j$ matrix $\mathbf{U}_j$ is defined according to the reduced rank parametrization: $\mathbf{T}_j \boldsymbol{\Phi}_j(0)^{1/2} = \mathbf{U}_j \mathbf{C}_j^H$ with the $N_R N_T W \times d_j$ matrix $\mathbf{T}_j$ containing the colums of $\mathbf{T}$ corresponding to the d_j paths of the jth subband. Similarly, matrices $\mathbf{S}_{\Phi,j}(\omega)$ and $\boldsymbol{\Phi}_j(0)$ contain respectively the spectra and the temporal correlation of the d_j paths of the jth subband. Comparing (4.48)-(4.49) with (4.32)-(4.33), it is clear that the analytical simplifications carried out in Sec. 4.7.2 for uniform Doppler spectrum can be applied to (4.48) as well, within each subband.

4.8 Numerical results: impact of channel and system parameters on the HCRB

Here we consider a MIMO system with time or frequency-domain transmission within a beamforming scenario. The $N_T = 4$ transmitting antennas and $N_R = 4$ receiving antennas belong to
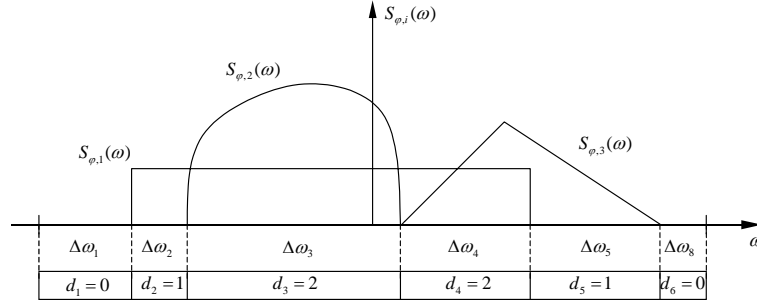
Figure 4.1: Illustration of the procedure employed in order to compute the asymptotic hybric CRB for paths having different Doppler spectra.

half-wavelength spaced linear antenna arrays. The training sequences are optimally designed and the noise is spatially white. The time-varying channel has temporal support $W = 8$ and is characterized by $d = 4$ paths with delays $\boldsymbol{\tau} = [1\ 3\ 4\ 6]T$, DODs and DOAs equally-spaced in the angular support $(-60, 60)\deg$. With these choices, all the paths are resolvable, i.e., the space-time signatures of different paths are linearly independent so that $r = r_{ST} = d = 4$. The length of the training sequences (i.e., size of the training burst in time domain or number of pilot subcarrier in MIMO-OFDM) is $L_P = N_T W = 32$, condition that guarantees the feasibility of UML estimation.

The asymptotic HCRB $MSE_{\hat{\mathbf{h}}_\ell}$ (4.46) is evaluated numerically in order to show the impact of Doppler spectra and power delay profiles on the channel estimation perfomance. The asymptotic HCRB is shown in fig. 4.2 versus the Doppler spread $f_D T_S$ in case the spectra and powers of the four paths are identical (uniform power-delay profile). Uniform (a), Clarke (b) and truncated Gaussian (c) spectrum profiles have been considered (in the latter case $f_D T_S$ is the 3dB cut-off frequency) as shown in the box. We notice that uniform and Clarke spectra lead approximately to the same performance, while the truncated Gaussian spectrum yields a larger channel estimation error, as it is defined over a larger support. For increasing Doppler spread $f_D T_S$, the channel estimator performance degrades as expected from (4.40). Moreover, as expected for $f_D = 0$, the lower bound is zero, showing that in case the channel is static (no Doppler variations), it can be asymptotically estimated with any accuracy.

In order to investigate the effect of different Doppler spectra and powers for different paths, fig. 4.3 shows the asymptotic HCRB versus SNR for four cases: uniform Doppler spectra equal for all paths with $f_D T_S = 0.5$ and uniform/non-uniform power delay profile (see boxes (a)-(b)); uniform Doppler spectra with different shapes for different paths (see box (c)) and uniform/non-uniform power delay profile. The solid curves represent results for uniform power delay profile for the four paths (as in box (a)), while dashed curves are derived when 80% of the power is associated with the path with the smallest Doppler bandwidth (as in box (b)). For
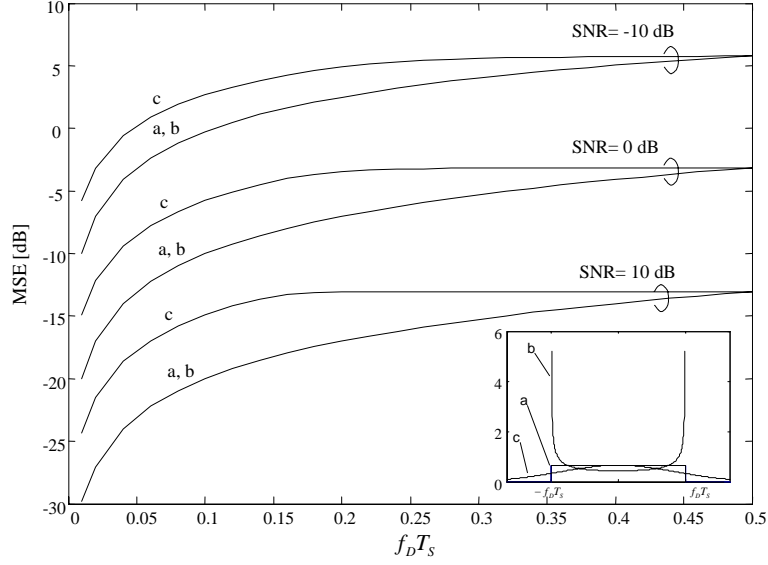
Figure 4.2: Asymptotic hybrid CRB $MSE_{\hat{\mathbf{h}}}$ versus Doppler spread $f_D T_S$ for different Doppler profiles and SNR.

reference, the perfomance of the UML estimator (4.6) is plotted. Introducing the topic of the next Chapter, we can conclude that the bound promises large perfomance improvements over the conventional UML estimator through careful design of the channel estimator. Moreover, as predicted by the analysis in the previous Section, for low SNR the MSE bound tends to the channel norm $E[||\mathbf{h}_\ell||^2] = N_R N_T = 16 = 12dB$ (see (4.39)) whereas for sufficiently large signal to noise ratios the MSE bound is linear in the SNR (see (4.40)). Finally, we remark that smaller a Doppler spread, as for the situation depicted in boc (c), leads to better channel estimation performance.

4.9 Relationship with the conventional CRB

In this Section, the relationship between the HCRB and the conventional CRB is discussed. The conventional CRB applies to the case where the vector to be estimated (in this case the channel vector) depends on deterministic parameters only. Starting from the minimum channel parametrization (4.26), if the amplitudes $\mathbf{d}_\ell$ are modelled as deterministic parameters, a lower bound on the MSE of any unbiased estimator can be obtained by means of the CRB. This reads [12] (compare with (4.22))

$$\mathbf{Q}_{\hat{\mathbf{h}}} = \frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{d}^T]} \cdot \mathbf{J}^\dagger \cdot E_{\mathbf{b}} \left[\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{d}^T]} \right]^H , \qquad (4.50)$$

Figure 4.3: Asymptotic hybrid CRB versus SNR for paths with same/different Doppler spectra and uniform/non uniform power delay profile. As a reference, the MSE of the LS channel estimate is shown.

where $\mathbf{J}$ is the Fisher information matrix of the parameters $\mathbf{u}$ and $\mathbf{d} = \text{vec}\{\mathbf{D}\}$ with $\mathbf{D} = [\mathbf{d}_1 \cdots \mathbf{d}_{N_B}]$. The computation of this bound is carried out in Appendix-C and results in
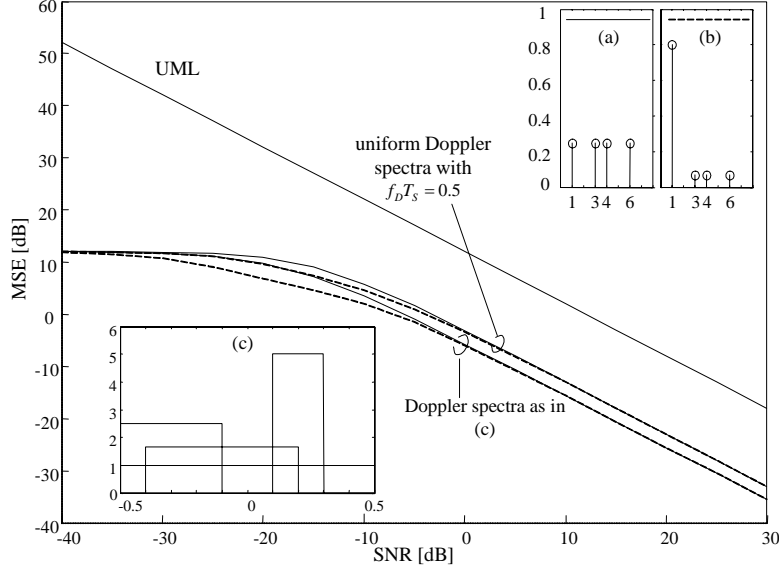
$$\mathbf{Q}_{\hat{\mathbf{h}}} = \left(\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{H/2} \right) \left(\mathbf{\Pi}_{\mathbf{D}^T}^{\perp} \otimes \mathbf{\Pi}_{\tilde{\mathbf{U}}} + \mathbf{\Pi}_{\mathbf{D}^T} \otimes \mathbf{I}_{N_R N_T W} \right) \left(\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{1/2} \right), \quad (4.51)$$

where $\tilde{\mathbf{U}} = \mathbf{Q}_{UML}^{-H/2} \mathbf{U}$. Recall that $\mathbf{\Pi}_{\mathbf{A}}$ for any tall matrix $\mathbf{A}$ is the projection matrix onto the subspace spanned by the columns of $\mathbf{A}$, $span(\mathbf{A})$, i.e., $\mathbf{\Pi}_{\mathbf{A}} = \mathbf{A}\mathbf{A}^{\dagger}$.

Let us now investigate the relationship between the CRB (4.51) and the asymptotic HCRB for uniform Doppler spectra (4.35) by letting $N_B \to \infty$ in (4.51). In this case, $\mathbf{\Pi}_{\mathbf{D}^T}^{\perp} \to \mathbf{I}_{N_B}$ and $\mathbf{\Pi}_{\mathbf{D}^T} \to \mathbf{0}$ since the ratio $r/N_B \to 0$. Accordingly, the CRB becomes

$$\mathbf{Q}_{\hat{\mathbf{h}}} = \mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{H/2} \mathbf{\Pi}_{\tilde{\mathbf{U}}} \mathbf{Q}_{UML}^{1/2}. \quad (4.52)$$

Now, consider the asymptotic HCRB (4.35) and let the signal to noise ratio be high enough (or the spatial correlation of the fading amplitudes small enough for a diversity scenario) so that $2f_D T_s \mathbf{R}_d^{-1} + \mathbf{R}_w \simeq \mathbf{R}_w$ and the fading amplitudes be temporally uncorrelated, $f_D T_S = 1/2$. It can be easily shown that the asymptotic HCRB under these assumptions coincides with the CRB (4.52).

From the discussion above, we can conclude that the CRB coincides with the HCRB when the following conditions are met: *i*) $N_B \to \infty$, *ii*) high SNR, *iii*) temporally (and spatially for a diversity scenario) uncorrelated fading. This result can be easily explained as follows. The conventional CRB deals with the performance of estimators that do not attempt to track

the fading amplitudes since they model the latter as a deterministic parameters. In general, this approach is suboptimal. However, it approaches optimality whenever the statistical information on the fading amplitudes can not be exploited to perform an effective tracking, i.e., when condition *ii*) and *iii*) are met. These considerations will be useful when designing channel estimators, as we will see in the next Chapter.

4.10 SISO/SIMO/MISO vs. MIMO approach to channel estimation

As we will thoroughly discuss in the next Chapter, the analysis of the asymptotic HCRB gives relevant directions on how to design an effective channel estimator. In this Section, we address from a theoretical standpoint the issue of complexity versus channel estimation performance. In particular, we compare three suboptimal and reduced complexity channel estimation strategies to the optimal MIMO approach (see Sec. 4.6). In fact, instead of performing channel estimation in a MIMO system by *jointly* considering all the (frequency-selective) SISO channels corresponding to each pair transmitting-receiving antenna, one could use sub-optimum approaches that estimates separately the SISO or the MISO/SIMO links. These suboptimal approaches are detailed below:

- SISO approach: separate estimation of the $N_R N_T$ SISO channels corresponding to each pair transmitting-receiving antennas;

- MISO approach: joint estimate of all the N_T SISO channels relative to the links between all the transmitting antennas and one receiving antenna (N_R separate channel estimates)

- SIMO approach: joint estimate of all the N_R SISO channels relative to the links between one transmitting antennas and all the receiving antenna (N_T separate channel estimates).

Example 2 *The UML estimator falls within the class of MISO channel estimation since it can be implemented as N_R separate channel estimates, one for each receiving antenna. Moreover, if the training sequences are designed so as to make the signal radiated by each transmitting antenna orthogonal, the UML reduces to a SISO channel estimator since it performs a separate estimate for each pair transmitting/receiving antenna.*

The asymptotic HCRB on the performance of the SISO approach can be easily obtained by evaluating the MSE (4.36) for $N_R = N_T = 1$ for each of the $N_R N_T$ SISO channels $\mathbf{h}_\ell^{(n_R, n_T)} = \mathbf{S}_T \boldsymbol{\beta}_k^{(n_R, n_T)} = \mathbf{U}_{SISO} \mathbf{d}_\ell^{(n_R, n_T)}$ composing the MIMO link. The modes $\mathbf{U}_{SISO}$, coincide with the temporal modes, $\mathbf{U}_{SISO} = \mathbf{U}_T$, defined within a diversity scenario in Sec. 4.4.2, as $\text{span}\{\mathbf{U}_{SISO}\} = \text{span}\{\mathbf{S}_T\}$. Moreover, $r_{SISO} = \text{rank}\{\mathbf{S}_T\} = r_T \leq W$ coincides with the number of temporally resolvable paths. If all the SISO channels share the same characteristics (as for transmitting and receiving antennas not too far apart), the MSE bound for the SISO

approach can be derived by specializing the asymptotic HCRB (4.36) for $N_R = N_T = 1$ (with obvious notation):

$$MSE_{SISO} = N_R N_T \cdot MSE_{MIMO}|_{N_R=1, N_T=1}. \tag{4.53}$$

According to the Sec. 4.7.2, for high SNR and spatially white noise the MSE (4.53) becomes

$$MSE_{SISO} = N_R N_T \cdot 2 f_D T_S \frac{\sigma_n^2}{L_P P} r_T \tag{4.54}$$

for both beamforming or diversity scenarios. For high SNR the comparison between (4.54) and (4.40) shows that for beamforming scenario the *joint* MIMO approach outperforms the SISO approach by

$$\frac{MSE_{SISO}}{MSE_{MIMO}} = \frac{N_R N_T r_T}{r_{MIMO}} \tag{4.55}$$

which is greater than one since in general $r_{MIMO} \ll N_R N_T W$. On the other hand, for the diversity scenario the SISO approach shows no degradation compared to the MIMO approach (recall (4.18)).

The asymptotic HCRB on the performance of the SIMO (or dually MISO) approach can be similarly obtained by evaluating the MSE (4.36) for $N_T = 1$ for each of the N_T SIMO channels $\mathbf{h}_\ell^{(n_T)} = \mathbf{U}_{SIMO} \mathbf{d}_\ell^{(n_T)}$ composing the MIMO link:

$$MSE_{SIMO} = N_T \cdot MSE_{MIMO}|_{N_T=1}. \tag{4.56}$$

Since the space-time modes are now collected by columns in matrix $\mathbf{U}_{SIMO}$ with $r_{SIMO} = \text{rank}\{\mathbf{U}_{SIMO}\} \leq N_R W$, the performance comparison reads

$$\frac{MSE_{SIMO}}{MSE_{MIMO}} = \frac{N_T \cdot r_{SIMO}}{r_{MIMO}} \tag{4.57}$$

that is larger than one for a beamforming scenario (MISO approach yields a similar conclusion). Under the said assumptions, the degree of improvement of MIMO approach to channel estimation compared to SISO or SIMO (or MISO) approaches depends on the number of temporal or space-time modes, closely related to the specific model. The number of modes (r_{SISO} or r_{SIMO}), i.e., the number of resolvable paths, increases when the multipath environment is dense in time and/or space. For a number of (well resolved in space and/or time) paths d large enough, we have $r_{MIMO} \simeq N_R N_T W$, $r_{SISO} \simeq W$ and $r_{SIMO} \simeq N_R W$ so that there is no practical advantage in estimating jointly the MIMO channel. These conclusions are validated numerically in the following Subsection.

4.10.1   Numerical examples: SISO/ MISO/ SIMO vs. MIMO channel estimation

To compare the suboptimal channel estimation strategies described in the previous Section, we study in fig. 4.4 the MSE degradation $MSE_{\text{degradation}} = (MSE - MSE_{\hat{\mathbf{h}}})/MSE_{\hat{\mathbf{h}}}$ of
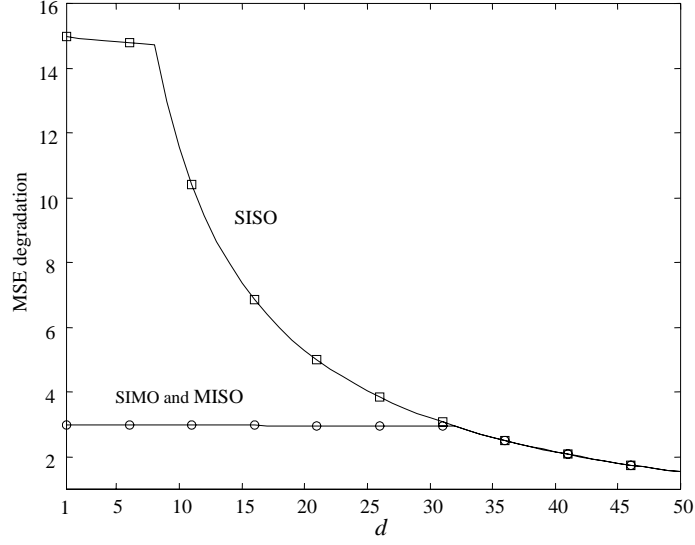
Figure 4.4: MSE degradation of the SIMO, MISO and SISO approaches compared to the MIMO estimation versus the number of paths d for the beamforming model.

SISO, MISO or SIMO channel estimation approaches compared to the more general MIMO approach, for the beamforming scenario and varying number of paths d. The channel and system parameters are selected according to the example in Sec. 4.8. In particular, the Doppler spectra of all paths and power delay profile are assumed to be uniform. The normalized Doppler shift is set to $f_D T_S = 0.1$ and $SNR = 10dB$.

As previously stated, the benefits of using a MIMO channel estimator compared to SISO, MISO or SIMO become smaller for increasing number of paths d when $d \geq W = 8$ for the SISO approach, $d \geq N_T W = 32$ for the MISO and $d \geq N_R W = 32$ for the SIMO approach. In addition, for $d \geq \max(N_R W, N_T W) = 32$ the SISO, MISO and SIMO approaches have the same performance. These results can be easily justified by recalling the discussion above. For instance, from (4.55) we have that $MSE_{SISO}/MSE_{\hat{\mathbf{h}}} \simeq N_R N_T \cdot \min(d, W)/d$, where the last (approximate) equality stems from the direct proportionality of the rank of $\mathbf{T}$ (or $\mathbf{U}$) and the number of paths d, as long as the space-time signatures of the d paths are linearly independent. It follows that for $d \geq W = 8$ the degradation of the SISO approach decreases as $1/d$, as confirmed by fig. 4.4. Similar considerations can be used to prove analogous conclusions for SIMO and MISO approaches.

4.11 Extension to multiuser systems

The extension of the HCRB to a multiuser setting, according to Sec. 3.4, is straightforward. In fact, let us at first consider the uplink. Recalling that the signal model (3.32),

it is enough to define a compound training matrix $\mathbf{X} = [\mathbf{X}^{(1)} \cdots \mathbf{X}^{(K)}]$ and channel vector $\mathbf{h}_\ell = [\mathbf{h}_\ell^{(1)T} \cdots \mathbf{h}_\ell^{(K)T}]^T$ in order to be able to apply all the results derived above for this setting. On the other hand, for the downlink, the training sequences are sent in broadcast to all users and according to the signal model (3.35), the analysis above can be applied separately for each user.

4.12   Conclusion

The HCRB for the estimation of frequency-selective MIMO channels for both a beamforming and a diversity scenario has been derived. As a by-product, this analysis led to the definition of an (asymptotically) optimal channel estimation strategy, based on the separate computation of long term and fast-varying channel features. In particular, the optimal strategy, to be investigated in the following Chapter, prescribes the consistent estimate of the long term features of the channel and the MMSE tracking of the fading amplitudes.

The properties of the HCRB have been thoroughly investigated in the asymptotic regime (i.e., for a large interval of stationarity of the long term features of the channel) from both an analytical standpoint and through numerical simulations. This study provided insight into the effects of system and channel parameters on the channel estimation error. Finally, the relationship with the conventional CRB has been discussed.

4.13   Appendix-A: proof of HCRB (4.22)

Let the unknown parameter vector to be estimated $\mathbf{h} = \mathbf{h}(\boldsymbol{\theta})$ depend on a vector of parameters $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1^T & \boldsymbol{\theta}_2^T \end{bmatrix}^T$ that has a deterministic part ($\boldsymbol{\theta}_1$) and a random part ($\boldsymbol{\theta}_2$) with probability density function (pdf) $p(\boldsymbol{\theta}_2)$. The observation vector $\mathbf{z}$ (i.e., any sufficient statistics for the estimation of [13] has pdf (conditioned on $\boldsymbol{\theta}$) $p(\mathbf{z}|\boldsymbol{\theta})$. Under appropriate assumptions of regularity of the pdf involved [12], the following equality holds

$$\iint p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}_2)d\mathbf{z}d\boldsymbol{\theta}_2 = 1, \tag{4.58}$$

can be differentiated, obtaining

$$\begin{aligned} 0 &= \iint \frac{\partial p(\mathbf{z}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}p(\boldsymbol{\theta}_2)d\mathbf{z}d\boldsymbol{\theta}_2 + \iint p(\mathbf{z}|\boldsymbol{\theta})\frac{\partial p(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}}d\mathbf{z}d\boldsymbol{\theta}_2 = \\ &= \iint \frac{\partial \ln p(\mathbf{z}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}_2)d\mathbf{z}d\boldsymbol{\theta}_2 + \iint \frac{\partial \ln p(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}}p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}_2)d\mathbf{z}d\boldsymbol{\theta}_2 \\ &= E[\boldsymbol{\Delta}_D] + E[\boldsymbol{\Delta}_P], \tag{4.59} \end{aligned}$$

where the expectation operator $E[\cdot]$ is defined with respect to the pdf $p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}_2) = p(\mathbf{z}, \boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$ and

$$\boldsymbol{\Delta}_D = \frac{\partial \ln p(\mathbf{z}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{4.60a}$$

$$\boldsymbol{\Delta}_P = \frac{\partial \ln p(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{0} \\ \frac{\partial \ln p(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} \end{bmatrix}. \tag{4.60b}$$

We also define $\boldsymbol{\Delta} = \boldsymbol{\Delta}_D + \boldsymbol{\Delta}_P$.

By differentiating the equality $\int p(\mathbf{z}|\boldsymbol{\theta})d\mathbf{z} = 1$, we can similarly obtain

$$\mathbf{0} = E[\boldsymbol{\Delta}_D|\boldsymbol{\theta}] = E[\boldsymbol{\Delta}_D], \tag{4.61}$$

that because of (4.59) implies

$$E[\boldsymbol{\Delta}_P] = \mathbf{0}. \tag{4.62}$$

Now, recalling that we are considering an unbiased estimate $\tilde{\mathbf{h}}$ of $\mathbf{h}$, we get

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}^H} = \frac{\partial E[\tilde{\mathbf{h}}|\boldsymbol{\theta}]}{\partial \boldsymbol{\theta}^H} = \frac{\partial}{\partial \boldsymbol{\theta}^H} \int \tilde{\mathbf{h}} p(\mathbf{z}|\boldsymbol{\theta})d\mathbf{z} = \int \tilde{\mathbf{h}} \frac{\partial \ln p(\mathbf{z}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^H} p(\mathbf{z}|\boldsymbol{\theta})d\mathbf{z} = E[\tilde{\mathbf{h}}\boldsymbol{\Delta}_D^H|\boldsymbol{\theta}]. \tag{4.63}$$

Then, defining $\mathbf{P} = E_{\boldsymbol{\theta}_2}\left[\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}^H}\right]$, we obtain

$$\mathbf{P} = E[\hat{\mathbf{h}}\boldsymbol{\Delta}_D^H] = E[(\tilde{\mathbf{h}} - \mathbf{h})\boldsymbol{\Delta}_D^H] = E[(\tilde{\mathbf{h}} - \mathbf{h})\boldsymbol{\Delta}^H], \tag{4.64}$$

where we used (4.61) and the fact that $E[(\tilde{\mathbf{h}} - \mathbf{h})\boldsymbol{\Delta}_P^H] = \mathbf{0}$, easily proved as follows

$$E[(\tilde{\mathbf{h}} - \mathbf{h})\boldsymbol{\Delta}_P^H|\boldsymbol{\theta}] = E[\tilde{\mathbf{h}} - \mathbf{h}|\boldsymbol{\theta}]\boldsymbol{\Delta}_P^H = \mathbf{0}. \tag{4.65}$$

We define the following matrix

$$E\left[\begin{bmatrix} \tilde{\mathbf{h}} - \mathbf{h} \\ \boldsymbol{\Delta} \end{bmatrix} \begin{bmatrix} (\tilde{\mathbf{h}} - \mathbf{h})^H & \boldsymbol{\Delta}^H \end{bmatrix}\right] = \begin{bmatrix} \mathbf{Q}_{\tilde{\mathbf{h}}} & \mathbf{D} \\ \mathbf{D}^H & \mathbf{J}_B \end{bmatrix} \tag{4.66}$$

with

$$\mathbf{J}_B = E[\boldsymbol{\Delta}\boldsymbol{\Delta}^H] = E[\boldsymbol{\Delta}_D\boldsymbol{\Delta}_D^H] + E[\boldsymbol{\Delta}_P\boldsymbol{\Delta}_P^H] = \mathbf{J}_D + \mathbf{J}_P, \tag{4.67}$$

for the second equality the relationship $E[\boldsymbol{\Delta}_D\boldsymbol{\Delta}_P^H] = \mathbf{0}$ has been used (proof is straightforward). The semi-positiveness of matrix (4.66) implies that [12]

$$\mathbf{Q}_{\tilde{\mathbf{h}}} \geq \mathbf{P}\mathbf{J}_B^{\dagger}\mathbf{P}^H, \tag{4.68}$$

proving the HCRB (4.21)-(4.22)

4.14   Appendix-B: computation of the HCRB (4.23)

From (4.1) and (4.15), the likelihood function for the estimation of parameters $\mathbf{U}$, $\mathbf{C}$ and $\mathbf{B}$ is (neglecting uninteresting constants)

$$
\begin{aligned}
\mathcal{L}(\mathbf{U}, \mathbf{C}, \mathbf{B}) &= \sum_{\ell=1}^{N_B} \left\| \mathbf{y}_\ell - \mathbf{X}\mathbf{U}\mathbf{C}^H \mathbf{b}_\ell \right\|_{\mathbf{R}_n^{-1}}^2 = \\
&= \left\| \mathbf{H}_{UML} - \mathbf{U}\mathbf{C}^H \mathbf{B} \right\|_{\mathbf{Q}_{UML}^{-1}}^2 ,
\end{aligned} \tag{4.69}
$$

where the second equality can be easily proved by using the definition of the $N_R N_T W \times N_B$ matrix that gathers the UML estimates (4.5) as $\mathbf{H}_{UML} = [\mathbf{h}_{UML,1} \cdots \mathbf{h}_{UML,N_B}]$. In other words, according to (4.69) the N_B UML estimates $\{\mathbf{h}_{UML,\ell}\}_{\ell=1}^{N_B}$ are sufficient statistics for the estimation of $\{\mathbf{h}_\ell\}_{\ell=1}^{N_B}$ [13]. Moreover, by introducing for the ensemble of N_B UML estimates the vector $\mathbf{h}_{UML} = \mathrm{vec}(\mathbf{H}_{UML})$, the likelihood function (4.69) becomes

$$
\mathcal{L}(\mathbf{U}, \mathbf{C}, \mathbf{B}) = \| \mathbf{h}_{UML} - (\mathbf{I}_{N_B} \otimes \mathbf{U}\mathbf{C}^H)\mathbf{b} \|_{\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{-1}}^2 . \tag{4.70}
$$

The HCRB for the estimation of $\mathbf{h}$ reads (4.22), where $\mathbf{J}_B$ is the (Bayesian) Fisher Information Matrix. According to [11] (see also Appendix-B), matrix $\mathbf{J}_B$ can be written as the sum of a term accounting for the information due to data $\mathbf{J}_D$ and a term accounting for prior knowledge $\mathbf{J}_P$

$$
\mathbf{J}_B = \mathbf{J}_D + \mathbf{J}_P . \tag{4.71}
$$

Notice that in our framework, the prior knowledge consists in the statistical properties of the randomly varying parameters $\mathbf{b}$. Since the observation model is Gaussian, we have [11]

$$
\mathbf{J}_D = E_{\mathbf{b}} \left[\left(\frac{\partial E[\mathbf{h}_{UML}|\mathbf{b}]}{\partial [\mathbf{u}^T \mathbf{c}^T \mathbf{b}^T]} \right)^H (\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{-1}) \left(\frac{\partial E[\mathbf{h}_{UML}|\mathbf{b}]}{\partial [\mathbf{u}^T \mathbf{c}^T \mathbf{b}^T]} \right) \right] , \tag{4.72}
$$

where $E_{\mathbf{b}}[\cdot]$ denotes the ensemble average with respect to the distribution of $\mathbf{b}$. The UML estimate is unbiased, $E[\mathbf{h}_{UML}|\mathbf{b}] = \mathbf{h}$, and

$$
\frac{\partial \mathbf{h}}{\partial \mathbf{u}^T} = \mathbf{B}^T \mathbf{C}^* \otimes \mathbf{I}_{N_R} \tag{4.73a}
$$

$$
\frac{\partial \mathbf{h}}{\partial \mathbf{c}^T} = \mathbf{B}^T \otimes \mathbf{U} \tag{4.73b}
$$

$$
\frac{\partial \mathbf{h}}{\partial \mathbf{b}^T} = \mathbf{I}_{N_B} \otimes \mathbf{U}\mathbf{C}^H \tag{4.73c}
$$

$$
\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{c}^T \mathbf{b}^T]} = \left[\begin{array}{ccc} \mathbf{B}^T \mathbf{C}^* \otimes \mathbf{I}_{N_R} & \mathbf{B}^T \otimes \mathbf{U} & \mathbf{I}_{N_B} \otimes \mathbf{U}\mathbf{C}^H \end{array} \right] . \tag{4.73d}
$$

It follows that

$$
\begin{aligned}
\mathbf{J}_D &= E_{\mathbf{b}} \left[\begin{array}{ccc}
\mathbf{C}^T \mathbf{B}^* \mathbf{B}^T \mathbf{C}^* \otimes \mathbf{Q}_{UML}^{-1} & \mathbf{C}^T \mathbf{B}^* \mathbf{B}^T \otimes \mathbf{Q}_{UML}^{-1} \mathbf{U} & \mathbf{C}^T \mathbf{B}^* \otimes \mathbf{Q}_{UML}^{-1} \mathbf{U}\mathbf{C}^H \\
\mathbf{B}^* \mathbf{B}^T \mathbf{C}^* \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} & \mathbf{B}^* \mathbf{B}^T \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U} & \mathbf{B}^* \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U}\mathbf{C}^H \\
\mathbf{B}^T \mathbf{C}^* \otimes \mathbf{C}\mathbf{U}^H \mathbf{Q}_{UML}^{-1} & \mathbf{B}^T \otimes \mathbf{C}\mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U} & \mathbf{I}_{N_B} \otimes \mathbf{C}\mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U}\mathbf{C}^H
\end{array} \right] = \\
&= \left[\begin{array}{ccc}
N_B \mathbf{C}^T \mathbf{C}^* \otimes \mathbf{Q}_{UML}^{-1} & N_B \mathbf{C}^T \otimes \mathbf{Q}_{UML}^{-1} \mathbf{U} & \mathbf{0} \\
N_B \mathbf{C}^* \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} & N_B \mathbf{I}_{N_R} \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}_{N_B} \otimes \mathbf{C}\mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U}\mathbf{C}^H
\end{array} \right] .
\end{aligned} \tag{4.74}
$$

Since from (4.11) $\mathbf{b} \backsim \mathcal{CN}(\mathbf{0}, \mathbf{R}_t \otimes \mathbf{I}_{N_R})$, the information matrix related to prior information becomes [11]

$$\mathbf{J}_P = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_t^{-1} \otimes \mathbf{I}_{N_R} \end{bmatrix}. \tag{4.75}$$

Non-singularity of $\mathbf{R}_t$ is assumed. Finally, the (Bayesian) Fisher Information Matrix is

$$\mathbf{J}_B = \begin{bmatrix} \mathbf{J}_a & \mathbf{J}_{ac} & \mathbf{0} \\ \mathbf{J}_{ca} & \mathbf{J}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_b \end{bmatrix} = \tag{4.76}$$

$$\begin{bmatrix} N_B \mathbf{C}^T \mathbf{C}^* \otimes \mathbf{Q}_{UML}^{-1} & N_B \mathbf{C}^T \otimes \mathbf{Q}_{UML}^{-1} \mathbf{U} & \mathbf{0} \\ N_B \mathbf{C}^* \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} & N_B \mathbf{I}_{N_R} \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_B} \otimes \mathbf{C} \mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U} \mathbf{C}^H + \mathbf{R}_t^{-1} \otimes \mathbf{I}_{N_R} \end{bmatrix}.$$

The non-uniqueness of the factorization (4.15) is accounted for by the rank-deficiency of the corresponding Fisher Information Matrix [12]:

$$\begin{aligned} rank\left\{ \begin{bmatrix} \mathbf{J}_a & \mathbf{J}_{ac} \\ \mathbf{J}_{ca} & \mathbf{J}_c \end{bmatrix} \right\} &= rank(\mathbf{J}_a) + rank(\mathbf{J}_c - \mathbf{J}_{ca} \mathbf{J}_a^{-1} \mathbf{J}_{ac}) = \\ &= rN_T + rank((\mathbf{I}_{N_R} - \mathbf{C}^*(\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T) \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U}) = \\ &= r(N_T + N_R - r). \end{aligned} \tag{4.77}$$

Averaging over the distribution of the amplitudes, we have

$$E_{\mathbf{b}} \left[\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{c}^T \mathbf{b}^T]} \right] = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_B} \otimes \mathbf{U} \mathbf{C}^H \end{bmatrix} \tag{4.78}$$

and using (4.76) in (4.22) the CRB (4.21) becomes

$$\mathbf{Q}_{\hat{\mathbf{h}}} \geq (\mathbf{I}_{N_B} \otimes \mathbf{U} \mathbf{C}^H) \mathbf{J}_b^{-1} (\mathbf{I}_{N_B} \otimes \mathbf{C} \mathbf{U}^H) \tag{4.79a}$$

that coincides with (4.23) as it can be easily shown by applying the matrix inversion lemma.

4.15 Appendix-C: computation of the CRB (4.51)

From (4.1) and (4.26), the negative log-likelihood function can be written as

$$\mathcal{L}(\mathbf{U}, \mathbf{D}) = \|\mathbf{H}_{UML} - \mathbf{U}\mathbf{D}\|_{\mathbf{Q}_{UML}^{-1}}^2, \tag{4.80}$$

The derivative of the channel vector $\mathbf{h}$ in (4.50) can be calculated as

$$\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{d}^T]} = \begin{bmatrix} \mathbf{D}^T \otimes \mathbf{I}_{N_R N_T W} & \mathbf{I}_{N_B} \otimes \mathbf{U} \end{bmatrix}. \tag{4.81}$$

The $r(N_R N_T W + N_B) \times r(N_R N_T W + N_B)$ Fisher Information Matrix of the parameters $\{\mathbf{u}, \mathbf{d}\}$ [11]

$$\mathbf{J} = \left[\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{d}^T]} \right]^H (\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{-1}) \left[\frac{\partial \mathbf{h}}{\partial [\mathbf{u}^T \mathbf{d}^T]} \right] \tag{4.82}$$

and can be partitioned as

$$\mathbf{J} = \left[\begin{array}{cc} \mathbf{J}_{uu} & \mathbf{J}_{ud} \\ \mathbf{J}_{ud}^{H} & \mathbf{J}_{dd} \end{array} \right] \tag{4.83}$$

where the $N_R N_T W r \times N_R N_T W r$ block $\mathbf{J}_{uu}$ and the $N_B r \times N_B r$ block $\mathbf{J}_{dd}$ depend on the space-time modes and the amplitudes, respectively. The diagonal blocks are obtained from (4.82):

$$\mathbf{J}_{uu} = \mathbf{D}^T \mathbf{D}^* \otimes \mathbf{Q}_{UML}^{-1} \tag{4.84a}$$

$$\mathbf{J}_{dd} = \mathbf{I}_{N_B} \otimes \mathbf{U}^H \mathbf{Q}_{UML}^{-1} \mathbf{U}. \tag{4.84b}$$

Note that the matrix $\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{-1}$ in (4.82) is positive definite so that, by using the standard result on the rank of a partitioned matrix [18], it can be shown that $\text{rank}\{\mathbf{J}\} = \text{rank}\{\mathbf{D}\} = r(N_R N_T W + N_B) - r^2$. The rank order of $\mathbf{J}$ can be also derived as the number of degrees of freedom in the channel model. Indeed, recalling that in N_B bursts the channel parametrization can be rewritten as $\mathbf{H} = \mathbf{UD}$ with $\mathbf{U}$ and $\mathbf{D}$ full rank matrices, it follows the number of degrees of freedom from the singular value decomposition of matrix $\mathbf{H}$: $[r N_R N_T W - r(r+1)/2]$ (left eigenvectors) $+ [r N_B - r(r+1)/2]$ (right eigenvectors) $+ [r]$ (eigenvalues) $= \text{rank}\{\mathbf{J}\}$.

As $\text{rank}\{\mathbf{J}\} < r(N_R N_T W + N_B)$ the Fisher Information Matrix is singular. Next, by defining the matrices $\mathbf{P}_1 = \mathbf{D}^T \otimes \mathbf{I}_{N_R N_T W}$, $\mathbf{P}_2 = \mathbf{I}_{N_B} \otimes \mathbf{U}$, and $\mathbf{P} = [\mathbf{P}_1 \; \mathbf{P}_2]$, the CRB (4.50) can be written as

$$\begin{aligned} \mathbf{Q}_{\hat{\mathbf{h}}} &= \mathbf{P} \left(\widetilde{\mathbf{P}}^H \widetilde{\mathbf{P}} \right)^{\dagger} \mathbf{P}^H \\ &= \left(\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{H/2} \right) \mathbf{\Pi}_{\widetilde{\mathbf{P}}} \left(\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{1/2} \right), \end{aligned} \tag{4.85}$$

$\widetilde{\mathbf{P}} = (\mathbf{I}_{N_B} \otimes \mathbf{Q}_{UML}^{-H/2})\mathbf{P}$ and $\mathbf{\Pi}_{\widetilde{\mathbf{P}}} = \widetilde{\mathbf{P}}(\widetilde{\mathbf{P}}^H \widetilde{\mathbf{P}})^{\dagger} \widetilde{\mathbf{P}}^H$ is the corresponding projector. Since $span\{\widetilde{\mathbf{P}}\} = span\{\widetilde{\mathbf{P}}_1\} \cup span\{\widetilde{\mathbf{P}}_2\}$, this can be equivalently decomposed into the orthogonal subspaces $span\{\widetilde{\mathbf{P}}\} = span\{\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp} \widetilde{\mathbf{P}}_2\} \cup span\{\widetilde{\mathbf{P}}_1\}$ such that $span\{\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp} \widetilde{\mathbf{P}}_2\} \cap span\{\widetilde{\mathbf{P}}_1\} = 0$. The the projection matrix reduces to

$$\mathbf{\Pi}_{\widetilde{\mathbf{P}}} = \mathbf{\Pi}_{\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp} \widetilde{\mathbf{P}}_2} + \mathbf{\Pi}_{\widetilde{\mathbf{P}}_1} \tag{4.86}$$

where $\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}, \mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp}$ and $\mathbf{\Pi}_{\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp} \widetilde{\mathbf{P}}_2}$ denote the orthogonal projections onto, respectively, $span\{\widetilde{\mathbf{P}}_1\}$, the orthogonal complement $span\{\widetilde{\mathbf{P}}_1^{\perp}\}$ and $span\{\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp} \widetilde{\mathbf{P}}_2\}$. According to the model exploited here, the projectors in (4.86) can be easily calculated as it follows:

$$\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1} = \mathbf{\Pi}_{\mathbf{D}^T} \otimes \mathbf{I}_{N_R N_T W}, \tag{4.87}$$

$$\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp} \widetilde{\mathbf{P}}_2 = \left(\mathbf{\Pi}_{\mathbf{D}^T}^{\perp} \otimes \mathbf{I}_{N_R N_T W} \right) \left(\mathbf{I}_{N_B} \otimes \widetilde{\mathbf{U}} \right) = \mathbf{\Pi}_{\mathbf{D}^T}^{\perp} \otimes \widetilde{\mathbf{U}}, \tag{4.88}$$

$$\mathbf{\Pi}_{\mathbf{\Pi}_{\widetilde{\mathbf{P}}_1}^{\perp} \widetilde{\mathbf{P}}_2} = \mathbf{\Pi}_{\mathbf{D}^T}^{\perp} \otimes \mathbf{\Pi}_{\widetilde{\mathbf{U}}}. \tag{4.89}$$

The CRB (4.85) can be now evaluated as (4.51).

Bibliography

[1] B. Hassibi, B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-963, April 2003.

[2] H. Holma and A. Toskala, *WCDMA for UMTS,* John Wiley & Sons, 2000.

[3] IEEE Std 802.11a-1999, "Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5Ghz Band".

[4] Q. Sun, D. C. Cox, H. C. Huang and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. on Wireless Comm.*, vol. 1, pp. 549-553, Oct. 2002.

[5] I. Barhumi, G. Leus and M. Moonen, "Optimal training design for MIMO in mobile wireless channels", *IEEE Trans. Signal Proc.*, vol. 51, no .6, pp.1615-1624, June 2003.

[6] T. L. Marzetta, "Blast training: estimating channel characteristics for high capacity space-time wireless," *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 958–966, September 1999.

[7] M. Cicerone, N. Geng, O. Simeone and U. Spagnolini, "Modal analysis/filtering to estimate time-varying MIMO-OFDM channels," in *Proc. ITG Workshop on Smart Antennas*, March 2004.

[8] T. Soderstrom and P. Stoica, *System identification*, Prentice Hall, 1989.

[9] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions", *IEEE Trans. Signal Processing,* vol. 44, no. 12, pp. 3069-3078, Dec. 1996.

[10] J. Picheral and U.Spagnolini, "Angles and delay estimation of space-time channels for multiuser detection," *IEEE Trans. on Wireless Communications*, vol. 3, no. 3, pp. 758-769, May 2004.

[11] H. L. Van Trees, *Optimum array processing*, J. Wiley Ed. 2002.

[12] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices", *IEEE Trans. Signal Processing*, vol. 49, no.1, pp. 87-90, Jan. 2001.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hill, 1993.

[14] J.-T. Chen, Member, Joonsuk Kim and Jen-Wei Liang, "Multichannel MLSE Equalizer with Parametric FIR Channel Identification," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1923-1935, Nov. 1999.

[15] Y. Li, "Optimum Training sequences for OFDM systems with multiple transmit antennas", in *Proc. IEEE GLOBECOM'00*, vol.3, pp. 1478-1482, 2000.

[16] V. K. Jones and G. G. Raileigh, "Channel estimation for wireless OFDM systems," in *Proc. IEEE GLOBECOM'98,* pp. 980-985.

[17] J. Baltersee, G. Flock and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2358-2368, Dec. 2001.

[18] G.H. Golub and C. F. Van Loan, *Matrix computations*, The Johns Hopkins University Press, 1996.

Chapter 5

Channel estimation by modal analysis/filtering

5.1 Introduction

IN the previous Chapter, investigation of the HCRB led to the derivation of an (asymptotically) optimal channel estimation strategy (see Sec. 4.6), which was proved to promise remarkable performance improvement as compared to conventional unstructured estimators, such as UML (see fig. 4.3). The optimal channel estimation strategy can be described by the following guidelines: *a)* long and short term channel parameters have to be estimated separately, *b)* long term parameters have to be estimated consistently, *c)* short term parameters (i.e., fast-varying amplitudes) have to be tracked according to the MMSE criterion. Notice that many known estimators proposed in the literature under simplified settings have (at least one of) the aforementioned properties. For instance, in a diversity scenario, Kalman filtering has been recently proposed for MMSE tracking of the fading amplitudes by assuming known sample-spaced delays (or equivalently $\mathbf{T}$ is diagonal and known) [1]. Additional references to existing work are discussed in the following.

In this Chapter, a novel class of channel estimators based on the optimal strategy described above is proposed. Our approach considers the parametrization of the channel vector $\mathbf{h}_\ell$ (see Sec. 4.4) in terms of the *space-time modes* $\mathbf{U}$ (for the beamforming scenario) or *temporal modes* $\mathbf{U}_T$ (for the diversity scenario):

$$\mathbf{h}_\ell = \mathbf{U}\mathbf{d}_\ell, \tag{5.1}$$

where for the diversity scenario, we have $\mathbf{U} = \mathbf{U}_T \otimes \mathbf{I}_{N_R N_T}$. For reference, it is useful to recall that the space-time modes (i.e., the columns of matrix $\mathbf{U}$) are defined as the $N_R N_T W \times 1$ vectors that span the long term subspace $range\{\mathbf{T}\} = range\{\mathbf{S}_{ST}\}$ (see Sec. 4.4.1). Similarly, the temporal modes (i.e., the columns of matrix $\mathbf{U}_T$) are defined as the $W \times 1$ vectors spanning the long term subspace $range\{\mathbf{S}_T\}$ (see Sec. 4.4.2).

```
┌─────────────────────────────────────────────┐
│ Algebraic channel model:  𝐡_ℓ = 𝐔𝐝_ℓ        │
└─────────────────────────────────────────────┘
                      │
                      ▼
        ┌───────────────────────────────┐
        │  Channel estimation strategy  │
        └───────────────────────────────┘
           ╱                      ╲
```

Space-time modal analysis   *Fast-fading tracking*
(estimate of **U**)   (estimate of $\mathbf{d}_\ell$)

• batch   • modal filtering   (Sec. 5.3.1)
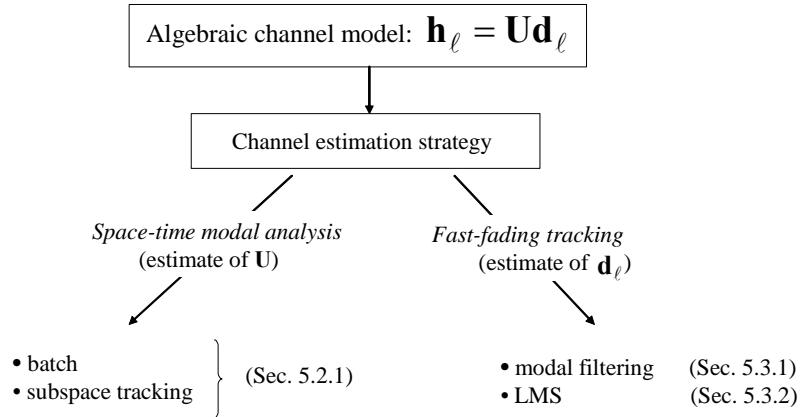• subspace tracking   (Sec. 5.2.1)   • LMS   (Sec. 5.3.2)

Figure 5.1: Taxonomy of channel estimators based on space-time modal analysis and tracking of the fast-fading amplitudes.

An illustration of the taxonomy of the proposed channel estimators is shown in fig. 5.1. The estimation of the long term features of the channel, i.e., of the stationary matrix $\mathbf{U}$, is disclosed in Sec. 5.2, where it is proposed both a batch approach and an adaptive approach based on subspace tracking. The latter method allows to alleviate the assumption of quasi-static variations of the long term features of the channel, allowing continuous (but still slow) fluctuations. The estimation of matrix $\mathbf{U}$ will be referred to as *space-time modal analysis*.

On the other hand, estimation of the fading amplitudes $\mathbf{d}_\ell$ is discussed in Sec. 5.3. At first, in Sec. 5.3.1 a block-by-block estimation that allows to set the channel estimation problem in a mathematically convenient ML framework is discussed. This first approach is referred to as *modal filtering*, since the corresponding channel estimator turns out to consist essentially in the projection of a preliminary UML estimate (recall Sec. 4.2) into the subspace spanned by the space-time modes of the channel. Then, tracking of the fading amplitudes based on the LMS algorithm is proposed in Sec. 5.3.2.

Moreover, in Sec. 5.2, it is also shown that, for a beamforming scenario, direct estimation (or tracking) of matrix $\mathbf{U}$ may have poor performance due to the large number of parameters to be estimated. In this case, an asymptotically suboptimum parameterization of the space-time modes into decoupled spatial and temporal modes is proposed. In particular, the parametrization $\mathbf{U} = \mathbf{U}_T^* \otimes \mathbf{U}_S$, where the matrix $\mathbf{U}_S$ spans the long term subspace spanned by the spatial signatures of multipaths, is advocated in Sec. 5.2.2. Accordingly, therein, an estimation algorithm based on the decoupled spatial and temporal that works in both batch and adaptive mode through subspace tracking is proposed (*decoupled spatial and temporal modal analysis*). The corresponding taxonomy of estimators based on this reduced-complexity approach is shown in fig. 5.2.

An analytical study of the performance of the proposed channel estimators is carried out

$$\boxed{\text{Reduced-complexity algebraic channel model: } \mathbf{h}_\ell = (\mathbf{U}_T^* \otimes \mathbf{U}_S)\mathbf{d}_\ell}$$

$$\boxed{\text{Channel estimation strategy}}$$

Decoupled spatial and temporal modal analysis
(estimate of $\mathbf{U}_T$ and $\mathbf{U}_S$)

Fast-fading tracking
(estimate of $\mathbf{d}_\ell$)

- batch
- subspace tracking   (Sec. 5.2.2)

- modal filtering   (Sec. 5.3.1)
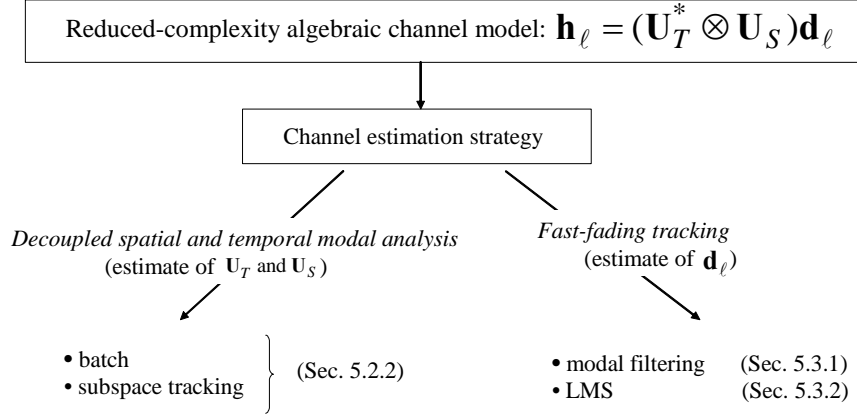- LMS   (Sec. 5.3.2)

Figure 5.2: Taxonomy of channel estimators based on decoupled spatial and temporal modal analysis and tracking of the fast-fading amplitudes.

in Sec. 5.5. This analysis is corroborated by an extensive numerical investigation for different propagation environment in Sec. 5.6. Finally, the extension of the proposed techniques to multiuser systems is discussed and two specific applications, namely MIMO-OFDM with an iterative receiver and the uplink of the third generation cellular standard TD-SCDMA, are considered.

Remark 5 *For simplicity of presentation, in the following, the noise spatial correlation $\mathbf{R}_n$ (recall the signal model (4.1)) will be considered as known. However, its estimation could be easily included in the algorithms presented below as discussed in Appendix-A.*

5.2 Estimation of the long-term features of the channel

The long-term parameters of the MIMO channel can be estimated by following either a structured or an unstructured approach. The structured approaches are parametric methods that estimate angles and delays according to the models in Chapter 2. Such techniques have been developed for SIMO systems in order to exploit the stationarity of angles/delays in matrix $\mathbf{T}$ for TDMA [2] [3] [4] or TD-CDMA systems [5]. The extension to MIMO system is conceptually trivial and it is just a matter of increased complexity when compared to SIMO systems. Basically, the angle and delay estimation methods are based on the minimization of non-linear objective functions in order to compute the d triplets AOD/AOA/delay (for the beamforming model) or the d delays (for the diversity model) according to the knowledge of the waveform $g(t)$ and the spatial manifolds. Even if these parametric methods can closely reach the performance limits derived in Chapter 4 (see [5] for the analytic derivation of the MSE for SIMO systems), there are several drawbacks that prevent their practical use. First, the need of regular spatial and temporal manifolds imposes strict constraints on array calibration errors and mod-
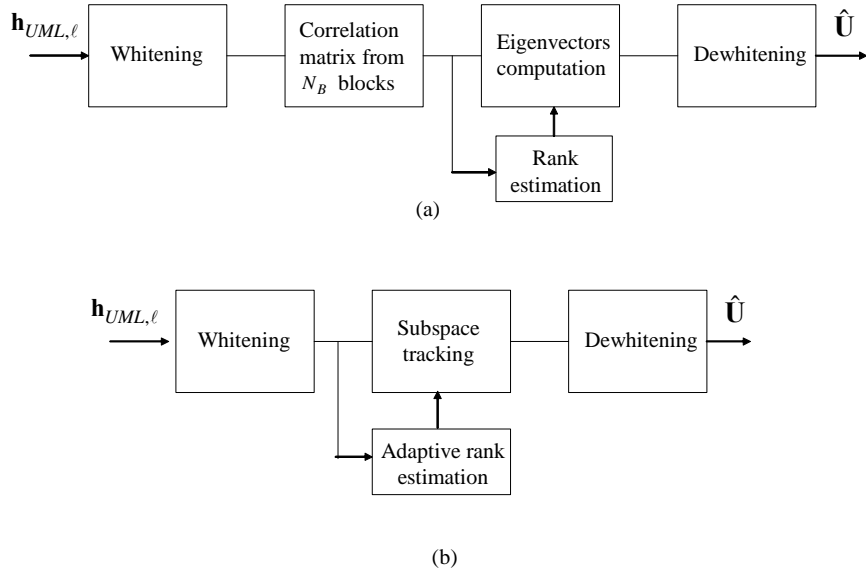
Figure 5.3: General block diagram of (a) batch or (b) subspace tracking-based modal analysis with rank estimation.

elling mismatches. Furthermore, angle and delay estimation suffers from threshold effects at low SNR's, typical of non-linear estimators.

5.2.1   Space-time modal analysis

Instead of estimating angles and delays in $\mathbf{T}$, it is possible to evaluate the space-time modes $\mathbf{U}$ (unstructured approach [6] [7]) directly. This choice not only poses less stringent requirements on array calibration and modelling accuracy (the relationship between $\mathbf{T}$ and angles/delays is not of concern) but also avoids the impairments of non-linear estimation since it reduces to a quadratic optimization problem. The estimation of modes will be referred to as *modal analysis*.

As it will be shown below, modal analysis amounts to the computation of the eigenvectors of a sample correlation matrix obtained from the UML estimates on different blocks. Fig. 5.3-(a) shows the corresponding block diagram of modal analysis (the need for whitening and dewhitening block will be clarified in the following). To have a glimpse of the main results, the reader is referred to equations (5.3) for the beamforming scenario and (5.7) for the diversity scenario. In order to reduce the computational complexity of the eigenvalue decomposition and accommodate more general models for the temporal variations of the channel, an adaptive computation of the eigenvectors is then proposed by means of subspace tracking. Fig. 5.3-(b) illustrates the block diagram of the modal analysis based on subspace tracking.

Batch estimation

Beamforming scenario   This Section considers the *beamforming scenario* since it is the most critical from the point of view of estimating the long term features of the channel. The diversity scenario will be addressed below. The aim is to show how the space-time modes can be estimated from the sample second-order statistics of the preliminary UML estimates. Toward this goal, in order to simplify the presentation, spatially uncorrelated noise and optimally designed training sequences are assumed at first (see previous Chapter) so that the error correlation matrix of UML estimation is diagonal, $\mathbf{Q}_{UML} = \sigma_n^2/(L_P P)\mathbf{I}_{N_R N_T W}$. It follows that the sample correlation matrix $1/N_B \sum_{\ell=1}^{N_B} \mathbf{h}_{UML,\ell}\mathbf{h}_{UML,\ell}^H$, computed from the set of N_B UML estimates $\mathbf{h}_{UML,\ell}$, reads for ergodicity

$$\frac{1}{N_B}\sum_{\ell=1}^{N_B}\mathbf{h}_{UML,\ell}\mathbf{h}_{UML,\ell}^H \underset{N_B\to\infty}{\to} E[\mathbf{h}_{UML,\ell}\mathbf{h}_{UML,\ell}^H] = \mathbf{R}_h + \frac{\sigma_n^2}{L_P P}\mathbf{I}. \qquad (5.2)$$

Notice that in (5.2), it is assumed that the fading is asymptotically uncorrelated (i.e., $\varphi_i(m) \to 0$ for $m \to \infty$).

Therefore, from (5.2) the space-time modes can be consistently (up to their ambiguity) estimated according to Method of Moments (MOM) principle [13] by considering the r leading eigenvectors of the sample correlation matrix of the UML estimates

$$\frac{1}{N_B}\sum_{\ell=1}^{N_B}\mathbf{h}_{UML,\ell}\mathbf{h}_{UML,\ell}^H = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^H. \qquad (5.3)$$

In case noise is not spatially white and/or the training sequences are not optimally designed, the UML estimates should be at first whitened through their error correlation matrix $\mathbf{Q}_{UML}$, i.e., $\tilde{\mathbf{h}}_{UML,\ell} = \mathbf{Q}_{UML}^{-H/2}\mathbf{h}_{UML,\ell}$, then the sample correlation $\frac{1}{N_B}\sum_{\ell=1}^{N_B}\tilde{\mathbf{h}}_{UML,\ell}\tilde{\mathbf{h}}_{UML,\ell}^H$ computed. The r leading eigenvectors $\widehat{\tilde{\mathbf{U}}}$ of the latter provide a consistent estimate of the space-time modes after de-whitening as

$$\hat{\mathbf{U}} = \mathbf{Q}_{UML}^{H/2}\widehat{\tilde{\mathbf{U}}}. \qquad (5.4)$$

The MOM estimator just presented can be justified from a ML standpoint as discussed in Sec. 5.3.1.

Diversity scenario   Similarly to the discussion above, for the *diversity scenario*, the temporal modes $\mathbf{U}_T$ can be estimated according to the MOM principle from the sample correlation matrix of the UML estimates. Again, at first spatially uncorrelated noise and optimally designed training sequences are assumed. Here, it is convenient to work with the $N_R N_T \times W$ channel matrix $\mathbf{H}_\ell$ defined in (2.7), from which we recall that the $N_R N_T W \times 1$ channel vector is obtained as $\mathbf{h}_\ell = \text{vec}\{\mathbf{H}_\ell\}$. According to the previously discussed relationships (2.26) and (4.16), it is easy to obtain that

$$E[\mathbf{H}_\ell^H\mathbf{H}_\ell] = N_R N_T \cdot \mathbf{S}_T\mathbf{S}_T^H, \qquad (5.5)$$

where we used the relationship $E[\boldsymbol{\mathcal{A}}_\ell^H \boldsymbol{\mathcal{A}}_\ell] = \sum_{i=1}^d \operatorname{tr}\{\mathbf{R}_i\} \cdot \mathbf{I}_d$. Now, denoting as $\mathbf{H}_{UML,\ell}$ the UML estimate of the channel matrix $\mathbf{H}_\ell$ ($\mathbf{h}_{UML,\ell} = \operatorname{vec}\{\mathbf{H}_{UML,\ell}\}$), the sample correlation matrix $1/N_B \sum_{\ell=1}^{N_B} \mathbf{H}_{UML,\ell}^H \mathbf{H}_{UML,\ell}$ satisfies

$$\frac{1}{N_B} \sum_{\ell=1}^{N_B} \mathbf{H}_{UML,\ell}^H \mathbf{H}_{UML,\ell} \underset{N_B \to \infty}{\to} E[\mathbf{H}_{UML,\ell}^H \mathbf{H}_{UML,\ell}] = N_R N_T \cdot \mathbf{S}_T \mathbf{S}_T^H + \frac{\sigma_n^2}{L_P P} \mathbf{I}, \ N_B \to \infty.$$
(5.6)

Therefore, a consistent estimate of the temporal modes $\mathbf{U}_T$ can be obtained by taking the r_T leading eigenvectors of the sample correlation matrix (5.6), i.e.,

$$\frac{1}{N_B} \sum_{\ell=1}^{N_B} \mathbf{H}_{UML,\ell}^H \mathbf{H}_{UML,\ell} = \hat{\mathbf{U}}_T \hat{\boldsymbol{\Lambda}}_T \hat{\mathbf{U}}_T^H.$$
(5.7)

Similarly to the previous Section, in case noise is not spatially white and/or the training sequences are not optimally designed, the UML estimates should be at first whitened through the error correlation matrix of the UML estimate $\mathbf{Q}_{UML}$. Recalling that (see (4.6) and (3.11))

$$\mathbf{Q}_{UML} = \mathbf{R}_{x,T}^{-1} \otimes \mathbf{R}_{x,S}^{-1} \otimes \mathbf{R}_n,$$
(5.8)

where $\mathbf{R}_{S,x}$ and $\mathbf{R}_{T,x}$ account for the spatial and temporal correlation of the training sequences, the whitening operation can be equivalently performed on the channel matrix $\mathbf{H}_{UML,\ell}$ as

$$\tilde{\mathbf{H}}_{UML,\ell} = (\mathbf{R}_{x,S}^{1/2} \otimes \mathbf{R}_n^{-H/2}) \mathbf{H}_{UML,\ell} (\mathbf{R}_{x,T}^{T/2}).$$
(5.9)

Once computed the sample covariance (5.7) and the corresponding r_T leading eigenvectors $\hat{\tilde{\mathbf{U}}}_T$, a consistent (up to its ambiguity) estimate of the temporal modes $\mathbf{U}_T$ can be obtained as

$$\hat{\mathbf{U}}_T = \mathbf{R}_{x,T}^{-*/2} \tilde{\mathbf{U}}_T,$$
(5.10)

and therefore an estimate of the space-time modes $\mathbf{U}$ as

$$\hat{\mathbf{U}} = \hat{\mathbf{U}}_T \otimes \mathbf{I}_{N_R N_T}$$

Remark 6 (*rank estimation*): *The number of modes r or r_T (i.e., the number of resolvable space-time or temporal signatures) is generally not known and hence should be estimated. This problem falls within the scope of the theory on model order selection. Recent advances of this field have moved the state of the art from hypothesis testing approaches, that require the definition of a somewhat arbitrary threshold, to information theoretic criteria such as Minimum Description Length (MDL) or Akaike Information Criterion [9]. The reader is referred to the cited references for further details.*

Subspace tracking adaptive estimation

So far, the quasi-static model of variations for the long term channel features in (4.2) has been taken into account. An adaptive (i.e., on a block-by-block basis) computation of the space-time modes $\mathbf{U}$ or temporal modes $\mathbf{U}_T$ that alleviates the computational burden of the eigenvalue decomposition and allow for a continuos (but still slow) variations of the channel space-time modes can be obtained through a subspace tracking algorithm [10].

Subspace tracking refers to a class of algorithms that manage to update the signal subspace (or alternatively the noise subspace) of a given sample correlation matrix whenever a novel measurement is available, with a lower computational complexity than an eigenvalue decomposition. If the sample correlation matrix has dimension $M \times M$ and signal subspace is of size s the order of complexity of an eigenvalue decomposition is $\mathcal{O}(M^3)$, whereas subspace trackers (with different properties of convergence and accuracy) have been proposed that reduce the computational burden up to $\mathcal{O}(Ms)$ [11]. Moreover, adaptive rank estimation can be included in subspace tracking procedures [10].

For the application proposed here, the subspace tracker proposed by [12] was proved to provide a good trade-off between complexity ($\mathcal{O}(Ms^2)$) and accuracy. Table 5.1 summarizes the subspace tracking algorithm with adaptive rank estimation: for this application, $\mathbf{v}_\ell = \tilde{\mathbf{h}}_{UML,\ell}$, $M = N_T N_R W$ and $s = r$ for beamforming scenario while $\mathbf{v}_\ell = \tilde{\mathbf{H}}_{UML,\ell}^H$, $M = W$ and $s = r_T$ for the diversity scenario. Notice that an upper bound $s_{\max}$ on the number of the space-time modes r or temporal modes r_T has to be predetermined, which may be derived from a priori knowledge about the channel. The coefficient γ rules the memory of the algorithm and can be adjusted to accommodate temporal variations of the channel modes. Moreover, the coefficient β is a threshold used for rank estimation purposes. Its value can be adjusted as a function of system parameters and SNR, as thoroughly explained in [13]. The estimate of the modal matrix for the ℓth block $\hat{\mathbf{E}}_\ell$ (that equals $\hat{\mathbf{U}}_\ell$ for the beamforming and $\hat{\mathbf{U}}_{T,\ell}$ for the diversity scenario) is obtained by taking the first $\hat{s}_\ell$ columns of $\mathbf{E}_\ell$. As a final remark, it is noticed that, according to [12] the subspace tracker could be implemented in a more computationally efficient way than the one illustrated in the table, still retaining the order of complexity $\mathcal{O}(Ms^2)$.

5.2.2 Decoupled spatial and temporal modal analysis

For the beamforming scenario, the estimation of the $N_R N_T W \times r$ long-term modes of the channel $\mathbf{U}$ as detailed above may suffer from high computational complexity and slow convergence due to the large size of the space-time modal matrix $\mathbf{U}$ whenever the number of transmitting/receiving antennas (N_T and N_R) and/or the temporal support of the channel W is large enough. Notice that in a diversity scenario, this problem does not hold since the temporal modes to be estimated are collected in a smaller size $W \times r$ matrix $\mathbf{U}_T$. Therefore, in the above mentioned cases, a suboptimal approach that retains the desired properties of com-

putational complexity and convergence should be preferred. Here the author proposes such an approach that considers separately the stationarity of angles and delays. The idea is the following: decouple the space-time modes ($\mathbf{U}$) into spatial ($\mathbf{U}_S$) *and* temporal ($\mathbf{U}_T$) modes, to be estimated separately. The advantage is that the corresponding modal matrices ($\mathbf{U}_S$ and $\mathbf{U}_T$) have reduced dimensions as compared to the space-time modal $\mathbf{U}$. Notice that the temporal modes have already been defined within the diversity scenario context.

The key is again to work with the $N_R N_T \times W$ channel matrix $\mathbf{H}_\ell$ (2.7). According to (2.23):

$$\mathbf{H}_\ell = \mathbf{S}_S \cdot \mathrm{diag}\{\boldsymbol{\beta}_\ell\} \cdot \mathbf{S}_T^H. \tag{5.11}$$

where the $N_R N_T \times d$ matrix $\mathbf{S}_S$ that collects by columns the spatial signatures of the d paths, $\mathbf{S}_S = \boldsymbol{\mathcal{A}}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})$. Following the lines of reasoning of Sec. 4.4, it can be concluded that the matrices that contain the spatial and temporal signatures, $\mathbf{S}_S$ and $\mathbf{S}_T$ respectively, are rank-deficient, $r_S = \mathrm{rank}\{\mathbf{S}_S\} \leq \min\{N_R N_T, N_F\}$ and $r_T = \mathrm{rank}\{\mathbf{S}_T\} \leq \{W, N_F\}$. Therefore, they can be parametrized as follows: $\mathbf{S}_S = \mathbf{U}_S \mathbf{L}_S^H$, where the full rank matrices $\mathbf{U}_S$ and $\mathbf{L}_S$ are $N_R N_T \times r$ and $d \times r$ respectively; $\mathbf{S}_T = \mathbf{U}_T \mathbf{L}_T^H$ (see Sec. 4.4) with full rank matrices $\mathbf{U}_T$ $W \times r_T$ and $\mathbf{L}_T$ $d \times r_T$. The r_S columns of the long term matrix $\mathbf{U}_S$ are defined as *spatial modes* of the channel whereas we recall that the r_T columns of the long term matrix $\mathbf{U}_T$ are defined as *temporal modes*. Notice that, without limiting the generality of the approach, the spatial and temporal modes are assumed to have unit norm. Therefore, the channel matrix can be restated in terms of spatial and temporal modes as

$$\mathbf{H}_\ell = \mathbf{U}_S \mathbf{D}_\ell \mathbf{U}_T^H, \tag{5.12}$$

with definition $\mathbf{D}_\ell = \mathbf{L}_S^H \mathrm{diag}\{\boldsymbol{\beta}_\ell\} \mathbf{L}_T$.

Batch estimation

A decoupled spatial and modal analysis estimates the spatial and temporal modes $\mathbf{U}_S$ and $\mathbf{U}_T$ according to the MOM principle from the correlation matrices of the UML estimates as done in the context of space-time modal analysis in Sec. 5.2.1. Such estimators are consistent (up to their ambiguities) and can be derived following the lines of Sec. 5.2.1. For instance, consider the estimation of spatial modes $\mathbf{U}_S$. The computation of $\mathbf{U}_T$ then follows a dual approach. From (5.11), it is

$$E[\mathbf{H}_\ell \mathbf{H}_\ell^H] = \mathbf{S}_S \boldsymbol{\Gamma}_S \mathbf{S}_S^H \tag{5.13}$$

where $\boldsymbol{\Gamma}_S = \mathbf{I}_d \odot (\mathbf{S}_T^H \mathbf{S}_T) = \mathbf{I}_d \odot \boldsymbol{\Omega}$ (assuming the approximate relationship[1] $||\mathbf{g}(\tau)||^2 = 1$). Assuming spatially white noise and optimal training sequences, for ergodicity the sample correlation matrix of the UML estimates $1/N_B \sum_{\ell=1}^{N_B} \mathbf{H}_{UML,\ell} \mathbf{H}_{UML,\ell}^H$ satisfies

$$\frac{1}{N_B} \sum_{\ell=1}^{N_B} \mathbf{H}_{UML,\ell} \mathbf{H}_{UML,\ell}^H \underset{N_B \to \infty}{\to} E[\mathbf{H}_{UML,\ell}^H \mathbf{H}_{UML,\ell}] = \mathbf{S}_S \boldsymbol{\Gamma}_S \mathbf{S}_S^H + \frac{\sigma_n^2}{L_P P} \mathbf{I}, . \tag{5.14}$$

---

[1] The relationship holds exactly only for sample spaced (T) delays.

Therefore, a consistent estimate of the spatial modes $\mathbf{U}_S$ can be obtained by taking the r_S leading eigenvectors of the sample correlation matrix (5.14), i.e.,

$$\frac{1}{N_B} \sum_{\ell=1}^{N_B} \mathbf{H}_{UML,\ell} \mathbf{H}_{UML,\ell}^H = \hat{\mathbf{U}}_S \hat{\mathbf{\Lambda}}_S \hat{\mathbf{U}}_S^H. \tag{5.15}$$

Moreover, in case noise is not spatially white and/or the training sequences are not optimally designed, the UML estimates should be at first whitened through the error correlation matrix of the UML estimate $\mathbf{Q}_{UML}$ according to (5.9). Then, once computed the sample covariance (5.14) and the corresponding r_S leading eigenvectors $\widehat{\widetilde{\mathbf{U}}}_S$, a consistent (up to its ambiguity) estimate of the temporal modes $\mathbf{U}_S$ can be obtained as

$$\hat{\mathbf{U}}_S = (\mathbf{R}_{x,S}^{-1/2} \otimes \mathbf{R}_n^{H/2}) \widehat{\widetilde{\mathbf{U}}}_S. \tag{5.16}$$

A consistent estimator for the temporal modes can be defined as in Sec. 5.2.1 from the sample correlation matrix of the whitened UML estimates $1/N_B \sum_{\ell=1}^{N_B} \mathbf{H}_{UML,\ell}^H \mathbf{H}_{UML,\ell}$ (5.10). Both the temporal and spatial analysis presented above can be justified from a ML standpoint, as it is shown in Sec. 5.3.1.
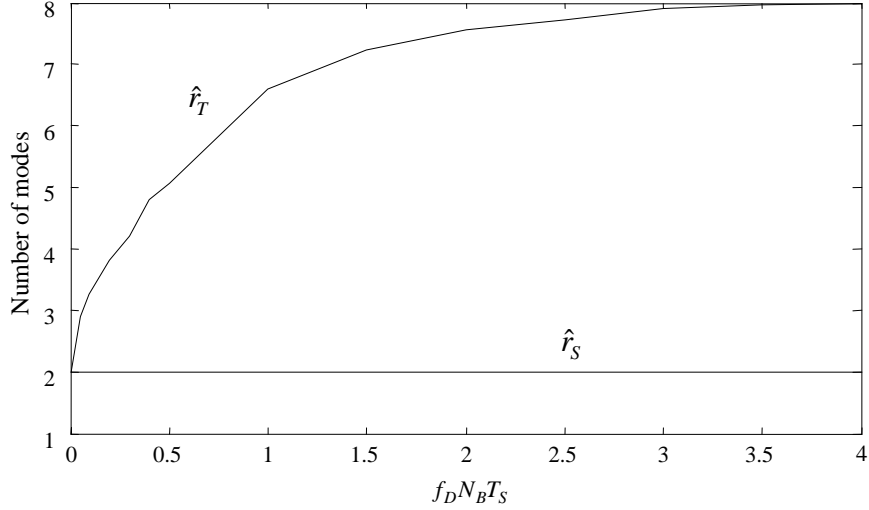
Example 3 *(rank estimation): In this example, the effect of the Doppler spread on the estimation of the spatial and temporal modes and in particular on their number r_S and r_T is studied through a numerical example. Let us consider a beamforming scenario and a space-time channel with either one transmit antenna ($N_T = 1$, SIMO channel) or one receiving antenna ($N_R = 1$, MISO channel), two resolvable angles ($r_S = 2$) and $r_T = 8$ resolvable delays. Other parameters are: $N_R = 8$ (or $N_T = 8$), $W = 15$ and $L_P = 40$. Moreover, the power delay profile is uniform and the Doppler spectrum is the same for all paths and is modelled according to the Clarke's model $\varphi(m) = J_0(2\pi f_D m T_S)$. Fig. 5.4 shows the average values of $\hat{r}_S$ and $\hat{r}_T$, carried out according to the MDL principle, obtained from 10^4 independent runs of fading versus the fading decorrelation within $N_B = 10$ blocks at large signal to noise ratios ($SNR = 20dB$). The choice of $N_B = 10$ is irrelevant as the result depends only on the product $f_D N_B T_S$ for N_B large enough. For a static channel ($f_D = 0$), fig. 5.4 shows that $\hat{r}_S = \hat{r}_T = \min(r_S, r_T) = 2$, that corresponds with the rank of the channel matrix $\text{rank}\{\mathbf{H}_\ell\}$ [14]. Moreover, for $f_D N_B T_S$ large enough to guarantee uncorrelated fading within N_B blocks, as expected the estimates are consistent: $\hat{r}_S \to r_S = 2$ and $\hat{r}_T \to r_T = 8$.*

Subspace tracking adaptive estimation

As explained in Sec. 5.2.1, an adaptive and low-complexity alternative to the computation of the eigenvalue decomposition required for spatial and temporal modal analysis is the implementation of a subspace tracking algorithm. This solution alleviates the assumption of quasi

Table 5.1: Subspace tracking algorithm with adaptive rank estimation.

$\underline{\text{Initialize:}}\ s_{\max};\ \mathbf{E}_0 = \begin{bmatrix} \mathbf{I}_{s_{\max}} \\ \mathbf{0} \end{bmatrix};\ \mathbf{\Theta}_0 = \mathbf{I}_{s_{\max}};\ \mathbf{C}_0 = \mathbf{0};\ p_0 = 0;\ 0 \leq \gamma \leq 1$

$\underline{\text{For each symbol } n:}$

input: $\mathbf{v}_\ell,\ M$

1. Subspace tracking:

$$\mathbf{Z}_\ell = \mathbf{E}_{\ell-1}^H \mathbf{v}_\ell$$
$$\mathbf{C}_\ell = \gamma \mathbf{C}_{\ell-1} \mathbf{\Theta}_{\ell-1} + \mathbf{v}_\ell \mathbf{Z}_\ell^H$$
$$\mathbf{C}_\ell = \mathbf{E}_\ell \mathbf{R}_\ell \quad \text{(QR factorization)}$$
$$\mathbf{\Theta}_\ell = \mathbf{E}_{\ell-1}^H \mathbf{E}_\ell$$

2. Adaptive rank estimation:

$$\hat{\lambda}_i = [\mathbf{R}_\ell]_{ii} \quad i = 1, 2, .., s_{\max}$$
$$p_\ell = \gamma p_{\ell-1} + \frac{1}{M} \operatorname{tr}\{\mathbf{v}_\ell \mathbf{v}_\ell^H\}$$
$$\hat{\sigma}^2 = \frac{M}{M-s_{\max}} p_\ell - \frac{1}{M-s_{\max}} \operatorname{tr}\{\mathbf{R}_\ell\}$$
$$\hat{s}_\ell = \operatorname{card}\{\hat{\lambda}_i : \hat{\lambda}_i > \beta \cdot \hat{\sigma}^2\}$$

updated basis: $\hat{\mathbf{E}}_\ell = [\mathbf{E}_\ell]_{:,1:\,\hat{s}_\ell}$

Figure 5.4: Estimated number of spatial and temporal modes $\{\hat{r}_S, \hat{r}_T\}$ versus the fading decorrelation $f_D N_B T_S$ ($r_S = 2$, $r_T = 8$).
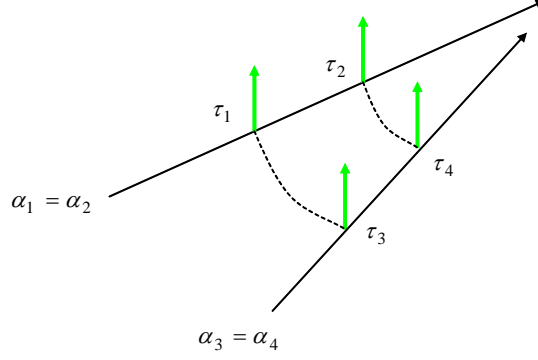
Figure 5.5: Path geometry in the angle/delay domain that guarantees the optimality of the separate spatial and temporal modal analysis.

static variations of the long term features of the channel, allowing continuos (but still slow) variations.

With reference to the subspace tracker in Table 5.1 (see Sec. 5.2.1 for further details), the input of the algorithm to track the spatial (or temporal) modes are $\mathbf{v}_\ell = \tilde{\mathbf{H}}_{UML,\ell}$ (or $\mathbf{v}_\ell = \tilde{\mathbf{H}}_{UML,\ell}^H$) and $M = N_R N_T$ (or $M = W$), whereas the output is $\hat{\mathbf{U}}_{S,\ell}$, equal to the first $\hat{r}_{S,\ell} = \hat{s}_\ell$ columns of $\mathbf{E}_\ell$ (and similarly for $\hat{\mathbf{U}}_{T,\ell}$).

Space-time versus decoupled spatial and temporal modal analysis

It is interesting to investigate the reasons why parametrization (5.12) will eventually lead to a (asymptotically with respect to N_B) suboptimum estimator. By stacking (5.12), we get

$$\mathbf{h}_\ell = vec\{\mathbf{H}_\ell\} = (\mathbf{U}_T^* \otimes \mathbf{U}_S)\mathbf{d}_\ell, \tag{5.17}$$

with the redefinition $\mathbf{d}_\ell = vec\{\mathbf{D}_\ell\}$. Recalling the channel vector parametrization (5.1), it is clear that the space-time modes $\mathbf{U}$ are here parametrized as the Kronecker product of the spatial modal matrix $\mathbf{U}_S$ and the temporal modal matrix $\mathbf{U}_T$ (apart from the inessential conjugate operation). In other terms, the space-time modes are assumed to be *separable* into spatial and temporal modes or equivalently that the number of space-time modes r can be written as $r = r_S r_T$. This condition only arises in very special situations such as in the example sketched in fig. 5.5. In this particular case (tailored for simplicity of representation on systems with $N_R = 1$ or $N_T = 1$, so that there is only one angle to be concerned with), there is a well resolved path for each couple angle/delay. In other words, the geometry of the paths in the angle/delay domain is separable over the two dimensions (and $r = 4 = r_S r_T$ where $r_S = r_T = 2$). In any other case, it is $r < r_S r_T$ and this condition implies a degraded performance of the estimator based on (5.12) for $N_B \to \infty$ as explained in Sec. 5.5.

5.3   Estimation of the fast-varying fading amplitudes

Once the modes $\mathbf{U}$ of the channel have been estimated, either through space-time modal analysis or decoupled spatial and temporal modal analysis (i.e., $\mathbf{U} = \mathbf{U}_T^* \otimes \mathbf{U}_S$), the fading amplitudes, modelled by vector $\mathbf{d}_\ell$, have to be evaluated. From the guidelines recalled at the beginning of this Chapter, the fading amplitude should be tracked according to the MMSE criterion, i.e., by Kalman filtering [15].

This thesis considers two suboptimal approaches: *i*) modal filtering: the amplitudes are modelled as deterministic parameters, therefore no tracking is performed; *ii*) LMS tracking of the fading amplitudes. In the first case, presented in Sec. 5.3.1 the structured channel estimation problem can be restated in a ML framework. The author will take the time necessary to present this aspect since as a by product it will be possible to justify the MOM estimators of the long term features of the channel presented above through the ML principle. On the other hand, the second case will be studied in Sec. 5.3.2.

5.3.1   Modal filtering

As a first solution to the problem of estimating the fading amplitudes, the simplest case where this estimation is performed separately on each block is addressed. More specifically, the amplitudes are modelled as deterministic parameters to be estimated anew on each block. This approach is apparently suboptimal but, as discussed in the previous Chapter (see Sec. 4.9), it approaches the optimal performance for spatially and temporally uncorrelated fading amplitudes and large signal to noise ratios. Under this framework since both the channel modes and the fading amplitudes are modelled as deterministic parameters, the channel estimation problem can be defined as a structured (or constrained) ML estimate. In Sec. 4.9 the corresponding CRB has been derived. Here, this analytical approach will be pursued and, as a by-product, it will be possible to justify the MOM estimators of the long term features of the channel presented above from a ML standpoint. For simplicity of presentation, in the following the treatment is focused on the beamforming scenario since the results concerning the diversity scenario can be easily derived following the same reasoning.

Space-time modal filtering

The ML estimation of the channel vector parametrized as (5.1) is obtained by minimizing the negative log-likelihood function (recall the signal model (4.1))

$$\mathcal{L}\left(\mathbf{U}, \mathbf{D}\right) = \sum_{\ell=1}^{N_B} \|\mathbf{y}_\ell - \mathbf{X}\mathbf{U}\mathbf{d}_\ell\|^2_{\mathbf{I}_{L_P} \otimes \mathbf{R}_n^{-1}} \tag{5.18}$$

with respect to the parameters $\{\mathbf{U}, \mathbf{D}\}$, where $\mathbf{D} = [\mathbf{d}_1 \cdots \mathbf{d}_{N_B}]$. The number of space-time modes r is assumed to be known within this section (see remark in Sec. 5.2.1). It can be

easily shown that the likelihood function (4.69) can be restated in terms of the whitened UML estimates $\tilde{\mathbf{h}}_{UML,\ell}$ (5.9) as

$$\mathcal{L}\left(\mathbf{U}, \mathbf{D}\right) = \sum_{\ell=1}^{N_B} \left\| \tilde{\mathbf{h}}_{UML,\ell} - \tilde{\mathbf{U}}\mathbf{d}_\ell \right\|^2, \qquad (5.19)$$

where $\tilde{\mathbf{U}} = \mathbf{Q}_{UML}^{-H/2}\mathbf{U}$. The UML estimates are thus sufficient statistics for the estimation of the parameters $\{\mathbf{U}, \mathbf{D}\}$ [8]. The optimization of $\mathcal{L}\left(\mathbf{U}, \mathbf{D}\right)$ in (5.19) can be carried out at first with respect to the amplitudes $\mathbf{D}$. The result is the estimate

$$\hat{\mathbf{d}}_\ell = \arg\min_{\mathbf{d}_\ell} \mathcal{L}\left(\mathbf{U}, \mathbf{D}\right) = \tilde{\mathbf{U}}^\dagger \tilde{\mathbf{h}}_{UML,\ell}. \qquad (5.20)$$

Substitution of (5.20) into (5.19) yields:

$$\mathcal{L}\left(\mathbf{U}, \hat{\mathbf{D}}\right) = \sum_{\ell=1}^{N_B} \left\| \tilde{\mathbf{h}}_{UML,\ell} - \mathbf{\Pi}_{\tilde{\mathbf{U}}} \tilde{\mathbf{h}}_{UML,\ell} \right\|^2, \qquad (5.21)$$

where $\mathbf{\Pi}_{\tilde{\mathbf{U}}} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\dagger$, i.e., $\mathbf{\Pi}_{\tilde{\mathbf{U}}}$ is the projection matrix onto the subspace spanned by the columns of $\tilde{\mathbf{U}}$, $range\{\tilde{\mathbf{U}}\}$. It follows that the ML estimation of the space-time modes reads

$$\begin{aligned} \hat{\mathbf{U}} &= \arg\min_{\mathbf{U}} \mathcal{L}\left(\mathbf{U}, \hat{\mathbf{D}}\right) = \\ &= \arg\max_{\mathbf{U}} tr\{\mathbf{\Pi}_{\tilde{\mathbf{U}}} \frac{1}{N_B} \sum_{\ell=1}^{N_B} \tilde{\mathbf{h}}_{UML,\ell} \tilde{\mathbf{h}}_{UML,\ell}^H\}. \end{aligned} \qquad (5.22)$$

Therefore, the ML estimation of the space-time modes coincides (up to its ambiguity) with the MOM estimator for the space-time modal analysis described in Sec. 5.2.1.

To sum up, once the estimate of the long term features of the channel has been obtained according to Sec. 5.2.1 (see (5.4)), the fading amplitudes are computed as (5.20), leading to the channel estimate $\hat{\mathbf{h}}_\ell = \hat{\mathbf{U}}\hat{\mathbf{d}}_\ell$, that can also be stated as

$$\hat{\mathbf{h}}_\ell = \mathbf{Q}_{UML}^{H/2} \hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}} \tilde{\mathbf{h}}_{UML,\ell}. \qquad (5.23)$$

From (5.23), this channel estimation approach is referred to as *modal filtering* since, apart from the whitening and dewhitening operations, it consists of a projection (filtering) of the preliminary UML estimate onto the estimated modal subspace. An example that illustrates the effects of modal filtering is proposed in the next Section.

As it will be proved analytically in Sec. 5.5, modal filtering leverages on the reduced rank properties of the space-time signatures matrix $\mathbf{S}_{ST} = \mathbf{T}$. In particular, the asymptotic ($N_B \rightarrow \infty$) gain in terms of MSE on the channel estimate as compared to UML estimation can be quantified in the ratio $r/(N_R N_T W)$. This behavior was already observed by studying the properties of the asymptotic HCRB (that under appropriate assumptions coincides with the CRB) for high SNR in Sec. 4.7.2.

Decoupled spatial and temporal modal filtering

The decoupled spatial and temporal modal analysis (for the beamforming scenario) is based on the parametrization of the channel matrix (5.12). Accordingly, in this context the channel estimation problem translates into the estimation of the spatial and temporal modes $\{\mathbf{U}_S, \mathbf{U}_T\}$ and the $r \times r$ weighting amplitudes $\mathbf{D}_\ell$. Here, following the idea of modal filtering, matrix $\mathbf{D}_\ell$ is considered as deterministic. Similarly to the previous Section, it can be shown that the UML estimates of the channel are sufficient statistics for the estimation of $\{\mathbf{U}_S, \mathbf{U}_T, \mathbf{D}\}$ defining for convenience $\mathbf{D} = [\mathbf{d}_1 \cdots \mathbf{d}_{N_B}])$ and their ML estimation can be evaluated by minimizing the negative log-likelihood function

$$\mathcal{L}(\mathbf{U}_S, \mathbf{U}_T, \mathbf{D}) = \sum_{\ell=1}^{N_B} \left\| \tilde{\mathbf{H}}_{UML,\ell} - \tilde{\mathbf{U}}_S \mathbf{D}_\ell \tilde{\mathbf{U}}_T^T \right\|^2 \tag{5.24}$$

with the definition of the whitened quantities $\tilde{\mathbf{U}}_S = (\mathbf{R}_{x,S}^{1/2} \otimes \mathbf{R}_n^{-H/2}) \mathbf{U}_S$ and $\tilde{\mathbf{U}}_T = \mathbf{R}_{x,T}^{*/2} \mathbf{U}_T$. The number of modes r_S and r_T are assumed to be known within this section (see remark in Sec. 5.2.1). The optimization of $\mathrm{L}(\mathbf{U}_S, \mathbf{U}_T, \mathbf{D})$ in (5.24) can be carried out at first with respect to $\mathbf{D}$. This result is in

$$\hat{\mathbf{D}}_\ell = \arg\min_{\hat{\mathbf{D}}_\ell} \mathcal{L}(\mathbf{U}_S, \mathbf{U}_T, \mathbf{D}) = \tilde{\mathbf{U}}_S^\dagger \tilde{\mathbf{H}}_{UML,\ell}(\ell) \tilde{\mathbf{U}}_T^{\dagger H}. \tag{5.25}$$

The substitution of (5.25) into (5.24) yields:

$$\mathcal{L}\left(\mathbf{U}_S, \mathbf{U}_T, \hat{\mathbf{D}}\right) = \sum_{\ell=1}^{N_B} \left\| \tilde{\mathbf{H}}_{UML,\ell} - \mathbf{\Pi}_{\tilde{\mathbf{U}}_S} \tilde{\mathbf{H}}_{UML,\ell} \mathbf{\Pi}_{\tilde{\mathbf{U}}_T} \right\|^2, \tag{5.26}$$

The estimation of $\tilde{\mathbf{U}}_S$ and $\tilde{\mathbf{U}}_T$ is turned into the minimization of (5.26) with respect to $\mathbf{\Pi}_{\tilde{\mathbf{U}}_S}$ and $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}$, constrained to be projection matrices of rank order r_S and r_T, respectively. By making use of the trace operator properties [16], the minimization of (5.26) can be equivalently stated as

$$\left\{\hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_S}, \hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_T}\right\} = \arg\max_{\{\mathbf{\Pi}_S, \mathbf{\Pi}_T\}} tr\left\{\mathbf{\Pi}_{\tilde{\mathbf{U}}_S} \sum_{\ell=1}^{N_B} \tilde{\mathbf{H}}_{UML,\ell} \mathbf{\Pi}_{\tilde{\mathbf{U}}_T} \tilde{\mathbf{H}}_{UML,\ell}^H\right\} \tag{5.27}$$

$$= \arg\max_{\{\mathbf{\Pi}_S, \mathbf{\Pi}_T\}} tr\left\{\mathbf{\Pi}_{\tilde{\mathbf{U}}_T} \sum_{\ell=1}^{N_B} \tilde{\mathbf{H}}_{UML,\ell}^H \mathbf{\Pi}_{\tilde{\mathbf{U}}_S} \tilde{\mathbf{H}}_{UML,\ell}\right\}. \tag{5.28}$$

The optimization (5.27) or (5.28) is non-linear. The separable characteristics of the objective function suggests that the minimization can be carried out iteratively by alternating the search for $\mathbf{\Pi}_{\tilde{\mathbf{U}}_S}$ (given $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}$) and $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}$ (given $\mathbf{\Pi}_{\tilde{\mathbf{U}}_S}$). For $N_B = 1$ (or equivalently for a static channel) the minimization (5.27) or (5.28) yields the reduced-rank estimate [17] [14], while for $N_B \to \infty$ the closed form solution can be based on a privileged choice of the initialization for the alternate search. Let us consider the optimization (5.27) with respect to $\mathbf{\Pi}_{\tilde{\mathbf{U}}_S}$ for any given

initialization $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T} = \mathbf{\Pi}_{\tilde{\mathbf{U}}_T}^{(0)}$ (the same reasoning applies dually for the optimization of (5.28)). Any choice $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}^{(0)}$ such that $range\{\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}^{(0)} \tilde{\mathbf{H}}_{UML,\ell}^H\} \subset range\{\tilde{\mathbf{H}}_{UML,\ell}^H\}$ affects the estimate of $\mathbf{\Pi}_{\tilde{\mathbf{U}}_S}$ since it reduces the set of solutions to those that are compatible to the initial choice $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}^{(0)}$. In order not to bias the final solution, here it is preferred to decouple the optimizations by relaxing the constraint on the temporal structure and choosing $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}^{(0)} = \mathbf{I}_W$. With this choice the estimate $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_S}$ can be obtained very easily from the leading r_S eigenvectors of the spatial correlation matrix $\frac{1}{N_B} \sum_{\ell=1}^{N_B} \tilde{\mathbf{H}}_{UML,\ell} \tilde{\mathbf{H}}_{UML,\ell}^H$. Therefore, the estimate of the spatial modes $\hat{\mathbf{U}}_S$ coincides with the MOM estimator for spatial modal analysis (5.16). Similarly, it can be easily shown that the estimate of the spatial modes $\hat{\mathbf{U}}_T$ coincides with the MOM estimator for temporal modal analysis (5.16). Notice that the procedure could be iterated by alternating the search once initialized as explained above but in practice there is no relevant improvement to justify the additional costs [17]. Moreover, these choices can be shown to coincide with the exact ML estimation for $N_B \to \infty$ (see Appendix-B).

To sum up, once the estimate of the spatial and temporal modes have been obtained according to (5.16) and (5.10), the fading amplitudes are computed as (5.25), leading to the channel estimate $\hat{\mathbf{H}}_\ell = \hat{\mathbf{U}}_S \hat{\mathbf{D}}_\ell \hat{\mathbf{U}}_T^H$, that can also be stated as

$$\hat{\mathbf{H}}_\ell = (\mathbf{R}_{x,S}^{-1/2} \otimes \mathbf{R}_n^{H/2}) \hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_S} \tilde{\mathbf{H}}_{UML,\ell} \hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_T} \mathbf{R}_{x,T}^{-T/2}, \quad \text{for } \ell = 1, 2, ..., N_B. \tag{5.29}$$

From (5.29), this channel estimation approach is referred to again as *modal filtering* since, apart from the whitening and dewhitening operations, it consists of a projection (filtering) of the preliminary UML estimate onto the estimated spatial and temporal modal subspaces.

Remark 7 *(**spatial or temporal modal filtering**): As it will be proved analytically in Sec. 5.5, spatial and temporal modal filtering leverages on the reduced rank properties of the path signatures matrices $\mathbf{S}_S$ and $\mathbf{S}_T$ respectively. In particular it can be proved that the asymptotic gain in terms of MSE on the channel estimate as compared to UML estimation can be quantified in the product $r_S/(N_R N_T) \cdot r_T/W$, where the first term accounts for the spatial gain and the second for the temporal gain. It is then clear that whenever the channel has a dense spatial scattering, i.e., $r_S \simeq N_R N_T$ there might be little to be gained from spatial modal filtering. Therefore, a good way to trade complexity for a slight decrease, if any, in performance would be the implementation of temporal modal filtering only. Similarly, if the channel has a dense temporal scattering, i.e., $r_S \simeq W$, it might be convenient to implement spatial modal analysis only. In [19] an estimator that performs spatial modal analysis for a frequency-flat MIMO is proposed whereas [18] presents an analogous technique based on temporal modal analysis. As a final remark, notice that the space-time analysis defined within a diversity scenario in Sec. 5.2.1 coincides with the temporal modal analysis discussed above.*

Example 4 *The effect of spatial and temporal modal filtering is illustrated by an example in fig. 5.6. The multipath channel is composed of $d = 5$ paths having power delay profile*

$\mathbf{\Omega} = \Omega_0 \operatorname{diag}\{[0\ -1.2057\ -2.3976\ -3.7239\ -5.9944][dB]\}$, *where Ω_0 is selected so as to guarantee channel normalization (2.35). For simplicity, we consider a SIMO system ($N_T = 1$), spatially white noise and optimal training sequences. Moreover, the number of blocks N_B is selected to be large enough to guarantee convergence of the estimates of the long term features of the channel. The geometry of the paths between the transmitter (TX) and the receiver (RX) is shown in fig. 5.6-a. The UML estimate $\mathbf{H}_{UML,\ell}$ is plotted in fig. 5.6-b through its power-delay-angle diagram $\mathcal{P}(i, \alpha)$. This diagram can be obtained as $\mathcal{P}(i, \alpha) = |[\mathbf{a}(\alpha)^H \mathbf{H}_{UML,\ell}]_{1,i}|^2$, where i runs over the discrete-time delay axis $i = 1, .., W$ and $\alpha \in [-\pi/3, \pi/3]$ span the DOA axis. Since $\alpha_1^{(R)} = \alpha_2^{(R)}$, $\alpha_3^{(R)} = \alpha_4^{(R)}$ and $\tau_4 = \tau_5$, the spatial and temporal modes are, respectively, $r_S = 3$ and $r_T = 4$. The projector $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_S}$ is calculated by using different values of the estimate $\hat{r}_S = 1 \div 3$. The same holds for $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_T}$ with $\hat{r}_T = 1 \div 4$. As shown in fig. 5.6-c, d and e, by projecting the space-time matrix $\mathbf{H}_{UML,\ell}$ the background noise is reduced. Notice that in fig. 5.6-c the temporal projector $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_T}$ with $\hat{r}_T = 1$ selects the path of the channel that has the largest mean power Ω_1, even though this is not the temporal component with the largest local power. The residual noise after the projection is that component that can no longer be eliminated as it belongs to the same subspace of the channel. By comparing the power-delay-angle diagram for the initial space-time matrix $\mathbf{H}_{UML,\ell}$ in fig. 5.6-a with the projected matrix $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_S} \mathbf{H}_{UML,\ell} \hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_T}$ in fig. 5.6-e, the visual inspection shows that the artifacts due to noise are reduced. This noise reduction depends on the ratio $r_S r_T / N_R W$. Indeed, it can be observed in Fig. 5.6-e that the double projection selects $r_S r_T = 12$ "intersections" in the space-time domain, i.e. the multipath components that have angles in $\{\alpha_1, \alpha_3, \alpha_5\}$ and delays in $\{\tau_1, \tau_2, \tau_3, \tau_4\}$. Since the initial UML in fig. 5.6-a contains all the $N_R W$ channel samples, the noise-reduction after the double projection is $r_S r_T / N_R W$.*

Modal filtering through subspace tracking

In order to get an adaptive implementation of modal filtering, it is enough to perform modal analysis according to Sec. 5.2.1 followed by the block-by-block estimation of the amplitudes (5.20) (space-time) or (5.25) (decoupled spatial and temporal modal filtering).

5.3.2 LMS tracking of the fading amplitudes

Based on the knowledge of the fading statistics and the estimate of the channel modes $\hat{\mathbf{U}}$, the optimal channel estimation approach described in the previous Chapter prescribes MMSE tracking of the fading amplitudes, e.g., by means of the classical Kalman filter. Alternatively, the suboptimal techniques proposed in [20] reduce the computational complexity of the Kalman filter with minor performance losses. Here, to get a computationally simpler estimator we investigate tracking of the fading process through the LMS algorithm. In [21] this solution was studied for the case $N_T = 1$ and spatial *or* temporal modal analysis.
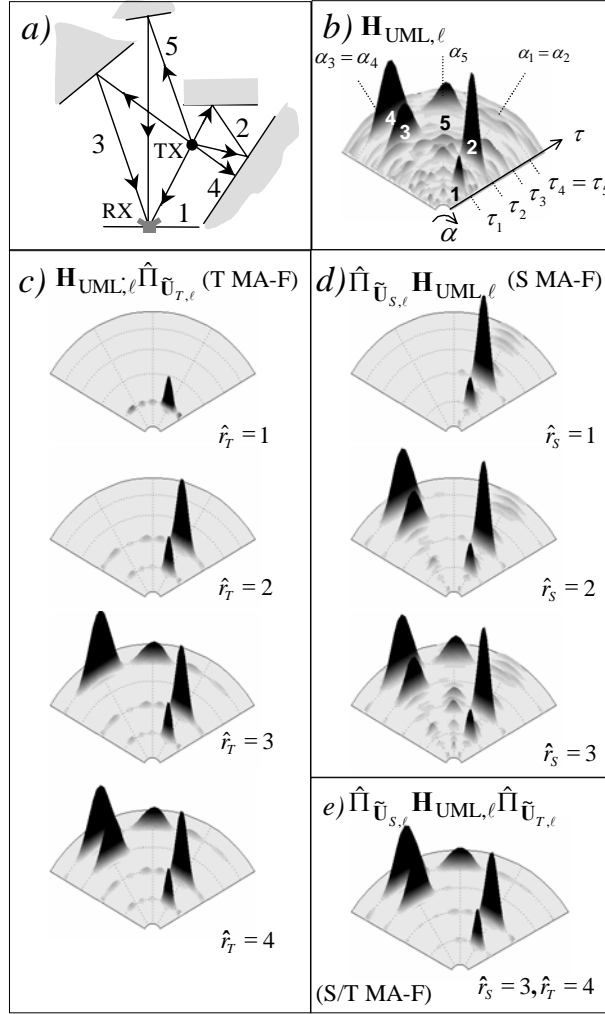
Figure 5.6: Example of space and/or time modal filtering for a channel with $d = 5$ paths and different estimated number of modes $\hat{r}_S$ and $\hat{r}_T$: *a*) multipath model; *b*) power-delay-angle diagram $\mathcal{P}(i, \alpha)$ from the UML estimate; temporal (c) and spatial (d) filtering of $\mathbf{H}_{UML,\ell}$ with increasing dimensions; e) spatial and modal filtering.

Recalling that the estimate of the channel vector is

$$\hat{\mathbf{h}}_\ell = \hat{\mathbf{U}}\hat{\mathbf{d}}_\ell \tag{5.30}$$

(where for decoupled spatial and temporal modal analysis we have $\hat{\mathbf{U}} = (\hat{\mathbf{U}}^*_{T,\ell} \otimes \hat{\mathbf{U}}_{S,\ell})$), it is straightforward to show that LMS tracking of the vector $\mathbf{d}_\ell$ can be obtained as follows [15] ($0 \leqslant \mu \leqslant 2$ in order to guarantee stability; $\mathbf{b}_0$ can be initialized as $\mathbf{b}_0 = \mathbf{0}$):

$$\begin{aligned}
\boldsymbol{\epsilon}_\ell &= \mathbf{h}_{UML,\ell} - \hat{\mathbf{U}}_\ell \hat{\mathbf{d}}_{-1} \tag{5.31a} \\
\hat{\mathbf{d}}_\ell &= \hat{\mathbf{d}}_{-1} + \mu \hat{\mathbf{U}}^H_\ell \boldsymbol{\epsilon}_\ell. \tag{5.31b}
\end{aligned}$$

Notice that, in case modal analysis through subspace tracking is implemented, the estimated number of space time modes $\hat{r}_\ell$ (for decoupled spatial and temporal modal analysis $\hat{r}_\ell = \hat{r}_{S,\ell}\hat{r}_{T,\ell}$) may vary as a function of the blocks ℓ. Accordingly, the number of amplitudes to be tracked, i.e., the size of $\hat{r}_\ell \times 1$ vector $\hat{\mathbf{d}}_\ell$, varies. Therefore, LMS tracking (5.31) should be performed on the space-time modes of size $\hat{r}_{\max}$ (for decoupled spatial and temporal modal analysis $\hat{r}_{\max} = \hat{r}_{S,\max}\hat{r}_{T,\max}$) defined in Table 5.1, thus updating a vector $\hat{\mathbf{d}}_\ell$ of size $\hat{r}_{\max} \times 1$ of which the first $\hat{r}_\ell$ entries correspond to the useful part.

5.4   Modal channel estimation: a summary

The proposed channel estimators can be summarized by means of the block diagrams in fig. 5.7 and 5.8. Notice that modal analysis can be implemented either following approach or through subspace tracking accordin to fig. 5.3. However, subspace tracking should be generally preferred since, as discussed in Sec. 5.2.1, it reduces the computational complexity of the batch approach with negligible performance degradation (see also Sec. 5.6 for numerical investigation).

- *Modal analysis/modal filtering* (fig. 5.7): apart from the necessary whitening and de-whitening operations, these estimators perform modal analysis, either by space-time (Sec. 5.2.1) or decoupled spatial and temporal modal analysis (Sec. 5.2.2), followed by modal filtering (Sec. 5.3.1). This approach is expected to be asymptotically optimum for spatially and temporally uncorrelated fading amplitudes and large SNR. According to table 5.2 the estimator based on space-time modal analysis/filtering will be referred to as MA-F whereas the estimator based on the decoupled computation of spatial and temporal modes will be referred to with the acronym S/T MA-F. The S/T MA-F algorithm is known to be suboptimum as compared to MA-F as explained in Sec. 5.2.2. Moreover, according to the remark Sec. 5.3.1, we define S MA-F and T MA-F the estimators based respectively on spatial or temporal modal analysis only.

- *Modal analysis with LMS tracking* (fig. 5.8): these estimators differ from the previous ones in that after modal analysis they perform LMS tracking of the fading amplitudes

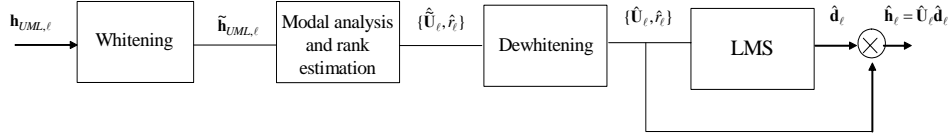Figure 5.7: Block diagram of modal analysis/modal filtering channel estimation.

Figure 5.8: Block diagram of modal analysis with LMS tracking channel estimation.

(Sec. 5.3.2). This structure is an approximation of the (asymptotically) optimal approach that prescribes MMSE tracking. Therefore, it is expected to bring some performance improvement as compared to the computationally simpler modal analysis/filtering channel estimation. According to table 5.2, the estimator based on space-time modal analysis/filtering will be referred to as MA-LMS whereas the estimator based on the decoupled computation of spatial and temporal modes will be referred to with the acronym S/T MA-LMS.

The two approaches above can be described by the same block diagram in fig. 5.9 if the noise is spatially white and the training sequences are optimally designed, i.e., if there is no need to introduce the whitening and dewhitening operations.

Table 5.2 classifies the channel estimators described above in terms of the propagation scenarios, either beamforming (bf) and or diversity (div), in which they can be applied. Notice that the decoupled spatial and temporal modal analysis S/T MA-F and the alternative approaches S MA-F and T MA-F have been defined for a beamforming scenario. However, from the discussion in Sec. 5.2.1 and 5.2.2, it is clear that, within a diversity scenario, the T MA-F algorithm coincides with the MA-F estimator.

5.5 Asymptotic MSE performance analysis

In this Section, the asymptotic MSE of the channel estimators in table 5.2 is computed and compared with the MSE of the conventional UML estimate and with the asymptotic HCRB derived in the previous Chapter. To simplify the analysis and according to the model used for derivation of the HCRB, the long term features of the channel are assumed to be constant over an infinite temporal horizon ($N_B \to \infty$). Moreover, it is assumed that the number of modes r

Table 5.2: Channel estimators based on modal analysis.

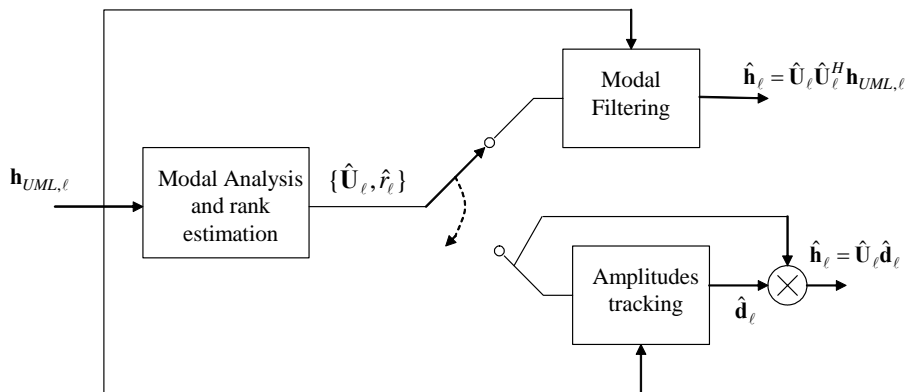| Acronym	Modal analysis	Fast fading amplitudes	Scenario
MA-F	space-time	modal filtering	bf, div
S/T MA-F	decoupled spatial and temporal	modal filtering	bf
S MA-F	spatial	modal filtering	bf
T MA-F	temporal	modal filtering	bf, div
MA-LMS	space-time	LMS	bf, div
S/T MA-LMS	decoupled spatial and temporal	LMS	bf

Figure 5.9: Block diagram of modal analysis channel estimation with modal filtering or LMS tracking under the assumption of spatially white noise and optimal training sequences.
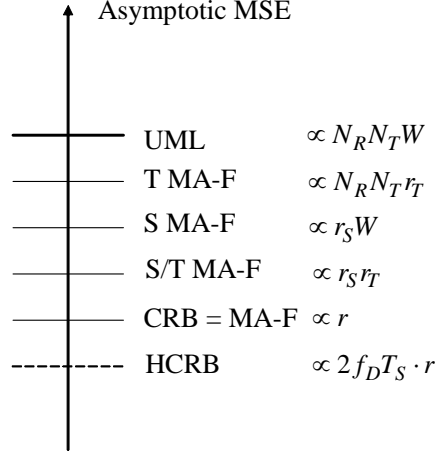
Figure 5.10: Scaling law (versus SNR) of the asymptotic MSE of different modal analysis/ filtering schemes for spatially white noise and optimally designed training sequences. For comparison between S MA-F and T MA-F, the realistic condition $r_S W < r_T N_R N_T$ is considered.

(or r_S and r_T) is accurately estimated, which is only possible for sufficiently high SNR, as it will be shown by simulations in Sec. 5.6. The definition "asymptotic" in this section has thus to be interpreted both as a function of the number of training OFDM symbols ($N_B \to \infty$) and of SNR ($SNR \to \infty$).

To get a preview of the results proved below, the reader is referred to fig. 5.10, where the scaling law (versus SNR) of the MSE's for different estimators is depicted for the case of spatially white noise and optimally designed training sequence. As it is expected, the MA-F estimator is able to reach the conventional CRB, that in turn coincides with the HCRB for $f_D T_s = 1/2$. Suboptimal techniques have degraded performance depending on the characteristics of the propagation environment. The performance analysis of modal analysis with LMS tracking could be pursued as shown in [22]. Here, the author points to this reference for further details.

To start with, the main results of interest previously derived are recalled. From (4.6), the MSE for the UML channel estimate reads (see also [23])

$$MSE_{UML} = E[||\mathbf{h}_{UML,\ell} - \mathbf{h}_\ell||^2] = tr\{\mathbf{Q}_{UML}\} = tr\{\mathbf{R}_x^{-1}\} tr\{\mathbf{R}_n\}, \qquad (5.32)$$

that for spatially white noise and optimally designed training sequences becomes:

$$MSE_{UML} = \frac{N_T N_R W}{L_P} \frac{\sigma_n^2}{P}. \qquad (5.33)$$

For reference, recall that under the same assumptions, for high SNR and uniform Doppler spectrum for all paths, the asymptotic HCRB derived in the previous Chapter is (see Sec. 4.7.2)

$$MSE_{\hat{\mathbf{h}}_\ell} \simeq \frac{2 f_D T_s r}{L_P} \frac{\sigma_n^2}{P}. \qquad (5.34)$$

Also recall that the asymptotic HCRB (5.34) coincides with the CRB for maximum normalized Doppler spread, $f_D T_s = 1/2$ (see Sec. 4.9).

In the following, the performance of modal analysis/filtering is investigated. Using the consistency of the estimation of the long term features of the channel (i.e., $\hat{\mathbf{U}}_\ell \rightarrow \mathbf{U}$ and $\hat{\mathbf{U}}_{S,\ell} \rightarrow \mathbf{U}_S$, $\hat{\mathbf{U}}_{T,\ell} \rightarrow \mathbf{U}_T$ for $N_B \rightarrow \infty$), it is easy to show that the MSE of the MA-F estimator is

$$MSE_{MA-F}(N_B \rightarrow \infty) = tr\{\mathbf{Q}_{UML}^{H/2}\mathbf{\Pi}_{\tilde{\mathbf{U}}}\mathbf{Q}_{UML}^{1/2}\}, \tag{5.35}$$

where the projection matrix $\mathbf{\Pi}$ is $\mathbf{\Pi} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\dagger$. Moreover, using the relationship $\tilde{\mathbf{U}} = \tilde{\mathbf{U}}_T^* \otimes \tilde{\mathbf{U}}_S$ valid for the S/T MA-F, that implies $\mathbf{\Pi} = \mathbf{\Pi}_T^* \otimes \mathbf{\Pi}_S$, the MSE for S/T MA-F is easily obtained:

$$\begin{aligned} MSE_{S/T\,MA-F}(N_B \rightarrow \infty) &= tr\{\mathbf{R}_{x,T}^{-1/2}\mathbf{\Pi}_{\tilde{\mathbf{U}}_T^*}\mathbf{R}_{x,T}^{-H/2}\} \\ &\quad \cdot tr\{(\mathbf{R}_{x,S}^{-1/2} \otimes \mathbf{R}_n^{H/2})\mathbf{\Pi}_{\tilde{\mathbf{U}}_S}(\mathbf{R}_{x,S}^{-H/2} \otimes \mathbf{R}_n^{1/2})\}. \end{aligned} \tag{5.36}$$

These results simplify for spatially white noise and optimally designed training sequences as

$$MSE_{MA-F}(N_B \rightarrow \infty) = \frac{r}{L_P}\frac{\sigma_n^2}{P} \tag{5.37}$$

and

$$MSE_{S/T\,MA-F}(N_B \rightarrow \infty) = \frac{r_S r_T}{L_P}\frac{\sigma_n^2}{P}. \tag{5.38}$$

Comparing (5.37) with (4.40), we can conclude that modal analysis/filtering allows to achieve the asymptotic HCRB for high SNR when the amplitudes are uncorrelated from block to block ($f_D T_s = 1/2$). In this case, as explained in Sec. 4.9, it attains the conventional CRB as well. Moreover, comparing with (5.33) we see that MA-F allows a reduction in the MSE with respect to the UML method equal to the ratio $N_R N_T W/r$.

On the other hand, inspection of (5.38) confirms that the suboptimality of the S/T MA-F approach can be quantified as discussed in Sec. 5.2.2 as $r/(r_S r_T)$. Moreover, its gain with respect to UML is given by the product of the spatial gain $N_R N_T/r_S$ and a temporal gain W/r_T. These gains quantify the reduction in the number of parameters to be estimated due to the exploitation of the structure of the channel vector.

The performance of modal filtering with spatial (S MA-F) or temporal (T MA-F) modal filtering only can be similarly computed, yielding

$$\begin{aligned} MSE_{S\,MA-F}(N_B \rightarrow \infty) &= tr\{(\mathbf{R}_{x,S}^{-1/2} \otimes \mathbf{R}_n^{H/2})\mathbf{\Pi}_{\tilde{\mathbf{U}}_S} \\ &\quad \cdot (\mathbf{R}_{x,S}^{-H/2} \otimes \mathbf{R}_n^{1/2})\}tr\{\mathbf{R}_{x,T}^{-1}\} \end{aligned} \tag{5.39}$$

$$MSE_{T\,MA-F}(N_B \rightarrow \infty) = tr\{\mathbf{R}_{x,T}^{-1/2}\mathbf{\Pi}_{\tilde{\mathbf{U}}_T^*}\mathbf{R}_{x,T}^{-H/2}\}tr\{\mathbf{R}_{x,S}^{-1}\}tr\{\mathbf{R}_n^H\}, \tag{5.40}$$

that for spatially white noise and optimally designed training sequences become:

$$MSE_{S\,MA-F}(N_B \rightarrow \infty) = \frac{r_S W}{L_P}\frac{\sigma_n^2}{P} \tag{5.41}$$

$$MSE_{T\,MA-F}(N_B \rightarrow \infty) = \frac{N_R N_T r_T}{L_P}\frac{\sigma_n^2}{P}. \tag{5.42}$$

The consideration made in the remark of Sec. 5.3.1 are confirmed by the analytical results (5.41)-(5.42). In particular, the gain of spatial modal filtering as compared to UML is quantified by the ratio $r_S/(N_R N_T)$ and the gain of temporal filtering by r_T/W.

5.6 Numerical results

5.6.1 Example 1: SISO channel

In this first example, a SISO channel ($N_T = 1, N_R = 1$) is simulated. Under this simple setting, it is easy to investigate the relationship between the proposed linear estimation of the long term features of the channel (modal analysis) and the structured techniques that are based on the non-linear computation of the multipath parameters (here, the delays) (recall Sec. 5.2, see fig. 5.11). Moreover, since the channel estimators under study are based on the invariance of the spatial and temporal channel manifolds [2], it is interesting to evaluate the effects of system stability over successive blocks. Here this investigation is carried out by modelling the stability of oscillators at the transmit and receive side by considering residual timing offsets (fig. 5.12). Notice that in a spatial (i.e., SIMO, MISO or MIMO) channel, stability of the calibration of the antenna arrays [24] could be taken into account in a similar way. As a final remark, the author recalls that in a SISO channel, there is no distinction between beamforming and diversity scenario, since this taxonomy is related to the spatial features of the channel. Moreover, the MA-F estimator coincides with S/T MA-F.

Simulation setting

The simulation setting is as follows: the channel length is $W = 15$ samples, the length of the training sequence, assumed to be optimally designed, is $L_P = 20$ (larger than the minimum value $L_P \geq W = 15$) and the fading amplitudes uncorrelated from block to block $\varphi_i(m) = \delta(m)$, $i = 1, ..., d$. Under this last assumption, as already discussed, there is nothing to be gained from tracking the fading amplitudes; therefore herein the investigation will be limited to the MA-F estimator.

In this first example the number of (temporal) modes will be assumed to be known, $\hat{r} = r = r_T$.

Structured versus unstructured estimate of the long term features of the channel

To simplify the problem at hand, herein a channel with a single propagation path ($d = 1$) with delay $5.1T$ is considered. Fig. 5.11 compares the MSE of MA-F computed through simulations for $N_B = 10, 30$ blocks with the asymptotic HCRB (dashed lines). Notice that the HCRB (4.25) coincides with its asymptotic value (4.36) for the case considered here of temporally uncorrelated fading amplitudes. It can be seen that for N_B large enough, the performance of the MA-F estimator reaches the HCRB (not shown in this figure for clarity,

see following examples) as expected from the analysis in Sec. 5.5. Moreover, in order to demonstrate that the linear channel estimator proposed here is not impaired by the threshold effects, typical of non-linear estimation problems, fig. 5.11 shows the normalized MSE (i.e., $E[||\hat{\mathbf{h}}_\ell - \mathbf{h}_\ell||^2]/E[||\mathbf{h}_\ell||]$) of a structured channel estimator based on the direct computation of multipath delays [4] [5]. The choice of a channel with a single delay implies that the ML delay estimator is $\hat{\tau} = \arg\max_\tau \mathbf{g}(\tau)^T (\sum_{\ell=1}^{N_B} \mathbf{h}_{UML,\ell}\mathbf{h}_{UML,\ell}^H)\mathbf{g}(\tau)$ (recall that for ideal training sequence, the additive noise on $\mathbf{h}_{UML}(\ell)$ is white) and the block-by-block amplitude estimate is $\hat{\beta}_\ell = (\mathbf{g}^T(\hat{\tau})/||\mathbf{g}(\hat{\tau})||^2) \cdot \mathbf{h}_{UML,\ell}$ [4]. Even though this constitutes a privileged scenario for delay estimation (as there are no resolution issues), the threshold effect causes the structured method to be outperformed by the proposed technique for $N_B > 10$ and sufficiently small $SNR's$ as shown in fig. 5.11. For higher SNR's the structured method attains the MSE bound. As expected, the unstructured technique has a slower convergence owing to the larger number of long term parameters that have to be estimated from the multiblock measurements.
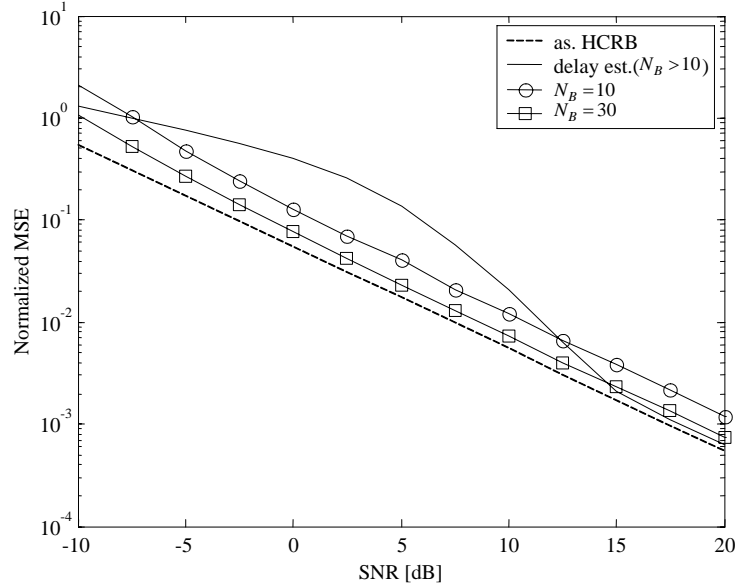

Figure 5.11: Normalized MSE of MA-F compared to the asymptotic HCRB and the MSE of a structured channel estimator based on the direct computation of delays (SISO channel).

Effects of residual timing offset

To make the multiblock approach of the proposed estimator valid in practice, the received signals $\{\mathbf{y}_\ell\}_{\ell=1}^{N_B}$ should be sampled with synchronized timing in each block in order to have equal multipath delays. Here we investigate the effect of timing errors characterized by a random offset independently selected in each block. The delays are selected as $\boldsymbol{\tau} = [5.1, 6.2, 6.8, 9.8]T$ and the power-delay profile is $\boldsymbol{\Omega} = \Omega_0 \, \mathrm{diag}\{[0, -3, -6, -9][dB]\}$ (Ω_0 is scaled to
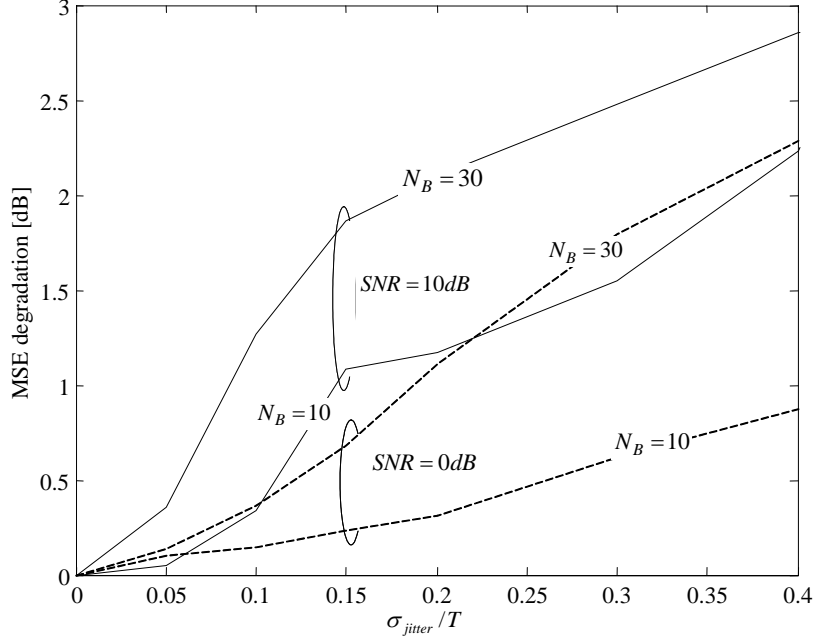
Figure 5.12: MSE degradation due to a residual random timing offset with standard deviation σ_{jitter} (SISO channel).

ensure the normalization (2.35)) so that $r_T = 4$. The matrix $\mathbf{G}(\boldsymbol{\tau})$ is modified by adding a time misalignment $\Delta\tau_\ell$ that is independent from block to block: $\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{g}(\tau_1 + \Delta\tau_\ell) \cdots \mathbf{g}(\tau_d + \Delta\tau_\ell)]$ with $\Delta\tau_\ell \sim \mathcal{N}(0, \sigma_{jitter}^2)$. Even though it is reasonable to assume that the variables $\Delta\tau_\ell$ are generally correlated (as it happens for instance due to the mismatch between the transmitter and receiver clocks), the simple model considered here constitutes an interesting worst-case scenario. Fig. 5.12 shows the MSE degradation compared to the MSE for $\sigma_{jitter} = 0$ versus the timing error normalized to the symbol interval (σ_{jitter}/T) for $SNR = 0, 10dB$. The modal analysis is robust with respect to random timing error as the MSE degradation can be quantified to be less than $3dB$ for $\sigma_{jitter}/T < 0.4$, $N_B \leq 30$ bursts and $SNR < 10dB$. Notice that since the gain of modal analysis is approximately $r_T/W \simeq 5.7dB$ with respect to UML estimation, it can be concluded that the proposed techniques is still advantageous for a wide range of timing offset standard deviations.

5.6.2 Example 2: SIMO channel (beamforming scenario)

Simulation setting

In this example, a SIMO system ($N_T = 1, N_R = 8$) within a beamforming scenario is considered. In particular, the receiver is equipped with a uniform linear antenna array of $N_R = 8$ elements with half-wavelength inter-element spacing and the temporal support of the channel is $W = 15$ symbols. The space-time channel is characterized by $d = 8$ paths

with with power-delay profile $\Omega_i = \Omega_0 \times (0.5)^{(i-1)}$ (Ω_0 is scaled so as to ensure channel normalization (2.35)). The paths are obtained from two main clusters, each corresponding to four paths with the same AOA and different delays: the first set of paths (first cluster) is characterized (where not stated otherwise) by the AOA $\alpha_i = \pi/3$ for $i = 1, \ldots, 4$ and the delays $[\tau_1 \cdots \tau_4] = [3.2, 5.1, 6.2, 6.8]T$ whereas the second has AOA $\alpha_i = \pi/6$ for $i = 5, \ldots, 8$ and delays $[\tau_5 \cdots \tau_8] = [9.8, 11.1, 11.9, 12.8]T$. The transmitted pulse $g(t)$ is a raised cosine with roll-off factor 0.2. The training sequence, optimally designed, has length $L_P = 40$ (larger than the minimum $L_P \geq W = 15$).

As in the previous example, fading amplitudes are assumed to be uncorrelated from block to block, $\varphi_i(m) = \delta(m)$, $i = 1, ..., d$, therefore the investigation focuses on the S/T MA-F estimator. The joint space-time modal analysis (MA-F estimator) will not be considered here for its poor convergence performance (see Example 4 below for further details). However, as practical alternatives to S/T MA-F, S MA-F and T MA-F will be considered.

Effect of spatially correlated noise (interference)

Here the aim is the evaluation of the performance of the estimator for spatially correlated noise. This models, the presence of out-of-cell interferers or in-cell interference for non-orthogonal signaling. We consider five interferers with AOA's $[\bar{\alpha}_1 \cdots \bar{\alpha}_5] = [-\pi/3, -\pi/6, 0, \pi/6, \pi/3]$ equally spaced within the angular support $[-60 \ 60] \deg$. They are modelled as Gaussian disturbance so that the noise correlation matrix reads

$$\mathbf{R}_n = (\sigma_i^2/5) \sum_{k=1}^{5} \mathbf{a}(\bar{\alpha}_k)\mathbf{a}(\bar{\alpha}_k)^H + \sigma_w^2 \mathbf{I}_M \tag{5.43}$$

and therefore the noise power is the sum of the contributions from white noise and interference, $\sigma_n^2 = \sigma_i^2 + \sigma_w^2$. Notice that in this case the signal to noise ratio SNR has to be interpreted as signal to noise plus interference ratio. Here we let $P/\sigma_w^2 = 50dB$ and let the SNR vary by modifying the power of interferers, i.e., the ratio P/σ_i^2.

In fig. 5.13 the simulations for the MSE on the channel estimate (normalized on the channel norm $E[|||\mathbf{h}_\ell|||^2] = N_R N_T = 8$) of S/T MA-F, S MA-F and T MA-F (markers) are compared with the analytical asymptotic MSE (5.36), (5.39) and (5.40) (dashed lines), respectively. As a reference, the performance of the UML estimator both from simulation and analysis (5.32) are shown. The number of blocks N_B is selected so as to guarantee convergence to the asymptotic results (in practice, here $N_B > 30$ is enough, see also Example 4). As for the top part of fig. 5.13, the angles of arrival of the user are aligned with those of the interferers: $\alpha_i = \bar{\alpha}_5 = \pi/3$ for $i = 1, \ldots, 4$, and $\alpha_i = \bar{\alpha}_4 = \pi/6$ for $i = 5, \ldots, 8$. At the bottom, the angles of arrival are slightly misaligned: $\alpha_i = \pi/4$ for $i = 1, \ldots, 4$, and $\alpha_i = \pi/8$ for $p = 5, \ldots, 8$ (see the box on both figures). Here, it is assumed that the number of modes is correctly estimated, $\hat{r}_S = r_S = 2$ and $\hat{r}_T = r_T = 8$. When the user angles are separated from those of the interferers (bottom), the spatial processing performed by the S/T MA-F and

S MA-F methods leads to a value of MSE that is independent on the interference level, but it is ruled by the background white noise. According to the analysis in Sec. 5.6, here the S MA-F method outperforms the T MA-F because in this propagation environment the spatial gain ($N_R N_T / r_S = 4$) is larger than the temporal gain ($W/r_T = 1.875$). To be precise, this argument applies to the case of spatially white noise, but it turns out to be a reliable rule of thumb to choose the modal estimators according to their expected performance in that case. The following simulations consider the angles of arrival for the user aligned with those of the interferers (top part of fig.5.13) as this corresponds to the worst case.

Estimate of the number of modes

The simulations presented above were obtained for an unbiased estimate by choosing the degree of diversity $\hat{r}_S = r_S = 2$ and $\hat{r}_T = r_T = 8$. However, it is well known that it may be advantageous for low SNR's to underparametrize the model of interest (i.e., to select a smaller number of modes) in order to trade some bias for a lower MSE [17]. This is shown in fig. 5.14 for the S/T MA-F method with $N_B = 10$ (similar results can be shown for all the methods). For simplicity, the number of spatial modes is fixed to $\hat{r}_S = r_S = 2$ while $\hat{r}_T$ ranges from $\hat{r}_T = 1$ to $\hat{r}_T = r_T = 8$ (dashed line). The normalized MSE versus SNR curves show that for $SNR < -10dB$ the MSE is minimized by choosing $\hat{r}_T = 1$, while for larger SNR values the MSE shows a floor due to the bias for under-parameterization. Larger values of $\hat{r}_T$ have to be used for increasing SNR in order to select the number of modes that minimizes the MSE. The performance of S/T MA-F with number of modes $\hat{r}_T$ selected according to the MDL principle [9] is shown in bold line. It can be concluded that MDL provides a rank estimation that approximately minimizes the MSE (see also [17]).

Subspace tracking versus batch modal analysis

Here the performance degradation of the computationally efficient subspace tracking implementation of modal analysis as compared to the batch approach is investigated. Toward this goal, in fig. 5.15 the performances of S/T MA-F through the subspace tracker proposed in [12] is evaluated in terms of MSE versus the number of blocks N_B. The exact batch implementation (dashed lines) and the subspace tracking implementation (bold lines) are compared for varying $SNR's$ (forgetting factor $\gamma = 1$, the number of modes is assumed to be known). The MSE corresponding to the UML estimate is also shown as a reference (dash-dotted lines) and confirms the accuracy of the subspace tracking methods for real time implementations.

5.6.3 Example 3: MIMO system (diversity scenario)

Simulation setting

In this example, we consider a MIMO system within a diversity scenario. The number of transmit antennas is $N_T = 4$ and the number of receiving antennas is $N_R = 4$. The multipath
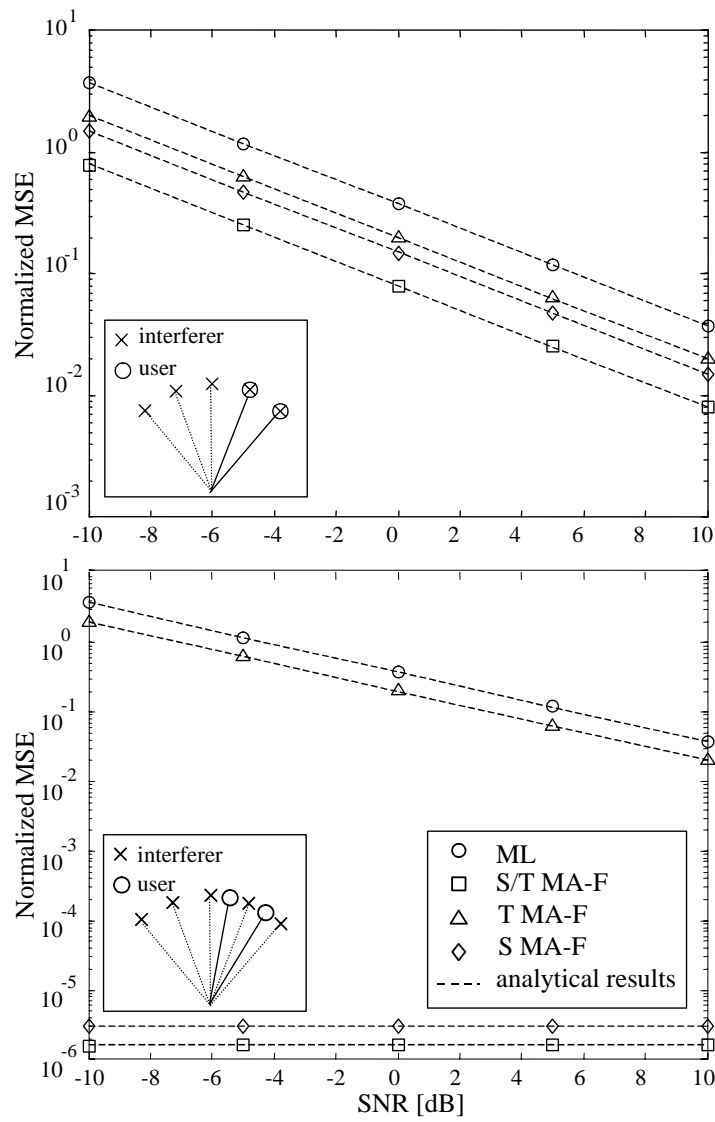
Figure 5.13: Normalized MSE of the S/T MA-F, S MA-F and T MA-F estimators for two different spatial configurations of user and interferers (SIMO channel, beamforming scenario).
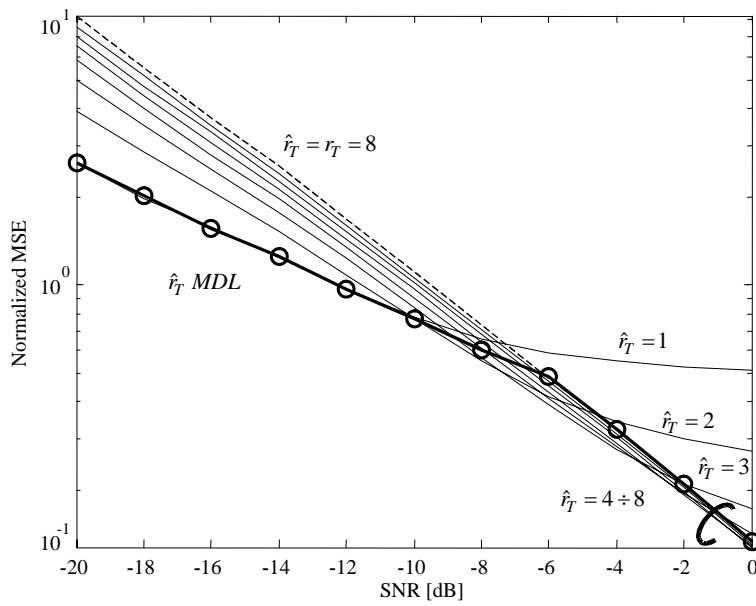
Figure 5.14: MSE of S/T MA-F versus SNR for different values of $\hat{r}_T = 1 \div r_T = 8$ and for $\hat{r}_T$ selected according to the MDL criterion (SIMO channel, beamforming scenario).
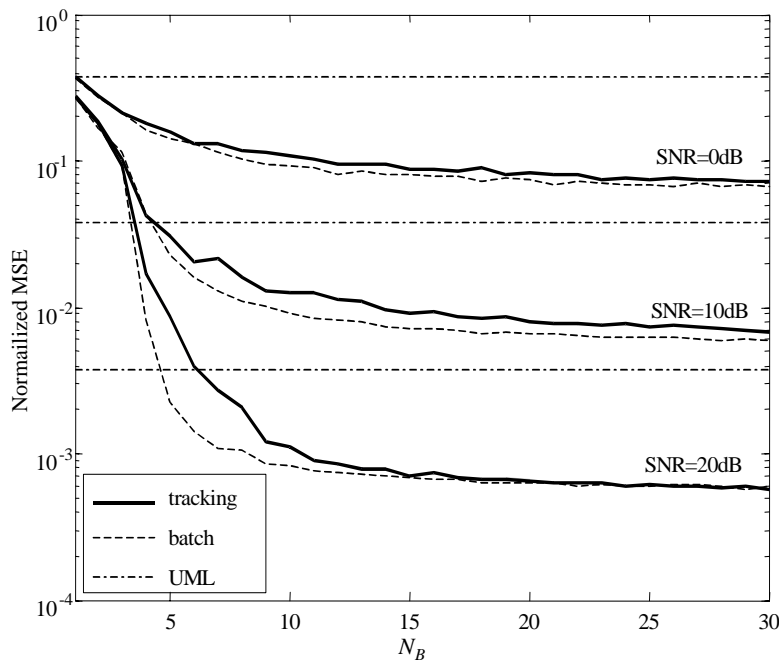

Figure 5.15: Normalized MSE versus the number of blocks N_B for the batch and the subspace tracking implementation of S/T MA-F (SIMO channel, beamforming scenario).

channel is characterized by a temporal support $W = 8$ samples and the number of paths is $d = 4$ with uniform power-delay profile and sample spaced delays $\tau_i/T = (i-1)$, $i = 1, ..., 4$. The paths have equal uniform Doppler spectrum selected according to the Clarke's model and equal separable spatial correlation $\mathbf{R} = \mathbf{R}^{(T)} \otimes \mathbf{R}^{(S)}$ where the transmit and receive side correlations $\mathbf{R}^{(T)}$ and $\mathbf{R}^{(S)}$ are assumed to be obtained from an autoregressive model so that they are Toeplitz matrices with first column $[1 \; \rho_T \cdots \rho_T^{N-1}]^T$ and $[1 \; \rho_R \cdots \rho_R^{N-1}]^T$ respectively. The correlation coefficients $0 \le \rho_T, \rho_R \le 1$ measure the transmit and receive side spatial correlation respectively. The noise is spatially white and the training sequences optimally designed with length $L_P = N_T W = 32$.

Effect of Doppler spread and spatial correlation

The asymptotic MSE of the MA-F and the MSE of the UML estimators are evaluated and compared with the asymptotic HCRB (4.36) in fig. 5.16 for varying normalized Doppler spread $f_D T_S$ and $SNR = 0, 10, 20 dB$ (upper part) and as a function of fading correlation $\rho_T = \rho_R$ for $f_D T_S = 0.1, 0.5$. As already discussed, performance degradation of these techniques occurs as these estimators fail to fully exploit the information on the MIMO channel matrix. Indeed the UML estimator does not use any deterministic or statistical information whereas the MA-F only capitalize on structural modelling of the channel (i.e., does not track the fading amplitudes). However when the information neglected by the estimator is not helpful in improving the performance, no degradation is expected. In this regard, consider the upper part of fig. 5.16. For increasing $f_D T_S$, i.e., increasingly uncorrelated fading amplitudes across different bursts, no benefits can be obtained by tracking the amplitudes and, as a consequence, the degradation of the two estimators decreases. Similarly, a decreasing spatial (ρ_T) or temporal (ρ_R) correlation renders the MMSE approach of the optimum estimator increasingly ineffective and the degradation decreases, as shown in the lower part of fig. 5.16. An increased SNR would make this effect less noticeable (not shown in the figure). The impact of the SNR, studied in the upper part of fig. 5.16 can be interpreted in the same way: an increasing SNR makes the MMSE approach of the optimum estimator closer in performance to a UML approach and, as a consequence, the degradation decreases.

The results presented above can be interpreted in the light of the discussion of Sec. 4.9 about the relationship between the asymptotic HCRB and the conventional CRB. In fact, the MA-F estimator, being a ML estimate, attains the conventional CRB for $N_B \to \infty$ (asymptotically efficient estimation [8]) and the latter coincides with the asymptotic HCRB for *i)* $N_B \to \infty$, *ii)* high SNR, *iii)* temporally (and spatially for a diversity scenario) uncorrelated fading.
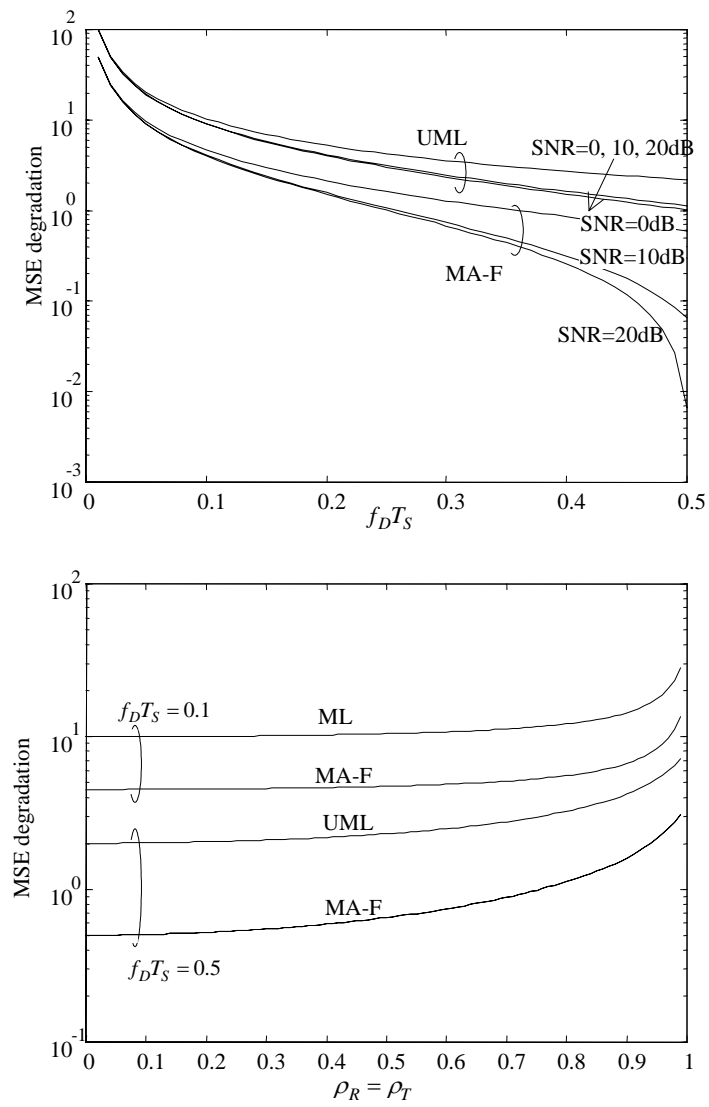
Figure 5.16: MSE degradation of UML and MA-F channel estimators as compared to the asymptotic hybrid CRB versus the Doppler spread (top) and the spatial correlation (bottom) (MIMO channel, diversity scenario).

5.6.4   Example 4: MIMO system (beamforming scenario)

Simulation setting

The performance of the proposed channel estimation methods is evaluated herein for a MIMO system within a beamforming environment. $N_T = 4$ transmitting antennas and $N_R = 4$ receiving antennas (half-wavelength spaced linear antenna arrays at both sides) communicate over a frequency selective channel with temporal support $W = 8$. The length of the training sequences, optimally designed, is $L_P = N_T W = 32$. Where not stated otherwise, the channel is characterized by $d = 4$ paths with uniform power delay profile, delays $\boldsymbol{\tau} = [1\ 3\ 4\ 6]T$, DODs and DOAs equally-spaced in the angular support $(-60, 60)\deg$. With these choices, all the paths are resolvable in both temporal and spatial domain and therefore the number of modes equals the number of paths, i.e., $r = 4$, $r_S = 4$ and $r_T = 4$. Moreover, Doppler spectra of different path are equal and selected according to the Clarke's model with normalized Doppler spread $f_D T_S = 0.03$. Validation of the proposed channel estimation algorithms for the channel model standardized by the 3GPP/3GPP2 SCM (spatial channel modeling) adhoc group [25] (that falls within the category of beamforming models) has been presented in [26].

Estimation of the number of modes and comparison between different techniques

Figure 5.17 compares the performance of the S/T MA-F and MA-F channel estimators implemented through subspace tracking with rank estimation in terms of MSE versus SNR with the UML channel estimate and the asymptotic HCRB. For the S/T MA-F, both the asymptotic results derived in Sec. 5.5 and the MSE obtained after $N_B = 50$ blocks through subspace tracking (with $\gamma = 0.999$, $\beta = 0.6$, $r_{S,\max} = 15$ and $r_{T,\max} = 7$) are shown (perfomance as a function of N_B is considered next). On the other hand, the performance of MA-F is plotted only in terms of the asymptotic MSE (5.35) given its prohibitive complexity and rate of convergence (see below for details). The lower bound is shown for the case $f_D T_S = 0.5$ (where it coincides with the conventional CRB) in order to confirm the analytical considerations presented in Sec. 5.5. The upper part of the figure shows the estimated number of spatial and temporal modes after $N_B = 50$ blocks.

Observing the figure from the upper curve, we can make the following considerations. *i)* The S/T MA-F estimator outperforms (asymptotically) the UML channel estimate by a factor $N_R N_T W/(r_S r_T) \simeq 9dB$ as expected from the analysis in Sec. 5.5. *ii)* For $N_B = 50$ blocks, the S/T MA-F essentially converges to the asymptotic perfomance. Moreover, for small SNR ($SNR < 0dB$) the number of modes is underestimated and this has the effect of lowering the MSE as compared to the asymptotic result (valid for large SNR). This phenomenon can be explained by noticing that reducing the number of modes implies trading a bias in the estimate (due to the underparametrization) for a reduced variance [6] (see also Example 2). *iii)* The asymptotic MSE of the MA-F technique coincides with the asymptotic HCRB for large SNR,
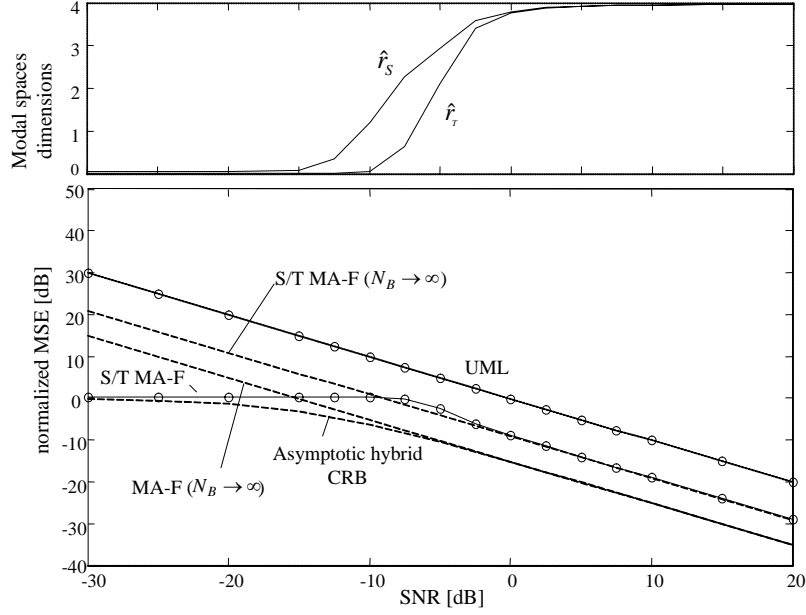
Figure 5.17: Normalized MSE of the S/T MA-F technique versus SNR for $N_B = 50$ compared to the limit performance given by the asymptotic hybrid CRB. The upper part of the figure shows the estimated number of spatial and temporal modes (MIMO channel, beamforming scenario).

as anticipated in Sec. 5.5. *iv*) The S/T MA-F technique has an asymptotic degradation of $r/(r_S r_T) = 6dB$. *v*) The qualitative behavior of the asymptotic HCRB for small SNR is analogous to that discussed above of S/T MA-F. In this case, the underparametrization of the channel (and consequent trade-off between bias and variance) is automatically performed by MMSE filtering. Accordingly, for very small SNR , the asymptotic HCRB tends to the channel norm (recall (4.39)).

Convergence to the asymptotic MSE

The MSE as a function of the number of processed blocks N_B for $SNR = 5dB$ is shown in fig. 5.18 for MA-F, S/T MA-F, MA-LMS and S/T MA-LMS. It can be seen that LMS amplitude tracking (with $\mu = 0.8$) allows a gain of approximately $2dB$ with respect to modal filtering. Moreover, convergence of the perfomance of S/T MA-F to the asymptotic value (5.35) is obtained (by a fraction of dB) for $N_B > 50$. On the other hand, M MA-F has a very slow convergence and becomes advantageous with respect to S/T MA-F only for $N_B > 80$. Finally, suboptimality of S/T MA-LMS can be quantified by approximately $5dB$ through comparison with the asymptotic HCRB $MSE_{\hat{\mathbf{h}}}$.
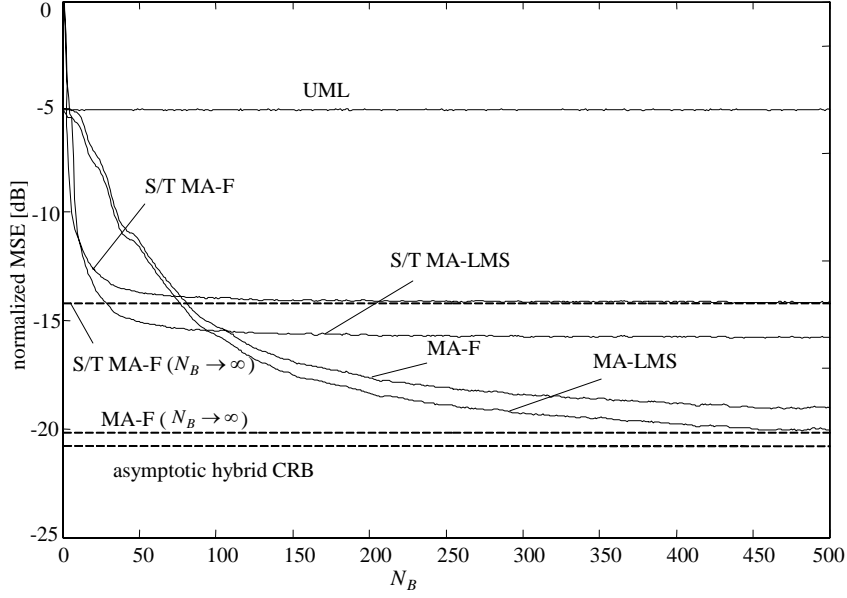
Figure 5.18: Normalized MSE of the S/T MA-MF technique versus N_B for $SNR = 5dB$ compared to the limit performance given by the asymptotic hybrid CRB (MIMO channel, beamforming scenario).

5.7 Extension to multiuser systems

The extension of the channel estimation algorithms presented above to a multiuser setting is straightforward for both downlink and uplink if the training sequences of different users are appropriately designed, i.e., if (recall the signal model in Sec. 3.4)

$$\mathbf{X}^{(n)H}\mathbf{X}^{(k)} \simeq \mathbf{0}, n \neq k. \tag{5.44}$$

Indeed, in this case, channel estimation corresponding to each user can be obtained (by the base station in uplink or by the users in downlink) separately, according to the algorithms presented above, with no additional interference. This condition on the training sequences, as already discussed is (approximately) met in both time-domain and frequency-domain transmission current standards. However, in order to account for possible suboptimality of training sequences, a novel strategy can be employed that reduces to the separate estimation of channels from different users when condition (5.44) is met. This strategy is here derived for MA-F and uplink since the generalization to other settings follows the same lines.

Recalling the signal model (3.32), the signal received by the base station can be written again as

$$\mathbf{y}_\ell = \mathbf{X}\mathbf{h}_\ell + \mathbf{n}_\ell, \tag{5.45}$$

where the channel vector and the training matrix contain the contributions from all users, i.e., $\mathbf{h}_\ell = [\mathbf{h}_\ell^{(1)T} \cdots \mathbf{h}_\ell^{(K)T}]^T$ and $\mathbf{X} = [\mathbf{X}^{(1)} \cdots \mathbf{X}^{(K)}]$. With these definitions, the negative log-likelihood function can thus still be written as (5.19). In order to separate the contribution

of each user in the likelihood (5.19), we partition the Cholesky factorization of the whitening matrix $\mathbf{Q}_{UML}$ into blocks $\mathbf{Q}_{ij}$ of size $N_R n_T^{(i)} W \times N_R n_T^{(j)} W$:

$$\mathbf{Q}_{UML}^{-H/2} = \begin{bmatrix} \mathbf{Q}_{11} & \cdots & \mathbf{Q}_{K1} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{Q}_{KK} \end{bmatrix} \tag{5.46}$$

yielding

$$\mathcal{L}(\mathbf{U}, \mathbf{D}) = \sum_{k=1}^{K} \sum_{\ell=1}^{N_B} \left\| \mathbf{Q}_{kk} \mathbf{U}^{(k)} \mathbf{d}_\ell^{(k)} - \left[\mathbf{Q}_{kk} \mathbf{h}_{UML,\ell}^{(k)} + \sum_{i=k+1}^{K} \left(\mathbf{Q}_{ki} \mathbf{h}_{UML,\ell}^{(i)} - \mathbf{U}^{(i)} \mathbf{d}_\ell^{(i)} \right) \right] \right\|^2. \tag{5.47}$$

The optimization of $\mathcal{L}(\mathbf{U}, \mathbf{D})$ in a closed form is not an easy task but the analysis of (5.47) suggests an approximated solution similar to the successive interference cancellation in multiuser detection [27]. This solution is obtained by minimizing separately the terms corresponding to each user, starting with the Kth user down to the first. The iterative scheme is (for $k = K, K-1, ..., 1$) (recall (5.23))

$$\hat{\mathbf{h}}_\ell^{(k)}(\ell) = \mathbf{Q}_{kk}^{-1} \widehat{\mathbf{\Pi}}^{(k)} \tilde{\mathbf{h}}_{UML,\ell}^{(k)} \tag{5.48}$$

where the pre-whitening accounts for the interference cancellation

$$\tilde{\mathbf{h}}_{UML,\ell}^{(k)} = \mathbf{Q}_{kk} \mathbf{h}_{UML,\ell}^{(k)} + \sum_{i=k+1}^{K} \mathbf{Q}_{ki} \left(\mathbf{h}_{UML,\ell}^{(i)} - \hat{\mathbf{h}}_\ell^{(k)} \right), \tag{5.49}$$

recall that $\widehat{\mathbf{\Pi}}^{(k)}$ is the projector onto the subspace spanned by the $\hat{r}^{(k)}$ principal eigenvectors of the correlation matrix $N_B^{-1} \sum_{\ell=1}^{N_B} \tilde{\mathbf{h}}_{UML,\ell}^{(k)} \tilde{\mathbf{h}}_{UML,\ell}^{(k)H}$. Let us remark that if the training sequences are mutually uncorrelated, the multiuser algorithm described by (5.48)-(5.49) reduces to the single-user case to be carried out on a user by user basis. In this case (which is closely approximated by third generation standards [28]), the multiuser approach does not increase the computational complexity of channel estimation per user. In addition, the near-far effect or the different correlation properties of the training sequences could be taken into account by sorting the users and thus optimizing the performances of this iterative scheme.

Further analysis on this iterative scheme can be found in [29].

5.8 Application 1: MIMO-OFDM system with BICM and Turbo-Equalization

In order to show the effectiveness of the proposed channel estimation technique in terms of probability of error of a practical system, here we consider a MIMO-OFDM system based on Bit Interleaved Coded Modulation [30] and Turbo-Equalization [31]. This combination has been recently recognized as a promising candidate for next generation wireless LAN [33].

With reference to fig. 3.4, the block diagram of the MIMO encoder/modulator for the considered system is shown in fig. 5.19-(a). The information bit stream d_m is passed through a convolutional encoder producing the encoded bits c_k, then interleaved (interleaved bits are denoted as b_k) and finally modulated into a M-QAM constellation and blocked into $L_D \times 1$ vectors $\mathbf{x}_\ell^{(n_T)}$ ($n_T = 1, ..., N_T$) to be transmitted by different transmitting antennas. No attempt of optimizing the interleaving operation over different transmitting antennas and frequencies is made here.

The receiver is depicted in fig. 5.19-(b). Channel estimation is not included in the block diagram for simplicity. The received signals on each data frequency $\mathbf{y}_\ell[k]$ ($k = 1, ..., L_D - 1$) are processed separately by a Soft Input Soft Output (SISO) BLAST decoder [31]. This performs MMSE linear filtering of the input signals and, based on the gaussian approximation of the residual interference, computes the a posteriori log-likelihood ratios (LLRs) $\Lambda_1(\mathbf{b}_k)$ of the encoded and interleaved bits $\mathbf{b}_k$ ($k = 0, ..., L_D - 1$). The $N_T \log_2 M \times 1$ vector $\mathbf{b}_k$ represents the encoded and modulated bits that are mapped onto the symbols collected in the $N_T \times 1$ vector $\mathbf{x}_\ell[k]$ transmitted on the kth subcarrier. After subtraction of the a priori LLR $\lambda_1(\mathbf{b}_k)$ (obtained from the SISO decoder, as explained below, and initialized to zero for the first iteration), the so obtained extrinsic LLR $\lambda_1^E(\mathbf{b})$ (where $\mathbf{b} = [\mathbf{b}_0^T ... \mathbf{b}_{L_D-1}^T]^T$) are deinterleaved producing the a priori LLR $\lambda_2(\mathbf{c})$ for the SISO decoder ($\mathbf{c}$ is the $L_D N_T \log_2 M \times 1$ vector obtained by deinterleaving of $\mathbf{b}$). In a similar way, from the a posteriori LLR $\Lambda_2(\mathbf{c})$ produced by the decoder, the a priori LLR $\lambda_2(\mathbf{c})$ are subtracted, yielding the extrinsic LLR $\lambda_2^E(\mathbf{c})$ that, interleaved, provide the a priori LLR for the SISO BLAST equalizer.

5.8.1 Simulation results

The proposed MIMO-OFDM system with the following specific features is tested: the convolutional encoder has rate $R = 1/2$ and generators $[7, 5]$, the interleaver is random, the modulation is $4-$QAM ($M = 4$) and the SISO decoder is log-MAP [32]. The BER of such a system is plotted in fig. 5.20 for the ideal case of perfect knowledge of the channel, for a UML channel estimate and for a S/T MA-F estimator (with $N_B = 50$ as in fig. 5.17) as a function of the equalization-decoding iteration (after the third only minor improvement are obtained). For a BER equal to 10^{-3}, S/T MA-F guarantees approximately $7dB$ gain as compared to the UML channel estimate and is only $3dB$ away from the case of ideal channel knowledge.

5.9 Application 2: Uplink of a TD-SCDMA system with a ZF block linear equalizer

The perfomance of S/T MA-F are validated here for the standard for 3rd generation cellular system TD-SCDMA [28]. We consider the uplink with a ZF block linear equalizer at the base station [34]. The latter is equipped with $N_R = 8$ antennas (inter-element spacing $\lambda/2$) whereas
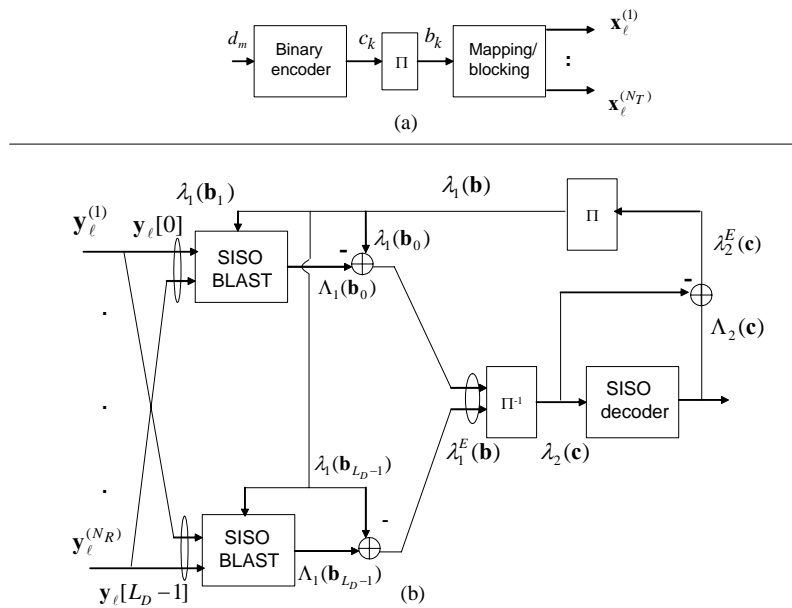
Figure 5.19: Block diagram of the transmitter (a) and receiver (b) of a MIMO-OFDM system basted on BICM and turbo equalization.
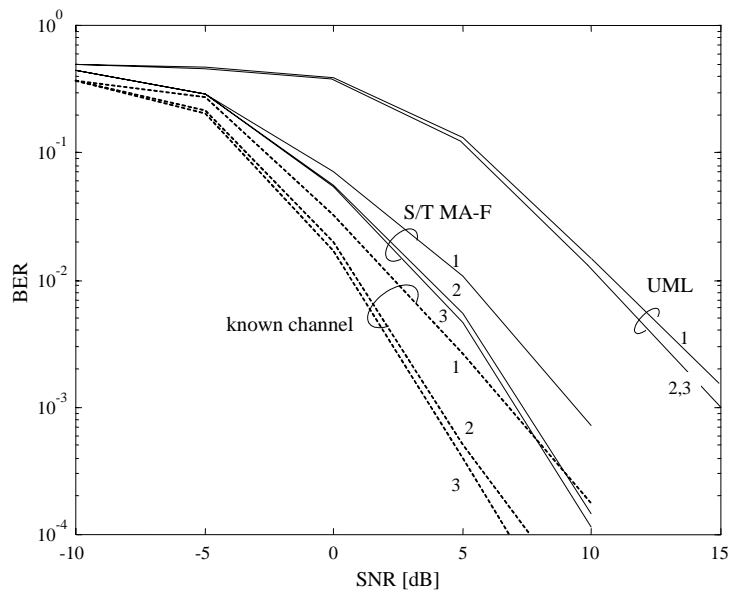

Figure 5.20: BER for a MIMO-OFDM system based on BICM and Turbo Equalization with channel estimation error for S/T MA-F ($N_B = 50$) and UML estimators (MIMO channel, beamforming scenario).

each user has $n_T^{(k)} = 1$ antenna. According to the specification , the parameters of interest are: $T_S = 10$ ms, $W = 16$, $L_P = 129$, roll-off equal to 0.22, carrier central frequency 1950 MHz. Moreover, each one of the $K = 8$ users transmits at the same rate using a spreading factor $Q = 16$: the data field thus contains 22 QPSK information symbols ($L_D = 352$ chips). The K training sequences are chosen according to the standard specifications. We consider a simple radio environment characterized by two clusters with angles randomly chosen within the angular support $[-60, 60]$ deg, four paths per cluster ($d^{(k)} = 8$) with angular dispersion of 5 deg and an exponential power profile: $\Omega_i = \Omega_0(0.5)^{(i-1)}$ (Ω_0 is scaled to ensure the normalization (2.35)) (paths $i = 1, .., 4$ correspond to the first cluster and paths $i = 5, .., 8$ to the second). In each cluster the delays are chip interval (T) spaced and the first delay is randomly selected in the interval $[2, 10]\, T$. The fading variation is simplified by assuming the same velocity v for all the users according to the Clarke's model. The covariance matrix of noise is (5.43) with 6 out-of-cell terminals with angles $\bar{\alpha}_k$ uniformly spaced in $[-60 \ 60]$ deg. The number of modes ($r_S^{(k)}$ and $r_T^{(k)}$) are estimated here by using the MDL criterion.

In fig. 5.21 the performance of the S/T MA-F algorithm is evaluated in terms of BER and MSE of the channel estimate. The modal analysis estimate for $N_B = 10$ is compared to the single-blocks techniques UML and Reduced Rank (RR, [17]). For any mobility of the terminal, from the pedestrian environment ($v = 3$ km/h), up to the vehicular ($v = 120$ km/h), the performance of the S/T MA-F method are the same. On the other hand, for a static channel ($v = 0$ km/h) the S/T MA-F technique performs better since the number of modes r_S and r_T can be reduced (see ref. [6]). Fig. 5.21 (upper figure) shows that for $N_B = 10$ and time-varying channel ($v \geq 3$ km/h) the MSE is very close to the asymptotical MSE. Compared to the ZF block linear equalizer based on the UML channel estimate, the modal analysis method shows a meaningful advantage in term of SNR (approximately 3-4 dB) that is practically independent on the variation of the faded amplitudes (lower figure). In addition, the loss with respect to the ideal case of known channel, or perfect CSI, (dashed line) is approximately $1.5dB$.

5.10   Conclusion

Based on the insight obtained from the analysis performed in the previous Chapter, practical channel estimators have been designed that perform close to the performance limit set by the HCRB. The proposed channel estimators compute the long term features by identifying the invariant, over multiple blocks, space-time modes of the channel (*modal analysis*). Modal analysis can be performed either through a batch or an adaptive approach based on subspace tracking. On the other hand, the fast varying fading amplitudes are either estimated block-by-block (*modal filtering*) or possibly tracked by using least squares techniques that exploit temporal (i.e., over multiple blocks) correlation of the fading process (*LMS tracking*). For the beamforming scenario, an alternative parametrization of the space-time modes into decoupled
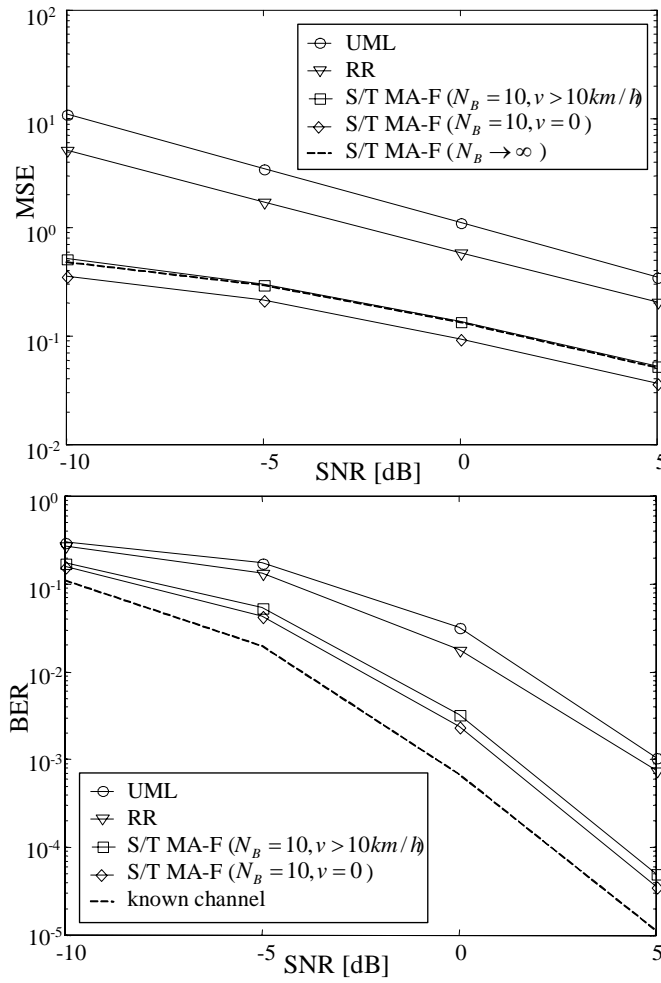
Figure 5.21: Normalized MSE (upper figure) and BER (lower figure) for the uplink of a TD-SCDMA standard system versus SNR for single-block techniques (UML and RR) and for S/T MA-F with $N_B = 10$ slot. The BER for known channel is shown as reference (multi-access MIMO channel, beamforming scenario).

spatial and temporal modes has been advocated and, accordingly, *decoupled spatial and temporal modal analysis* proposed.

Thorough numerical investigation has validated the performance of the proposed techniques in different propagation scenarios and systems. Moreover, the extension to multiuser system has been considered. Finally, two applications have been disclosed, namely MIMO-OFDM with Bit Interleaved Coded Modulation and the uplink of a TD-SCDMA system, in order to evaluate the impact of channel estimation on BER.

Remark 8 *The investigation of modal analysis and filtering through subspace tracking for uplink combined with transmit beamforming for downlink has been studied in [35] in the context of CDMA systems.*

5.11 Appendix-A: estimation of the spatial noise covariance matrix

From the signal models (3.6) (time-domain transmission) and (3.22) (frequency-domain transmission), the noise matrix $\mathbf{N}_\ell$ can be estimated, once the UML channel estimate has been computed, as

$$\hat{\mathbf{N}}_\ell = \mathbf{Y}_\ell - \breve{\mathbf{B}}_{UML,\ell}\boldsymbol{\mathcal{X}}^T, \tag{5.50}$$

where $\breve{\mathbf{B}}_{UML,\ell}$ reads $\breve{\mathbf{H}}_{UML,\ell}$ or $\breve{\mathbf{F}}_{UML,\ell}$ (i.e., the channel matrices estimated according to UML) for time and frequency-domain transmission respectively. Then, an estimate of the noise spatial correlation matrix can be evaluated by the following sample average:

$$\hat{\mathbf{R}}_n = \frac{1}{N_B L_P} \sum_{\ell=1}^{N_B} \hat{\mathbf{N}}_\ell \hat{\mathbf{N}}_\ell^H. \tag{5.51}$$

This estimate can be easily shown to be consistent under the condition that $rank(\hat{\mathbf{N}}_\ell) = rank(\mathbf{R}_n) = N_R$. But, recalling that $\breve{\mathbf{B}}_{UML,\ell} = \mathbf{Y}_\ell\boldsymbol{\mathcal{X}}^*(\boldsymbol{\mathcal{X}}^*)^\dagger$, the estimate (5.50) can be stated as $\hat{\mathbf{N}}_\ell = \mathbf{Y}_\ell(\mathbf{I}_{L_P} - \boldsymbol{\Pi}_{\boldsymbol{\mathcal{X}}^*}) = \mathbf{Y}_\ell\boldsymbol{\Pi}_{\boldsymbol{\mathcal{X}}^*}^\perp$. Therefore, the condition stated above entails $L_P - rank(\boldsymbol{\mathcal{X}}^*) \geq N_R$ so that

$$L_P \geq N_T W + N_R. \tag{5.52}$$

In other words, while for channel estimation only is enough to allocate $L_P \geq N_T W$ training (or pilot) symbols per block, when the spatial noise covariance matrix is of interest, condition (5.52) should be satisfied.

5.12 Appendix-B: asymptotic optimality of S/T MA-F

The exact ML estimation of spatial modes should be obtained by evaluating the first r_S eigenvectors of the sample correlation matrix $1/N_B \sum_{\ell=1}^{N_B} \tilde{\mathbf{H}}_{UML,\ell}\boldsymbol{\Pi}_T\tilde{\mathbf{H}}_{UML,\ell}^H$ (and similarly for

the temporal modes). However, the proposed S/T MA-F estimator computes this estimate as the first r_S eigenvectors of the sample correlation matrix $1/N_B \sum_{\ell=1}^{N_B} \tilde{\mathbf{H}}_{UML,\ell} \tilde{\mathbf{H}}_{UML,\ell}^H$. Let us for simplicity consider spatially white noise and optimal designed training sequences. The proof can be easily extended to the more general case of correlated noise on the UML estimate. When $N_B \to \infty$ for ergodicity (recall the discussion in Sec. 5.2 and in particular equation (5.14)) we have

$$\frac{1}{N_B} \sum_{\ell=1}^{N_B} \tilde{\mathbf{H}}_{UML,\ell} \mathbf{\Pi}_{\tilde{\mathbf{U}}_T} \tilde{\mathbf{H}}_{UML,\ell}^H \to E[\tilde{\mathbf{H}}_{UML,\ell} \mathbf{H}_{UML,\ell}^H] = \mathbf{S}_S \check{\mathbf{\Gamma}}_S \mathbf{S}_S^H + \frac{\sigma_n^2}{L_P P} \mathbf{I}, \qquad (5.53)$$

where $\check{\mathbf{\Gamma}}_S = \mathbf{I}_d \odot (\mathbf{S}_T^H \mathbf{\Pi}_{\tilde{\mathbf{U}}_T} \mathbf{S}_T)$. Therefore the subspaces spanned by the r_S leading eigenvectors of the correlation matrices (5.14) and (5.53) coincide if $[\mathbf{S}_T^H \mathbf{\Pi}_{\tilde{\mathbf{U}}_T} \mathbf{S}_T]_{kk} \neq 0 \ \forall k$. Since a sufficient condition is $range\{\mathbf{\Pi}_{\tilde{\mathbf{U}}_T}\} \supseteq range\{\mathbf{S}_T\}$, the choice $\mathbf{\Pi}_{\tilde{\mathbf{U}}_T} = \mathbf{I}_W$ does not preclude the optimality so that the ML estimate $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{U}}_S}$ can be obtained from the eigenvectors of (5.14) for $N_B \to \infty$.

Bibliography

[1] C. Komninakis, C. Fragouli, A. H. Sayed and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing,* vol. 50, no. 10, pp. 1065-1076, May 2002.

[2] A. Van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint Angle and Delay Estimation Using Shift-Invariance Techniques", *IEEE Trans. Signal Processing*, vol 46, pp 405-418, Feb. 1998.

[3] Yung-Yi Wang, Jiunn-Tsair Chen and Wen-Hsien Fang, "TST-MUSIC for Joint AOA-Delay Estimation," *IEEE Trans. Signal Processing*, vol. 49, pp. 721-729, April 2001.

[4] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing,* vol. 46, no. 2, pp. 449-462, Feb. 1998.

[5] J. Picheral and U.Spagnolini, "Angles and delay estimation of space-time channels for multiuser detection," *IEEE Trans. on Wireless Communications*, vol. 3, no. 3, pp. 758-769, May 2004.

[6] M. Nicoli, O. Simeone, U. Spagnolini, "Multislot estimation of fast-varying space-time channels", *IEEE Trans. Signal Processing*, vol.51, no.5, pp.1184-1195, May 2003.

[7] F. Dietrich, M.T. Ivrlac and J.A. Nossek, "On performance limits of optimum reduced rank channel estimation," *Proc. GLOBECOM*, vol. 1, pp. 345-349, 2002.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[9] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Signal Processing*, vol. 33, no.2, pp. 387-392, April 1985.

[10] D. J. Rabideau, "Fast, rank-adaptive subspace tracking and applications," *IEEE Trans. Signal Processing*, vol. 44, no. 9, pp. 2229-2244, Sept. 1996.

[11] S. Attallah, K. Abed-Meraim, "Fast algorithms for subspace tracking," IEEE Signal Processing Letters, vol. 8, no. 7, pp. 203-206, July 2001.

[12] P. Strobach, "Low-rank adaptive filters", *IEEE Trans. Signal Processing*, vol.44, no.12, pp.2932-2947, Dec. 1996.

[13] A. Kavcic, B. Yang, "Adaptive rank estimation for spherical subspace trackers", *IEEE Trans. Signal Processing*, vol.44, no.6, pp.1573-1579, June 1996.

[14] M. Nicoli, M. Sternad, U. Spagnolini and A. Ahlen, "Reduced-rank channel estimation and tracking in time-slotted CDMA systems," *Proc. International Conference on Communications (ICC) 2002*, vol. 1, pp. 533-537, May 2002.

[15] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: theory and experiments," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2538-2546, Oct. 2002.

[16] R. Horn and C. Johnson, *Matrix analysis*, Cambridge University Press, 1985.

[17] M. Nicoli, *Multi-user Reduced Rank Receivers for TD/CDMA Systems*, Ph.D. Thesis, Politecnico di Milano, Italy, Dec. 2001.

[18] J. Du, Y. Li, "MIMO-OFDM channel estimation based on subspace tracking", *Proc. IEEE VTC 2003*, vol.2 , pp.1084-1088, April 2003.

[19] M.Stege, P.Zillmann, G.Fettweis, "MIMO channel estimation with dimension reduction", *Proc. IEEE International Symposium on WPMC 2002*, vol.2, pp.417-421, Oct 2002.

[20] L. Lindbom, M. Sternad and A. Ahlén, "Tracking of the time-varying mobile radio channels-Part I: the Wiener LMS algorithm," *IEEE Trans. Comm.*, vol. 49, no. 12, pp. 2207-2217, Dec. 2001.

[21] M. Nicoli, M. Sternad, U. Spagnolini, A. Ahlen, "Reduced-rank channel estimation and tracking in time-slotted CDMA systems," in *Proc. ICC 2002*, vol. 1, pp. 533-537, May 2002.

[22] M. Cicerone, *Modal analysis/filtering to estimate time-varying MIMO-OFDM channels*, Master Thesis, Politecnico di Milano, Italy, March 2004.

[23] Y. Li, "Optimum Training sequences for OFDM systems with multiple transmit antennas", in *Proc. IEEE GLOBECOM'00*, vol.3, pp. 1478-1482, 2000.

[24] I. J. Gupta, J. R. Baxter, S. W. Ellingson, H.-G. Park, H. Seo Oh and M. G. Kyeong, "An experimental study of antenna array calibration," *IEEE Trans. Antennas and Propagation*, vol. 51, no. 3 , pp. 664-667, March 2003.

[25] 3GPP TR 25.996 V.6.0,"Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations (Release 6)", *3GPP Technical Specification Group Radio Access*, June 2003.

[26] M. Cicerone, N. Geng, O. Simeone and U. Spagnolini, "Modal analysis/filtering to estimate time-varying MIMO-OFDM channels," *Proc. ITG Workshop on Smart Antennas*, March 2004.

[27] S. Verdu, *Multiuser Detection*, Cambridge University Press, 1998.

[28] H. Holma and A. Toskala, *WCDMA for UMTS,* John Wiley & Sons, 2000.

[29] M. Nicoli, O. Simeone and U. Spagnolini, "Multi-slot estimation of fast-varying frequency-selective channels", *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1337-1347, Aug. 2003.

[30] G. Caire, G. Taricco and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927-946, May 1998.

[31] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: theory and experiments," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2538-2546, Oct. 2002.

[32] Jun Tan, G. L. Stuber, "New SISO decoding algorithms," *IEEE Trans. Commun.* vol. 51 , no. 6 , pp. 845-848, June 2003.

[33] Kee-Bong Song, S.A. Mujtaba, "A Low Complexity Space-Frequency BICM MIMO-OFDM System for Next-Generation WLANs," *Proc. IEEE Global Telecommunications Conference 2003 (GLOBECOM '03)*, vol. 2, pp. 1059 - 1063, Dec. 2003.

[34] G. K. Kaleh, "Channel equalization for block transmission systems," *IEEE J. Select. Areas Commun.,* vol. 13, no. 1, pp. 110-121, Jan. 1995.

[35] O. Simeone, M. Nicoli, U. Spagnolini, "Subspace tracking for uplink/downlink array processing in CDMA systems", *Proc. IEEE Telecommunications Conference (GLOBECOM 2003)*, vol. 2, pp. 824-828, san Francisco, Dec. 1-5, 2003.

Chapter 6

Information rate with imperfect CSI

6.1 Introduction

ACCORDING to the principle of synchronized detection, after channel estimation (i.e., CSI acquisition) has been performed on the training part of the block (see fig. 3.1 for time-domain transmission and fig. 3.3 for multicarrier transmission), the detection process is carried out on the data symbols by considering the estimated channel as if it was the real channel [1]. So far, this thesis focused on evaluating at first the theoretical limit perfomance of the channel estimation process (Chapter 4) and then on designing novel channel estimators able to approach the theoretical bound (Chapter 5). Here the analysis focuses on the detection phase by pursuing an information theoretic analysis of a MIMO system with imperfect CSI. Toward this goal, channel estimators that attain the asymptotic limit performance derived in Chapter 4 will be considered.

To start with, Sec. 6.2t addresses the case of a single user MIMO system over a frequency-selective channel and, in order to simplify the analysis, therein it is assumed that no CSI is available at the transmitter. Similar investigations have been proposed in [2] within the context of a frequency-flat channel in a diversity scenario and recently in [3], where the study was limited to UML channel estimation. The capacity of a frequency-flat MIMO system in presence of imperfect CSI at the transmitter has been investigated in [4].

Under the assumption of no CSI at the transmitter, the extension of the analysis presented in Sec. 6.2 to a multiuser system is straightforward (see Sec. 6.4 and [5] for an overview on the information theoretic analysis of MIMO systems). However, in case of (generally imperfect) CSI available at the transmitter, the analysis requires more care. In fact, while for the uplink the capacity region and the sum-capacity in case of perfect CSI at the transmitter are well studied, only recently an achievable rate region and the sum-capacity have been determined for the downlink through a duality result [6]. This duality result is based on a non-linear precoding technique proposed in [7], that requires perfect CSI at the transmitter. Therefore, while the study of uplink under the assumption of imperfect CSI is a relatively easy task and

follows the lines of Sec. 6.4, an investigation of the information rate of the downlink under the same condition is an open problem. Thus, in this Chapter (Sec. 6.4), we focus on the information rate of a specific linear precoding scheme for the downlink of a multiuser MIMO system in presence of imperfect CSI. The scheme under consideration has been proposed in [8] and essentially enforces a zero inter-user interference constraint through orthogonal linear precoding.

6.2 Lower bound on the information rate for a single-user MIMO link

The impact of imperfect channel knowledge on the system performance is investigated in [9] for a SISO link in terms of capacity. In particular, a lower bound on the information rate assuming a gaussian input distribution is derived. Notice that the choice of a gaussian input distribution might not lead to the maximization of the information rate and thus to the capacity of the system, but it greatly simplifies the mathematical analysis. The received signal (3.12) can be written in terms of the channel estimate $\hat{\mathbf{h}}$ as (we drop the temporal dependence on the block ℓ for simplicity of notation):

$$\mathbf{y} = \boldsymbol{\mathcal{H}}\mathbf{x} + \mathbf{n} = \widehat{\boldsymbol{\mathcal{H}}}\mathbf{x} + ((\boldsymbol{\mathcal{H}} - \widehat{\boldsymbol{\mathcal{H}}})\mathbf{x} + \mathbf{n}), \tag{6.1}$$

where $\widehat{\boldsymbol{\mathcal{H}}}$ is the block-convolution matrix (3.13) for time-domain transmission or the block-diagonal matrix (3.30) for multicarrier transmission built from the channel estimate $\hat{\mathbf{h}}$. According to [9], a lower bound on the information rate $I(\mathbf{y}, \mathbf{x}|\hat{\mathbf{h}})$ between the transmitted signal $\mathbf{x}$ and the received signal $\mathbf{y}$ can be obtained by treating the equivalent disturbance term in (6.1) due to the channel estimation error, $(\boldsymbol{\mathcal{H}} - \widehat{\boldsymbol{\mathcal{H}}})\mathbf{x}$, as a gaussian additive noise. Letting the input distribution be Gaussian, as stated above, with $E[\mathbf{x}\mathbf{x}^H] = P/N_T\mathbf{I}$, we get that the covariance matrix of the additive noise is $P/N_T\mathbf{Q}_{\widehat{\boldsymbol{\mathcal{H}}}}$, where we defined the $N_R L_D \times N_R L_D$ correlation matrix $\mathbf{Q}_{\widehat{\boldsymbol{\mathcal{H}}}} = E[(\widehat{\boldsymbol{\mathcal{H}}} - \boldsymbol{\mathcal{H}})(\widehat{\boldsymbol{\mathcal{H}}} - \boldsymbol{\mathcal{H}})^H]$. Now, recall that the capacity for perfect CSI at the receiver (and no CSI at the transmitter) for a block of length L reads

$$C = I(\mathbf{y}, \mathbf{x}|\mathbf{h}) = \frac{1}{L}\log_2\left|\mathbf{I} + \frac{P}{N_T}(\mathbf{I}_{L_D} \otimes \mathbf{R}_n^{-1})\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^H\right| \quad [bit/s/Hz]. \tag{6.2}$$

Notice that, from Chapter 3, the block length reads $L = L_D + L_O$, where L_O accounts for the overhead necessary for channel estimation and for ensuring absence of inter-block (and inter–carrier for multicarrier transmission) interference. Then, according to the discussion above, a lower bound on the information rate

$$C \geq I(\mathbf{y}, \mathbf{x}|\hat{\mathbf{h}}) \geq I_{lb} \tag{6.3}$$

can be computed by adding to the noise covariance matrix $\mathbf{R}_n$ the additional term due to the channel estimation error $P/N_T\mathbf{Q}_{\widehat{\boldsymbol{\mathcal{H}}}}$ as in

$$I_{lb} = \frac{1}{L}\log_2\left|\mathbf{I} + \frac{P}{N_T}[(\mathbf{I}_{L_D} \otimes \mathbf{R}_n) + \frac{P}{N_T}\mathbf{Q}_{\widehat{\boldsymbol{\mathcal{H}}}}]^{-1}\widehat{\boldsymbol{\mathcal{H}}}\widehat{\boldsymbol{\mathcal{H}}}^H\right| \quad [bit/s/Hz]. \tag{6.4}$$

Notice that in (6.4) the power transmitted over the data part of the block is assumed to be equal to the power transmitted over the training symbols P, see [10] [3] and Sec. 6.4 for a discussion on the benefits of exploiting the degree of freedom in the choice of different power for data and training. Under the assumption $N_B \to \infty$ (consistent with the infinite temporal horizon of information theory), the signal model (4.28) is linear in the fading amplitudes and similarly to [2] [9] [10] we can consider the matrix $\widehat{\mathcal{H}}$ as obtained from the MMSE estimator of the fading amplitudes described in Chapter 4 so that $\mathbf{Q}_{\widehat{\mathcal{H}}}$ can be computed from the error correlation matrix $\mathbf{Q}_{\hat{\mathbf{h}}}$, as shown in the Appendix. Notice that a different choice for the channel estimator would decrease the lower bound (6.4) so that the inequality $I(\mathbf{y}, \mathbf{x}|\hat{\mathbf{h}}) \geq I_{lb}$ holds for a channel estimator that reaches the performance bound (4.46).

6.3 Numerical examples: single user MIMO link

6.3.1 Simulation setting

The simulation setting is in Table 6.1. To gain some insight into the effects of system parameters and channel characteristics on the information rate, the correlation matrices involved in signal modelling are assumed to be obtained from an autoregressive model: $\mathbf{Z}_N(\rho)$ denotes a $N \times N$ Toeplitz matrix with first column $[1 \; \rho \; \cdots \rho^{N-1}]^T$. Here ρ_n accounts for the spatial correlation of noise ($\mathbf{R}_n$), ρ_l and ρ_a for auto- and mutual-correlation of the training sequences of different antennas (training sequences in practical systems are usually designed to have $\rho_l = \rho_a \simeq 0$ as explained in Chapter 4). For the beamforming scenario, the arrays at both ends are assumed to be uniform linear with half-wavelength inter-element spacing. For the diversity scenario the spatial correlation matrices of different paths $\mathbf{R}_i = \mathbf{R}$, $i = 1, ..., d$, are assumed to be equal and separable (see (4.44)) and the correlation coefficient ρ_T and ρ_R characterize the spatial correlation of the fading at the transmitter and receiver side, respectively. The Doppler spectrum of all paths is uniform with (normalized) Doppler spread $f_D T_S$.

The influence of system parameters and channel characteristics on the lower bound on the information rate in presence of imperfect CSI (6.4) and on the capacity for a known channel (6.2) is investigated numerically. The computation of the bound (6.4) is carried out by aver-

Table 6.1: Default parameter settings

N_R	4	$\mathbf{Q}$	$P/N_T \mathbf{Z}_{N_R}(\rho_n)$
N_T	4	$\mathbf{R}_x$	$P/N_T \mathbf{Z}_W(\rho_l) \otimes \mathbf{Z}_{N_T}(\rho_a)$
W	8	$\mathbf{R}$	$\mathbf{Z}_{N_T}(\rho_T) \otimes \mathbf{Z}_{N_R}(\rho_R)$
d	4	ρ_n	0
τ_i/T	$i-1$	ρ_l	0
$\alpha_i^{(T)}, \alpha_i^{(R)}$	$-60 + 120/d \cdot (i-1)$ deg	ρ_a	0

aging over the distribution of fading amplitudes (recall that angles, delays and power-delay profile are deterministic) by using 10^3 runs of Monte Carlo simulations. Similarly to the experimental analysis carried out in [2], the estimate $\hat{\mathbf{d}}$ of the fading amplitudes (recall Sec. 4.7) has a circularly symmetric Gaussian distribution with covariance matrix $\mathbf{Q}_{\hat{\mathbf{d}}}$, i.e., $\hat{\mathbf{d}} \backsim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\hat{\mathbf{d}}})$. The size of the data part of the block is $L_D = 30$ and $SNR = P/\sigma_n^2 = 10dB$.

6.3.2   Optimal training length

The optimal choice of the length L_P of the training sequences is investigated first. The study here focuses on the diversity scenario since the beamforming scenario would lead to similar conclusions. Fig. 6.1 shows the ergodic capacity (dashed lines) and the lower bound on the information rate (solid lines) versus L_P for $\rho_T = \rho_R = 0.3$, different Doppler shifts ($f_D T_S = 0.1, 0.5$) and number of paths $d = 4, 8$. Increasing L_P has two opposite effects on the information rate: it increases the overhead L_O (which reduces both the scaled capacity and the information rate), and it decreases the channel estimation error (which has a beneficial impact on the information rate). Therefore, the bound I_{lb} versus L_P shows an optimum trade-off between transmission overhead and channel estimation error as in fig. 6.1. The optimum training sequence length L_P depends on both the number of paths d and the Doppler shift. For $f_D T_S = 0.1$ the information rate I_{lb} peaks at approximately $L_P \simeq d$, whereas for larger $f_D T_S$ the maximum moves towards slightly higher values of L_P. This can be simply interpreted by pointing out that in a block-fading channel, it is advisable to increase L_P in order to experience a constant channel for a longer interval if the channel itself varies rapidly over consecutive bursts. Interestingly, under the assumption of known long term features of the channel, the optimum L_P is much smaller than the number of training symbols required to obtain the UML estimator (i.e., $L_P \geq N_T W$).

6.3.3   Effect of spatially correlated fading and number of paths

Fig. 6.2 shows the degradation of the average information rate I_{lb} compared to the case of known channel as a function of normalized Doppler spread $f_D T_S$, spatial correlation $\rho_T = \rho_R$ (for the diversity scenario) and number of paths d (for the beamforming scenario). The degradation is measured as $(C - I_{lb})/C$ (notice that being $I_{lb} \leq I \leq C$ the selected quantity is an upper bound on $(C - I)/C$) and set $L_P = 32$. The values of the capacity C are in the table in the upper left corner.

For decreasing Doppler shifts, the degradation decreases and vanishes for $f_D = 0$ since a static channel can be estimated with any accuracy as $N_B \to \infty$ (see Sec. 4.7.2). Moreover, the degradation decreases with increasing spatial correlation (for the diversity scenario) since higher correlations lead to a smaller channel estimation error. In order to keep the example consistent, as $d = 4$ for the diversity scenario, the delays are selected as $\tau_i/T = \lfloor (i-1)/4 \rfloor$ for $i = 1, .., d$ with $d = 4, 12, 20$ for the beamforming scenario (i.e., the delays span the
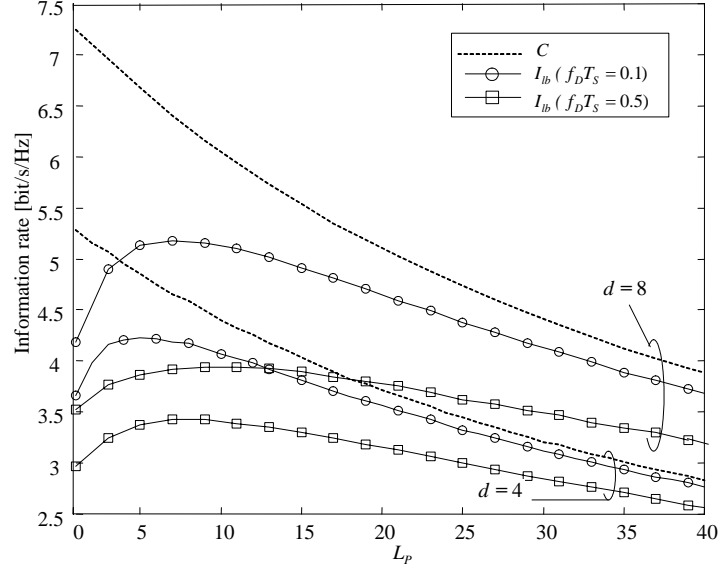
Figure 6.1: Ergodic capacity with perfect CSI C and average lower bound on the information rate with imperfect channel state information I_{lb} versus the length of the training sequence L_P (diversity model, $\rho_T = \rho_R = 0.3$).

first four samples of the channel). With this choice, increasing the number of paths in the beamforming scenario has qualitatively the same effect as decreasing the spatial correlation in the diversity scenario, as shown in fig. 6.2. As a final remark, the degradation for the beamforming scenario is smaller than for the diversity scenario as in the first case the number of parameters (amplitudes) to be tracked is $r \simeq d$ whereas in the latter is larger ($r \simeq N_R N_T d$).

6.3.4 Effect of the correlation properties of the training sequences

The effect of the correlation properties of the training sequences (ρ_a and ρ_l) on the system performance is addressed in terms of the information rate degradation as in the previous example. The results for the diversity scenario with $f_D T_S = 0.1$ and $L_P = 32$ ($C = 3.1$ bit/sec/Hz as shown in figures 6.1 and 6.2) are depicted in fig. 6.3. The degradation of the average information rate versus ρ_l with $\rho_a = 0$ and versus ρ_a with $\rho_l = 0$ show that both ρ_a and ρ_l have a similar impact on the information rate. The correlations range from 0 to 0.9 since a complete correlation ($\rho_a = 1$ or $\rho_l = 1$) would not be compatible with the assumption of a consistent estimate of the long term channel features.
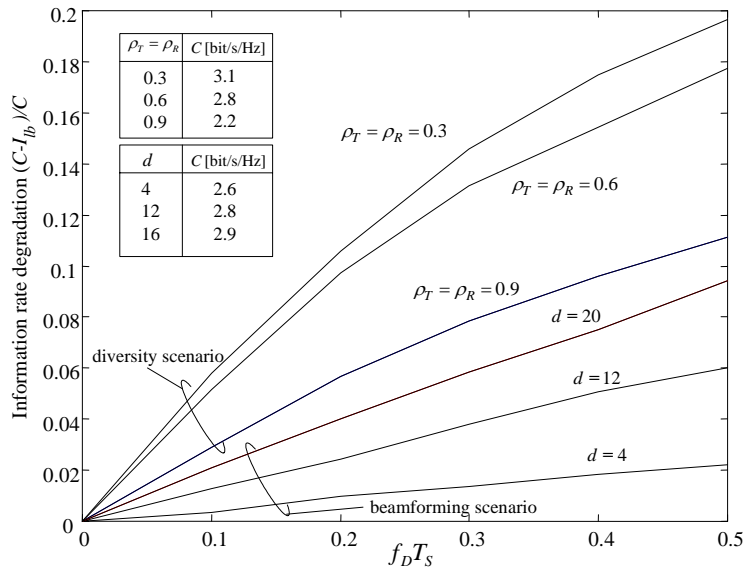
Figure 6.2: Degradation of the average information rate compared to the ergodic capacity with perfect CSI versus Doppler shift $f_D T_S$ for different values of spatial correlation $\rho_T = \rho_R$ (diversity model) and number of paths d (beamforming model).
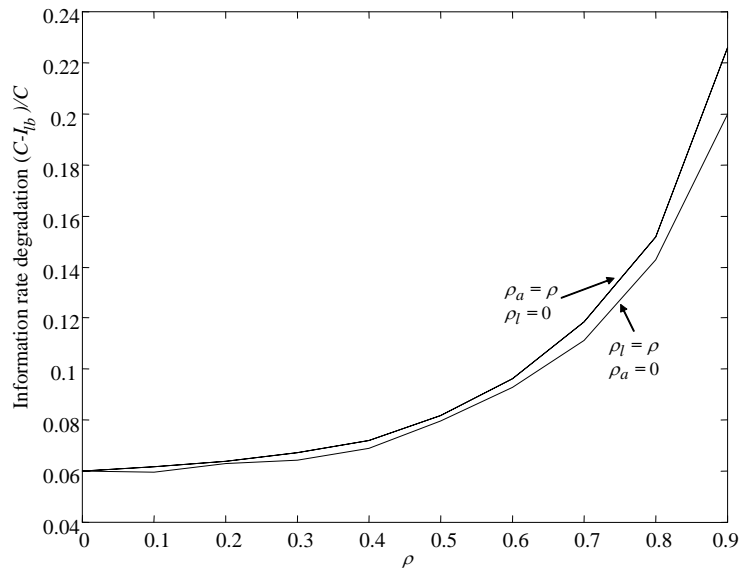

Figure 6.3: Degradation of the average information rate compared to the ergodic capacity with perfect CSI versus correlation ρ: *i)* $\rho_l = \rho$ and $\rho_a = 0$; *ii)* $\rho_a = \rho$ with $\rho_l = 0$.
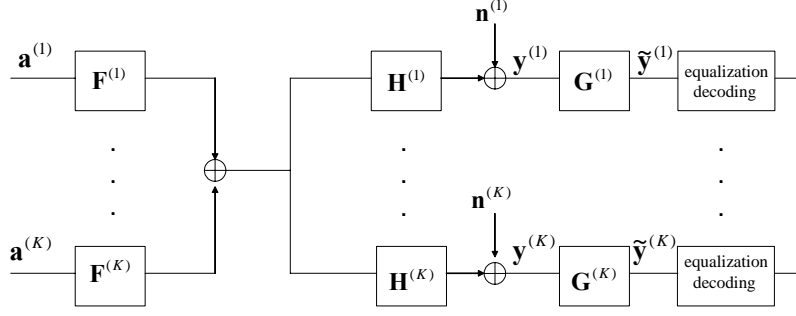
Figure 6.4: Block diagram of the downlink of a MIMO system with linear precoding.

6.4 Lower bound on the information rate for downlink with orthogonal precoding

As explained in Sec. 6.1, the more challenging setting in which to study the information rate in presence of imperfect CSI at the transmitter (and the receiver) is the downlink of a multiuser MIMO system. In order to simplify the investigation, in this Section a specific linear precoding scheme that ensures zero inter-user interference through orthogonal precoding [8] is assumed. Moreover, as a further simplification, this Section considers a frequency-flat channel $(W = 1)$ within a diversity scenario with spatially and temporally uncorrelated fading, $\mathbf{R} = \mathbf{I}_{N_R N_T}$. Notice that in this case $L = L_D + L_P$ (i.e., $L_O = L_P$). Finally, noise is assumed to be spatially white and the training sequences optimally designed.

6.4.1 Orthogonal spatial precoding

From (3.4) and (3.36), the signal received by the kth user within the data part of the ℓth block can be written as the $N_R \times 1$ vector $\mathbf{y}_\ell^{(k)}[m]$

$$\mathbf{y}_\ell^{(k)}[m] = \mathbf{H}_\ell^{(k)} \mathbf{x}_\ell^{(k)}[m] + \mathbf{H}_\ell^{(k)} \sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{x}_\ell^{(i)}[m] + \mathbf{n}_\ell^{(k)}[m], \quad m = 0, ..., L_D - 1. \tag{6.5}$$

In (6.5), the second term on the right hand side represents the interference from other users. Notice that for a frequency-flat channel, time and frequency-domain transmission strategies give rise to the same signal model (6.5). In the following we drop for simplicity of notation the dependence on the time sample m and the block ℓ. A general block diagram for downlink with linear precoding is shown in fig. 6.4.

Accordingly, the $N_T \times 1$ vector $\mathbf{x}_\ell^{(k)}[m]$ transmitted by the kth user is obtained by linearly precoding the $p^{(k)} \times 1$ data vector $\mathbf{a}^{(k)}$ output by the modulator through multiplication by the

$N_T \times p^{(k)}$ matrix $\mathbf{F}^{(k)}$:

$$\mathbf{x}^{(k)} = \mathbf{F}^{(k)}\mathbf{a}^{(k)}. \tag{6.6}$$

It is important to emphasize that in this Section we are assuming that the number of spatial channels $p^{(k)}$ assigned to each user is selected as $p^{(k)} = rank(\mathbf{H}_\ell^{(k)}) = n_R^{(k)}$ ($N_T \geq n_R^{(k)}$) and that condition

$$N_T \geq \sum_{k=1}^{K} n_R^{(k)} \tag{6.7}$$

is satisfied. The condition (6.7) makes it possible to calculate the set of orthogonal precoding matrices $\mathcal{F} = \{\mathbf{F}^{(k)}\}_{k=1}^{K}$ as explained below. This assumption will be removed in Chapter 8.

Orthogonal precoding proposed in [8] is obtained by maximizing the sum-capacity of the system under a total power constraint and under the condition of zero inter-user interference:

$$\mathcal{F} = \underset{\mathcal{F}}{\operatorname{argmax}} \sum_{i=1}^{K} C^{(i)}(\mathcal{F}) \tag{6.8a}$$

$$\text{s.t. } \sum_{k=1}^{K} tr(\mathbf{F}^{(k)}\mathbf{F}^{(k)H}) \leq \frac{L(1-\upsilon)}{(L-N_T)}\bar{P}, \tag{6.8b}$$

$$\mathbf{H}^{(i)}\mathbf{F}^{(j)} = \mathbf{0} \ \ if \ \ i \neq j, \tag{6.8c}$$

with the capacity of each user being

$$C^{(i)}(\mathcal{F}) = \log_2 |\mathbf{I}_{d_i} + \mathbf{R}_n^{(i)-1}(\mathbf{H}^{(i)}\mathbf{F}^{(i)}\mathbf{F}^{(i)H}\mathbf{H}^{(i)H})| \tag{6.9}$$

and the noise (interference) covariance matrix for each user (recall (6.5))

$$\mathbf{R}_n^{(i)} = \sigma_n^2 \mathbf{I}_{n_{Ri}^{(k)}} + \sum_{k \neq i} \mathbf{H}^{(i)}\mathbf{F}^{(k)}\mathbf{F}^{(k)H}\mathbf{H}^{(i)H}. \tag{6.10}$$

The power constraint (6.8b) requires more explanation. The total power available at the base station (averaged over the block length L) is $\bar{P}$, which has to be distributed among the training (P_P) and data (P_D) part as

$$\bar{P} = P_P\frac{L_P}{L} + P_D\frac{L_D}{L} = \bar{P}_P + \bar{P}_D. \tag{6.11}$$

Notice that in (6.11) we are distinguishing between the instantaneous power and the average (over the block length) power by the upperscore. In the following we consider a fixed amount of training symbols, i.e., the minimum amount needed in order to obtain a meaningful channel estimate, $L_P = N_T$ (see Chapter 4 and 5), and share the average power $\bar{P}$ as $\bar{P}_P = \nu\bar{P}$ and $\bar{P}_D = (1-\upsilon)\bar{P}$, $0 \leq \upsilon \leq 1$. From this discussion, the average power constraint (6.8b) easily follows. Moreover, equation (6.8c) enforces the zero inter-user interference constraint.

The optimization problem (6.8) is solved in [8] by assuming perfect channel knowledge at the transmitter as explained below. Let us consider the matrix gathering all the channel matrices

except the ith user's matrix $\bar{\mathbf{H}}^{(i)} = [\mathbf{H}^{(1)H}.., \mathbf{H}^{(i-1)H}, \mathbf{H}^{(i+1)H},.., \mathbf{H}^{(K)H}]^H$. Now, in order to guarantee the zero-interference constraint (6.8c) the subspace spanned by the columns of the precoding matrix $\mathbf{F}^{(i)}$ must lie into the null space of matrix $\bar{\mathbf{H}}^{(i)}$, that can be computed from the (full) singular value decomposition

$$\bar{\mathbf{H}}^{(i)} = \begin{bmatrix} \bar{\mathbf{U}}^{(i)} & \bar{\mathbf{U}}_0^{(i)} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\Lambda}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}^{(i)} & \bar{\mathbf{V}}_0^{(i)} \end{bmatrix}^H \tag{6.12}$$

as the subspace spanned by the columns of the orthonormal matrix $\bar{\mathbf{V}}_0^{(i)}$. Since $\bar{\mathbf{H}}^{(i)}\bar{\mathbf{V}}_0^{(j)} = \mathbf{0}$ for $j \neq i$, the precoding matrix can be written as $\mathbf{F}^{(i)} = \bar{\mathbf{V}}_0^{(i)}\breve{\mathbf{F}}^{(i)}$, where the first term $\bar{\mathbf{V}}_0^{(i)}$ nulls inter-user interference while $\breve{\mathbf{F}}^{(i)}$ is used for further optimization. Notice that in general the condition (6.7) is required for $\bar{\mathbf{V}}_0^{(i)}$ to exist. To obtain $\breve{\mathbf{F}}^{(i)}$, let us consider the (interference-free) channel matrix for user i be $\mathbf{H}^{(i)}\bar{\mathbf{V}}_0^{(i)}$ and its singular value decomposition (limited to the non-zero eigenvalues)

$$\mathbf{H}^{(i)}\bar{\mathbf{V}}_0^{(i)} = \mathbf{U}^{(i)}\boldsymbol{\Lambda}^{(i)}\mathbf{V}^{(i)H}. \tag{6.13}$$

The precoding matrices $\mathbf{F}^{(i)}$ that maximize the capacity (8.6) can then be expressed as ($\breve{\mathbf{F}}^{(i)} = \mathbf{V}^{(i)}\boldsymbol{\Phi}^{(i)}$, see also [5])

$$\mathbf{F}^{(i)} = \bar{\mathbf{V}}_0^{(i)}\mathbf{V}^{(i)}\boldsymbol{\Phi}^{(i)}, \tag{6.14}$$

where $\boldsymbol{\Phi}^{(i)}$ is a diagonal matrix that defines power allocation over the channel modes according to waterfilling:

$$|[\boldsymbol{\Phi}^{(i)}]_{kk}|^2 = (\xi - \frac{\sigma_n^2}{[\boldsymbol{\Lambda}^{(i)}]_{kk}^2})^+, \tag{6.15}$$

where $(x)^+ = \max(x, 0)$ and ξ is such that the power constraint (6.8b) is satisfied.

6.4.2 Imperfect CSI at the transmitter

The CSI available at the transmitter to calculate the precoding matrices $\{\mathbf{F}^{(i)}\}_{i=1}^K$ in (6.14) consists of a set of estimates $\{\hat{\mathbf{H}}^{(i)}\}_{i=1}^K$ of the channel matrices $\{\mathbf{H}^{(i)}\}_{i=1}^K$, generally fedback by the receivers via a dedicate link [11]. Notice that in case the channel is reciprocal as in a Time Division Duplex link, the CSI could be directly acquired by the base station via measurements on the uplink (as long as the differences in the electronics at the transmitter and receiver can be estimated and compensated for). However, in general, the actual and estimated channel matrices are mismatched because of the effects of *i*) channel estimation errors and *ii*) feedback delay that outdates the estimated channels when the propagation environment is time varying.

In order to simplify the analysis, we consider the channel estimates obtained by the UML channel estimator, whose performance are recalled here for reference (see Sec. 4.2):

$$\mathbf{Q}_{UML} = \frac{\sigma_n^2}{L_P P_P}\mathbf{I}_{N_R N_T}. \tag{6.16}$$

We emphasize that this is a minor limitation to our analysis since as shown in Sec. 4.7.2, the UML estimator for the setting considered here attains the conventional CRB. Moreover, it

achieves the HCRB as well for high SNR and temporally uncorrelated fading amplitudes. In the following, we define for simplicity of notation the normalized (on the channel norm) MSE on channel estimation as $\overline{MSE} = \sigma_n^2/(L_P P_P)$. Accordingly, the estimates $\hat{\mathbf{H}}^{(i)}$ can be written as the sum of the real channel and the estimation error: $\hat{\mathbf{H}}^{(i)} = \mathbf{H}^{(i)} + \mathbf{H}_e^{(i)}$, where the entries of $\mathbf{H}_e^{(i)}$ are independent identically distributed $[\mathbf{H}_e^{(i)}]_{n_R,n_T} \sim \mathcal{CN}(0, \overline{MSE})$.

6.4.3 Lower bound with channel estimation error

Under the assumption of imperfect CSI available at the transmitter, even with orthogonal precoding all the users interfere with each other as, in general, $\mathbf{H}^{(i)}\hat{\mathbf{F}}^{(j)} \neq \mathbf{0}$ for $i \neq j$, with $\hat{\mathbf{F}}^{(j)}$ being the precoding matrix for the jth user computed according to the channel estimates $\{\hat{\mathbf{H}}^{(i)}\}_{i=1}^K$: $\hat{\mathbf{H}}^{(i)}\hat{\mathbf{F}}^{(j)} = \mathbf{0}$ for $i \neq j$. The received signal in (6.5) then reads

$$
\begin{aligned}
\mathbf{y}^{(k)} &= \mathbf{H}^{(k)}\hat{\mathbf{F}}^{(k)}\mathbf{a}^{(k)} + \mathbf{H}^{(k)}\sum_{\substack{i=1 \\ i \neq k}}^{K} \hat{\mathbf{F}}^{(i)}\mathbf{a}^{(i)} + \mathbf{n}^{(k)} = \\
&= \hat{\mathbf{H}}^{(k)}\hat{\mathbf{F}}^{(k)}\mathbf{a}^{(k)} - \mathbf{H}_e^{(k)}\hat{\mathbf{F}}\mathbf{a} + \mathbf{n}^{(k)},
\end{aligned} \tag{6.17}
$$

where $\hat{\mathbf{F}} = [\hat{\mathbf{F}}^{(1)} \cdots \hat{\mathbf{F}}^{(K)}]$ and $\mathbf{a} = [\mathbf{a}^{(1)T} \cdots \mathbf{a}^{(K)T}]^T$. In (6.17) the channel estimate $\hat{\mathbf{H}}^{(k)}$ plays the role of the real channel and the term $(-\mathbf{H}_e^{(i)}\hat{\mathbf{F}}\mathbf{a} + \mathbf{n}^{(k)})$ of additive noise.

In order to evaluate a lower bound on the information rate for orthogonal precoding $\sum_{k=1}^K I(\mathbf{y}^{(k)}, \mathbf{a}|\hat{\mathbf{h}}) \geq I_{lb}$, according to [9], we need to evaluate the distribution of the additive noise term in (6.17) conditioned on the channel estimates $\hat{\mathbf{H}}^{(k)}$. Toward this goal, it can be easily proved that the entries of the channel estimation error $\mathbf{H}_e^{(k)}$, conditioned on $\hat{\mathbf{H}}^{(k)}$, are uncorrelated and distributed as

$$
[\mathbf{H}_e^{(k)}]_{n_R,n_T}|_{\hat{\mathbf{H}}^{(k)}} \sim \mathcal{CN}\left(\mu_e \cdot [\hat{\mathbf{H}}^{(k)}]_{n_R,n_T}, \sigma_e^2\right) \tag{6.18}
$$

where $\mu^e = (\overline{MSE})/(1 + \overline{MSE})$ and $\sigma_e^2 = \overline{MSE}/(1 + \overline{MSE})$. The interference term can then be written by separating the mean error $\mu_e \hat{\mathbf{H}}^{(k)}$ and the zero mean remaining term $\bar{\mathbf{H}}_e^{(k)}$ as $[\bar{\mathbf{H}}_e^{(k)}]_{n_R,n_T}|_{\hat{\mathbf{H}}^{(k)}} \sim \mathcal{CN}\left(0, \sigma_e^2\right)$:

$$
\mathbf{H}_e^{(k)}\hat{\mathbf{F}}\mathbf{a} = (\mu^e\hat{\mathbf{H}}^{(k)} + \bar{\mathbf{H}}_e^{(k)})\hat{\mathbf{F}}\mathbf{a} = \mu^e\hat{\mathbf{H}}^{(k)}\hat{\mathbf{F}}\mathbf{a} + \bar{\mathbf{H}}_e^{(k)}\hat{\mathbf{F}}\mathbf{a}, \tag{6.19}
$$

It should be noticed that the term $\mu_e \hat{\mathbf{H}}^{(k)}\hat{\mathbf{F}}\mathbf{a}$ in (6.19) is correlated with the desired part of the signal and accounts for an equivalent power loss with respect to the perfect channel estimation case. On the other hand, the residual error $\bar{\mathbf{H}}_e^{(k)}\hat{\mathbf{F}}\mathbf{a}$ is uncorrelated with the useful part of the signal. Even though the latter is not Gaussian distributed, the assumption of gaussianity (with the same covariance matrix) provides the worst-case scenario in terms of capacity loss [9]. A lower bound on the information rate for each user can then be easily evaluated as [9] [12]

$$
I(\mathbf{y}^{(k)}, \mathbf{a}|\hat{\mathbf{h}}) \geq I_{lb}^{(k)} = \frac{L_D}{L}\log_2|\mathbf{I}_{n_R^{(k)}} + (\mathbf{\Upsilon}_e^{(k)})^{-1}(1 - \mu_e)^2\hat{\mathbf{H}}^{(k)}\hat{\mathbf{F}}^{(k)}\hat{\mathbf{F}}^{(k)H}\hat{\mathbf{H}}^{(k)H}|), \tag{6.20}
$$

where $\mathbf{\Upsilon}_e^{(k)}$ is the noise covariance matrix

$$\mathbf{\Upsilon}_e^{(k)} = (\sigma_n^2 + \frac{L(1-\upsilon)}{(L-N_T)}\bar{P}\sigma_e^2)\mathbf{I}_{n_R^{(k)}}. \tag{6.21}$$

A lower bound on the downlink information rate is then obtained as $\sum_{k=1}^{K} I(\mathbf{y}^{(k)}, \mathbf{a}|\hat{\mathbf{h}}) \geq \sum_{k=1}^{K} I_{lb}^{(k)}$.

6.4.4 Lower bound with channel estimation error and feedback delay

As discussed above, in practice, the channel estimates $\{\hat{\mathbf{H}}^{(i)}\}_{i=1}^{K}$ must be fed back by the K users to the base station before being used for the optimization of spatial precoding. Here, the delay related to this transmission is assumed equal to one block, which amounts to optimizing transmission over the channel realization $\mathbf{H}_\ell^{(i)}$ using the estimate of the previous block $\hat{\mathbf{H}}_{\ell-1}^{(i)}$. The precoding matrices are hence selected so that $\hat{\mathbf{H}}_{\ell-1}^{(i)}\hat{\mathbf{F}}_\ell^{(j)} = \mathbf{0}$ for $i \neq j$. In this case, following the same steps of the previous Section, the received signal in (6.5) reads,

$$\mathbf{y}_\ell^{(k)} = \hat{\mathbf{H}}_\ell^{(k)}\hat{\mathbf{F}}_\ell^{(k)}\mathbf{a}_\ell^{(k)} - \mathbf{H}_{e,\ell}^{(k)}\hat{\mathbf{F}}_\ell^{(k)}\mathbf{a}_\ell^{(k)} + \mathbf{H}_\ell^{(k)}\sum_{\substack{i=1\\i\neq k}}^{K}\hat{\mathbf{F}}_\ell^{(i)}\mathbf{a}_\ell^{(i)} + \mathbf{n}_\ell^{(k)}, \tag{6.22}$$

where the channel estimate $\hat{\mathbf{H}}_\ell^{(k)}$ plays the role of the real channel and the remaining term of additive noise. Let the fading amplitudes have a temporal correlation following an autoregressive model of order one with correlation coefficient $0 \leq \rho \leq 1$. By the same arguments used in the previous section and after some more tedious calculations, the entries of $\mathbf{H}_{e,\ell}^{(k)}$ conditioned on a realization of $\{\hat{\mathbf{H}}_\ell^{(k)}, \hat{\mathbf{H}}_{\ell-1}^{(k)}\}$ are now distributed as

$$[\mathbf{H}_{e,\ell}^{(k)}]_{n_R,n_T}|_{\{\hat{\mathbf{H}}_\ell^{(k)},\hat{\mathbf{H}}_{\ell-1}^{(k)}\}} \sim \mathcal{CN}\left(\mu_d'\left[\hat{\mathbf{H}}_\ell^{(k)}\right]_{n_R,n_T} + \mu_d''\left[\hat{\mathbf{H}}_{\ell-1}^{(k)}\right]_{n_R,n_T}, \sigma_d^2\right) \tag{6.23}$$

with $\mu_d' = \dfrac{\overline{MSE}\left(1+\overline{MSE}\right)}{\left(1+\overline{MSE}\right)^2 - \rho^2}$, $\mu_d'' = \dfrac{-\overline{MSE}\rho}{\left(1+\overline{MSE}\right)^2 - \rho^2}$ and $\sigma_d^2 = \dfrac{\overline{MSE}\left(1-\rho^2+\overline{MSE}\right)}{\left(1+\overline{MSE}\right)^2 - \rho^2}$.

Therefore, after decomposing the interference term as (following the same notation as in the previous Section)

$$\mathbf{H}_{e,\ell}^{(k)} = \mu_d'\hat{\mathbf{H}}_\ell^{(k)} + \mu_d''\hat{\mathbf{H}}_{\ell-1}^{(k)} + \bar{\mathbf{H}}_{e,\ell}^{(k)}, \tag{6.24}$$

and defining $\bar{\mathbf{H}}_\ell^{(k)} = (1-\mu_d')\hat{\mathbf{H}}_\ell^{(k)} - \mu_d''\hat{\mathbf{H}}_{\ell-1}^{(k)}$, equation (6.22) can be rearranged as

$$\mathbf{y}_\ell^{(k)} = \bar{\mathbf{H}}_\ell^{(k)}\hat{\mathbf{F}}_\ell^{(k)}\mathbf{a}_\ell^{(k)} - \bar{\mathbf{H}}_{e,\ell}^{(k)}\hat{\mathbf{F}}_\ell^{(k)}\mathbf{a}_\ell^{(k)} + (1-\mu_d')\hat{\mathbf{H}}_\ell^{(k)}\sum_{\substack{i=1\\i\neq k}}^{K}\hat{\mathbf{F}}_\ell^{(i)}\mathbf{a}_\ell^{(i)} + \mathbf{n}_\ell^{(k)}. \tag{6.25}$$

A lower bound on information rate for the kth user can then be computed as [9]

$$I(\mathbf{y}^{(k)}, \mathbf{a}|\hat{\mathbf{h}}) \geq I_{lb}^{(k)} = \frac{L_D}{L}\log_2\left|\mathbf{I}_{n_R^{(k)}} + (\mathbf{\Upsilon}_d^{(k)})^{-1}\bar{\mathbf{H}}_\ell^{(k)}\hat{\mathbf{F}}_\ell^{(k)}\hat{\mathbf{F}}_\ell^{(k)H}\bar{\mathbf{H}}_\ell^{(k)H}\right|, \tag{6.26}$$

where the correlation matrix $\mathbf{\Upsilon}_d^{(k)}$ reads

$$\mathbf{\Upsilon}_d^{(k)} = \mathbf{\Upsilon}_e^{(k)} + (1 - \mu_d')^2 \hat{\mathbf{H}}_\ell^{(k)} \hat{\mathbf{F}}_\ell^{(k)} \hat{\mathbf{F}}_\ell^{(i)H} \hat{\mathbf{H}}_\ell^{(k)H}, \tag{6.27}$$

that accounts for both the channel estimation error ($\mathbf{\Upsilon}_e^{(k)}$) and the feedback delay ($\hat{\mathbf{H}}_\ell^{(k)} \hat{\mathbf{F}}_\ell^{(k)} \neq \mathbf{0}$ in presence of feedback delay as explained above).

6.5 Numerical examples: downlink with orthogonal precoding

The lower bound on the downlink information rate derived above is here evaluated through computer simulations under different assumptions on the CSI available at the transmitter as detailed in the following.

1. *Open loop (no CSI at the transmitter)*: channel estimation is carried out at the receiver, but there is no feedback of the estimate to the base station. Therefore, orthogonal precoding cannot be used and user separation must be carried on in the time (or frequency/code) domain [13]. Spatial multiplexing of data is still possible for a single user by transmitting equal power on each antenna. In this case, a lower bound on the information rate for user k can be calculated as (6.4) that, adapted to the setting considered herein, reads

$$I_{lb}^{(k)} = \frac{c^{(k)}}{L} \log_2 \left| \mathbf{I} + \frac{P_D}{N_T} [\mathbf{\Upsilon}_e^{(k)}]^{-1} \mathbf{H}^{(i)} \mathbf{H}^{(i)H} \right|, \tag{6.28}$$

   where $c^{(k)}$ is the number of time instants assigned to user k for transmission out of the L_D available ($\sum_{k=1}^{K} c^{(k)} = L_D$). Recall that all the users share the training part of the block.

2. *Closed loop transmission with perfect CSI*: The receiver performs both channel estimation and feedback of the estimate to the base station and the latter employs orthogonal precoding. It is assumed that perfect CSI is available at the transmitter, so that $\hat{\mathbf{H}}^{(k)} = \mathbf{H}^{(k)}$ and $\mathbf{H}_e^{(k)} = \mathbf{0}$. Moreover the transmission delay is zero. The information rate for user k is then

$$I^{(k)} = \frac{L_D}{L} \log_2 \left| \mathbf{I} + \sigma_n^{-2} \mathbf{H}^{(k)} \mathbf{F}^{(k)} \mathbf{F}^{(k)H} \mathbf{H}^{(k)H} \right|. \tag{6.29}$$

3. *Closed loop transmission with estimation error*: feedback delay is not taken into account, whereas estimation error is handled as explained in Sec. 6.4.3.

4. *Closed loop transmission with estimation error and feedback delay:* feedback delay and estimation error are taken both into account, according to the results in Sec. 6.4.4.

Fig. 6.5 shows contour plots of the lower bound I_{lb} (averaged over the fading distribution) on the sum rate as a function of υ (the fraction of power used for channel estimation) and
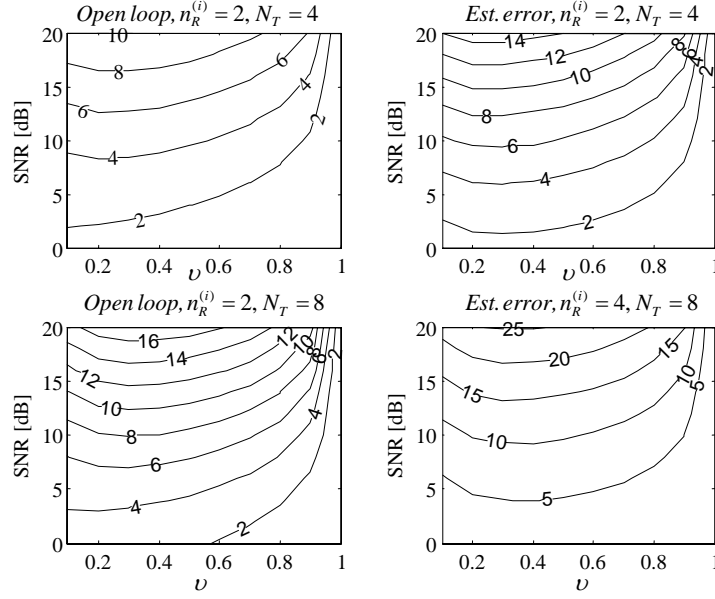
Figure 6.5: Average sum rate I_{lb} versus SNR and υ for open loop (left column), and closed loop transmission with channel estimation error (right column).

SNR for open loop ("Open loop") and closed loop transmission with estimation error ("Est. error"). The downlink considered consists of $K = 2$ users, each with $n_R^{(i)} = 2$ (top) and 4 (bottom) receiving antennas, and a base station equipped with $N_T = 4$ (top) and 8 (bottom) antennas. Using orthogonal precoding results in approximately a 50% increase of the system throughput. Moreover, it can be seen that the optimal fraction of power dedicated for channel estimation for a wide range of SNR lingers around $\upsilon = 0.25$, value used in the following simulation.

In fig. 6.6 it is shown the lower bound on the sum rate I_{lb} as a function of SNR for $K = 2$ users, $n_R^{(i)} = 2$ and $N_T = 4$. The impact of estimation error alone is not severe, less than $2dB$ uniformly with respect to SNR. On the other hand, channel outdating due to feedback delay has a remarkable effect, due to inter user interference accounted for by the rightmost term in (6.27). From fig. 6.6, for a channel correlation of $\rho = 0.8$, the simpler open loop strategy performs uniformly better than orthogonal precoding, due to the levelling effect on capacity of inter user interference growing at the same rate as useful signal power.

6.6 Conclusion

A lower bound on the information rate for a single user MIMO system over a frequency-selective channel with imperfect CSI at the receiver (and no CSI at the transmitter) has been adapted from the SISO counterpart proposed by Medard in [9]. The bound on the information rate has been thoroughly investigated through numerical results in order to show the impact
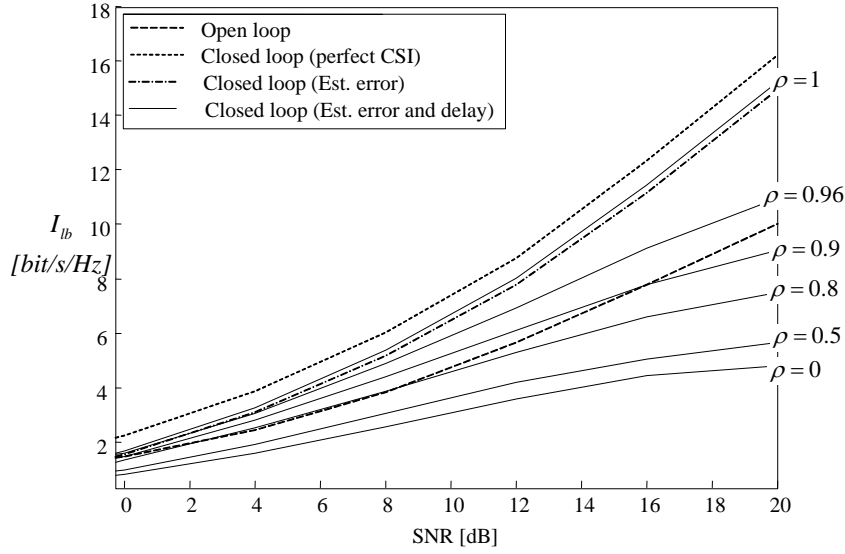
Figure 6.6: Average sum rate I_{lb} versus SNR for various coding strategies and temporal correlation coefficient of the fading process ρ.

of system parameters (e.g., length of the training sequence versus size of the data block, correlation properties of the training sequences) and channel characteristics (e.g., Doppler shift, spatial correlation) on the system performance.

Under the assumption of imperfect CSI at the receiver and no CSI at the transmitter, the extension of the analysis discussed above to a multiuser setting is straightforward. Therefore, here we focused on the downlink of a MIMO system in case of noisy and outdated CSI available at the transmitter. The selected case is of particular importance since, as opposed to the uplink, an achievable rate region and the sum-capacity for the downlink have been determined through a duality result based a non-linear precoding technique proposed by Costa in [7], that requires perfect CSI at the transmitter. Therefore, while the study of uplink under the assumption of imperfect CSI is a relatively easy task, an investigation of the information rate of the downlink under the same condition is an open problem. Thus, in this Chapter we derived a lower bound on the information rate of a specific linear precoding scheme for the downlink. The scheme under consideration has been proposed in [8] and essentially enforces a zero inter-user interference constraint through orthogonal linear precoding. The impact of the considered deviations from the ideal condition of perfect and instantaneous CSI have been considered both analytically and through simulations in order to test the robustness of this transmission strategy. From this analysis, performance of orthogonal precoding strongly depends on the correlation between successive channel realizations suggesting to restrict its use to channels varying slowly enough. Possible performance gain from using more complex channel modeling, estimation

and prediction remains an open subject for further study.

6.7   Appendix: computation of $\mathbf{Q}_{\widehat{\mathcal{H}}}$ from $\mathbf{Q}_{\hat{\mathbf{h}}}$

Let us at first consider time-domain transmission. The matrix $L_D N_R \times L_D N_T$ matrix $\widehat{\mathcal{H}}$ is the block-convolution matrix (here shown for $L_D = 5$ and $W = 3$ for illustration purposes):

$$\widehat{\mathcal{H}} = \begin{bmatrix} \hat{\mathbf{H}}[0] & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hat{\mathbf{H}}[1] & \hat{\mathbf{H}}[0] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hat{\mathbf{H}}[2] & \hat{\mathbf{H}}[1] & \hat{\mathbf{H}}[0] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{H}}[2] & \hat{\mathbf{H}}[1] & \hat{\mathbf{H}}[0] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \hat{\mathbf{H}}[1] & \hat{\mathbf{H}}[0] \end{bmatrix}. \tag{6.30}$$

The $L_D N_R \times L_D N_R$ covariance matrix $\mathbf{Q}_{\widehat{\mathcal{H}}} = E[(\widehat{\mathcal{H}} - \mathcal{H})(\widehat{\mathcal{H}} - \mathcal{H})^H]$ can be written as the block matrix

$$\mathbf{Q}_{\widehat{\mathcal{H}}} = \begin{bmatrix} \mathbf{Q}[0] & \mathbf{Q}[1] & \mathbf{Q}[2] & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}[1] & 2\mathbf{Q}[0] & 2\mathbf{Q}[1] & \mathbf{Q}[2] & \mathbf{0} \\ \mathbf{Q}[2] & 2\mathbf{Q}[1] & 3\mathbf{Q}[0] & 2\mathbf{Q}[1] & \mathbf{Q}[2] \\ \mathbf{0} & \mathbf{Q}[2] & 2\mathbf{Q}[1] & 3\mathbf{Q}[0] & 2\mathbf{Q}[1] \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}[2] & 2\mathbf{Q}[1] & 3\mathbf{Q}[0] \end{bmatrix} \tag{6.31}$$

where the $N_R \times N_R$ blocks $\mathbf{Q}[k]$ are defined as $\mathbf{Q}[k] = E[(\hat{\mathbf{H}}[i] - \mathbf{H}[i])(\hat{\mathbf{H}}[i+k] - \mathbf{H}[i+k])^H]$. For the general case with any L_D and W, the (i,j)th $N_R \times N_R$ block of the covariance matrix $\mathbf{Q}_{\widehat{\mathcal{H}}}$ reads: $\min\{\max(i,j), W - |i-j|)\mathbf{Q}[|i-j|]\}$. The blocks $\mathbf{Q}[k]$ can be computed from the HCRB matrix $\mathbf{Q}_{\hat{\mathbf{h}}}$. In fact, $\mathbf{Q}_{\hat{\mathbf{h}}}$ is a $W \times W$ block matrix such that the (i,j)th block, denoted as $\mathbf{Q}_{\hat{\mathbf{h}}}(i,j)$, is the $N_R N_T \times N_R N_T$ matrix $\mathbf{Q}_{\hat{\mathbf{h}}}(i,j) = E[\text{vec}\{\hat{\mathbf{H}}[i] - \mathbf{H}[i]\} \text{vec}\{\hat{\mathbf{H}}[j] - \mathbf{H}[j]\}^H]$, for $i,j = 0, .., W - 1$. $\mathbf{Q}_{\hat{\mathbf{h}}}(i,j)$ can in turn be written as a $N_T \times N_T$ block matrix composed of $N_R \times N_R$ blocks such that $E[(\hat{\mathbf{H}}[i] - \mathbf{H}[i])(\hat{\mathbf{H}}[j] - \mathbf{H}[j])^H]$ is equal to the sum of the blocks on its main diagonal.

For multicarrier transmission, the matrix $\widehat{\mathcal{H}}$ is block-diagonal and, following the lines of the Appendix of Chapter 3, can be written in terms of the corresponding quantity in time domain (6.30) as (denoting with a tilde the matrix in the frequency domain) $\widetilde{\widehat{\mathcal{H}}} = (\mathbf{\Theta} \otimes \mathbf{I}_{N_R})\widehat{\mathcal{H}}(\mathbf{\Theta}^H \otimes \mathbf{I}_{N_T})$, with $\mathbf{\Theta}$ representing here the $L_D \times L_D$ DFT matrix corresponding to the data subcarriers. Accordingly, the covariance matrix $\mathbf{Q}_{\widehat{\mathcal{H}}}$ for multicarrier transmission is readily obtained using the result above for time-domain transmission.

Bibliography

[1] H. Meyr, M. Moeneclaey and S. Fechtel, *Digital communication receivers*, Wiley, 1998.

[2] J. Baltersee, G. Flock and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2358-2368, Dec. 2001.

[3] H. Vikalo, B. Hassibi, B. Hochwald and T. Kailath, "On the capacity o frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2572-2583, Sept. 2004.

[4] D. Samardzija and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Trans. Signal Proc.*, vol. 51, no. 11 , pp. 2882-2890, Nov 2003.

[5] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684-702, June 2003.

[6] N. Jindal, S. Vishwanath and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 768-783, May 2004.

[7] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[8] Q. H. Spencer, A. L. Swindlehurst and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Proc.*, vol. 52, no. 2, pp. 461-471.

[9] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933-946, May 2000.

140   *Information rate with imperfect CSI*

[10] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-963, April 2003.

[11] T. Marzetta and B. Hochwald, "Fast Transfer of Channel State Information in Wireless Systems," submitted to *IEEE Trans. Signal Proc*. [also available on http://mars.bell-labs.com/cm/ms/what/mars/papers/channel_estimation/]

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

[13] J. G. Proakis, *Digital communications*, McGraw-Hill, 1995.

Chapter 7

Linear and non-linear precoding/equalization with long-term CSI at the transmitter

7.1 Introduction

IN order to achieve the high spectral efficiencies promised by the information theory over a MIMO link (see previous Chapter for details and references), different practical approaches have been proposed, e.g., space-time codes [1] and V-BLAST [2]. This Chapter is concerned with the optimization of the transceiver structure shown in fig. 7.1, that operates over a (single user) frequency-flat MIMO channel, under the constraint that the CSI available at the transmitter is limited to the long term features of the channel. In particular, as discussed in Chapter 2, this implies that the transmitter is provided with the second order statistics (i.e., the correlation matrix) of the channel. We refer to this condition as long-term CSI or, herein in short, LT-CSI, see also [3] [4] [5]. The assumption of LT-CSI at the transmitter is of crucial relevance for systems in which the fading amplitudes are sufficiently fast-varying to make the condition of instantaneous CSI (in short I-CSI) at the transmitter not realistic, as anticipated by the analysis in the previous Chapter (Sec. 6.4). The LT-CSI can be acquired by the transmitter either directly from measurements of the opposite link [7] or by feedback from the receiver. On the other hand, in the design of the receiver, the instantaneous realization of the channel matrix $\mathbf{H}$ (i.e., I-CSI) is assumed known (effects of channel estimation errors are studied by means of simulations).

The design of linear/non-linear precoding/equalization for a multiuser MIMO system based on LT-CSI at the transmitter is an open problem. The next Chapter is devoted to the study of linear precoding/equalization for the downlink of a MIMO system with channel aware scheduling and I-CSI at the transmitter.

Linear and non-linear precoding and equalization are considered in the scheme of fig. 7.1.
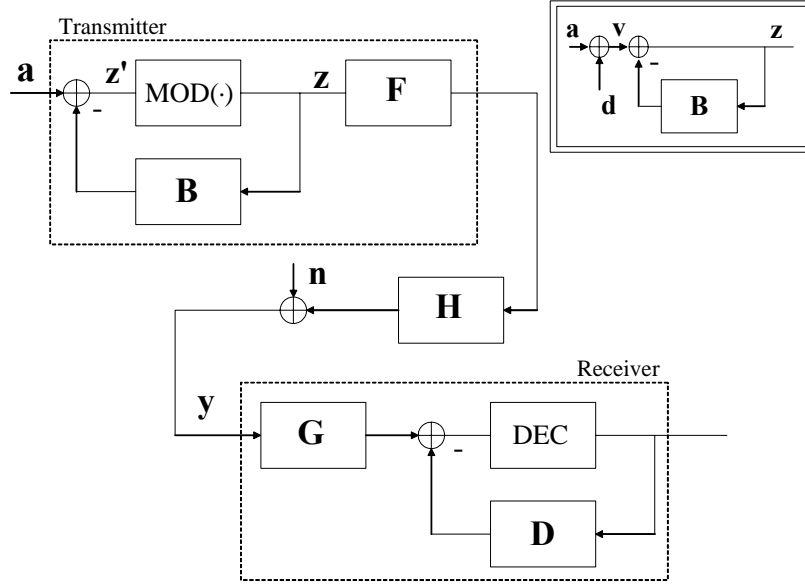
Figure 7.1: Block diagram of linear/non-linear precoding/equalization for a single user MIMO system.

As explained below, the non-linear block at the transmitter, named MOD(.), limits the dynamic range of the precoded sequence whereas the non-linear block at the receiver, DEC, is the traditional decision device. The structure reduces to known systems for specific constraints on the matrices $\{\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}\}$. For instance, imposing $\mathbf{B} = \mathbf{0}$ and $\mathbf{F} = \mathbf{I}$ the scheme reduces to a decision feedback equalizer (or equivalently to the V-BLAST receiver without optimal ordering [2] [8]); for $\mathbf{B} = \mathbf{0}$ and $\mathbf{D} = \mathbf{0}$ we have the linear precoding-linear equalization structure (LP-LD), studied in [9] [10], under the assumption of I-CSI at both the transmitter and the receiver; for $\mathbf{F} = \mathbf{I}$ and $\mathbf{D} = \mathbf{0}$ the Tomlinson-Harashima precoding (THP) structure proposed in [6] and studied for LT-CSI at the transmitter in [4] is obtained.

7.2 Linear/non-linear precoding

We focus on a MIMO wireless link with an equal number of transmit and receive antennas $N_T = N_R = N$. The $N \times 1$ data vector $\mathbf{a}$ (the time dependence of all the variables is implied) is composed of complex symbols taken from the $M-$QAM constellation, i.e., each entry a_i ($i = 1, .., N$) belongs to the set $\mathcal{A} = \{a^I + ja^Q | a^I, a^Q \in \{\pm 1, \pm 3, .., \pm \sqrt{M} - 1\}\}$. The data vector is passed through the non-linear part of the precoder defined by the $N \times N$ strictly upper triangular matrix $\mathbf{B}$ (i.e., $[\mathbf{B}]_{ii} = 0$). In order to stabilize the precoder, or equivalently to limit the dynamic range of the precoded sequence, a non-linear modulo-arithmetic operation (MOD) is introduced, as it is done in THP (see, e.g., [12] [13]). This operation performs a periodic mapping (or modulo reduction) of its input $\mathbf{z}'$ on the square region of the complex plane that

contains $\mathcal{A}$ and has side length $2\sqrt{M}$, i.e., $\mathcal{R} = \{z^I + jz^Q | z^I, z^Q \in (-\sqrt{M}, \sqrt{M}]\}$. In other words, $z_i = MOD(z_i') = z_i' + 2\sqrt{M}k_i$, where the real and imaginary parts of k_i are integers chosen to reduce $z_i \in \mathcal{R}$. Notice that there is only one k_i that satisfies this condition. It follows that the non-linear part of the precoder can be equivalently redrawn by deleting the block MOD and adding at the input an input-dependent vector $\mathbf{d}$ (see box in fig. 7.1) such that $d_i = 2k_i\sqrt{M}$. Therefore, the effective symbols input to the non-linear precoder are $\mathbf{v} = \mathbf{a} + \mathbf{d}$. After linear precoding with the $N \times N$ matrix $\mathbf{F}$ and propagation through the $N \times N$ radio channel $\mathbf{H}$, the received signal on each time-instant or each frequency of the data part of the block can be written as (we drop the dependence on the block ℓ for simplicity of notation):

$$\mathbf{y} = \mathbf{HFz} + \mathbf{n}, \tag{7.1}$$

where we recall that the circularly symmetric Gaussian noise has correlation $E[\mathbf{nn}^H] = \mathbf{R}_n$ and we set $E[\mathbf{zz}^H] = \mathbf{I}_N$. Notice that the latter assumption, also made in [6] and [4] to make the problem tractable, is not rigorously satisfied when $\mathbf{B} \neq \mathbf{0}$ since in this case $E[\mathbf{zz}^H]$ is a function of the unknown (i.e., design target) matrix $\mathbf{B}$. When optimizing the scheme of fig. 7.1, the power constraint $E[|||\mathbf{Fz}||^2] = \text{tr}\{\mathbf{FF}^H\} \leq P$ will be imposed. This condition is clearly satisfied when no linear precoding is employed ($\mathbf{F} = \mathbf{I}$). Notice that we are assuming equal power within the training and data part of the block.

7.3 Channel model: a brief review

According to Chapter 2, the beamforming and diversity scenarios adapted to a frequency-flat setting as the one considered here yield to:

1) *Diversity scenario:* the channel matrix $\mathbf{H}$ is assumed to be zero-mean (Rayleigh fading) circularly symmetric complex Gaussian distributed with a separable spatial correlation function

$$\mathbf{H} = [\mathbf{R}^{(R)}]^{H/2}\mathbf{H}_w[\mathbf{R}^{(T)}]^{1/2}, \tag{7.2}$$

where the correlation matrices $\mathbf{R}^{(R)}$ and $\mathbf{R}^{(T)}$ account for receive and transmit side spatial correlation and $\mathbf{H}_w$ is a matrix of independent identically distributed circularly symmetric complex Gaussian variables with unit power. For simplicity, the numerical evaluation of the performance of the presented algorithms will be carried out for an autoregressive model of the spatial correlation (see also Sec. 5.6.3). For later use, we remark that the correlation matrix of the channel (that is related to the LT-CSI available at the transmitter, see next Section) is

$$E[\mathbf{H}^H\mathbf{H}] = \mathbf{R}^{(T)} \cdot \text{tr}\{\mathbf{R}^{(R)}\} = N\mathbf{R}^{(T)}.$$

2) *Beamforming scenario:* the channel matrix $\mathbf{H}$ is

$$\mathbf{H} = \mathbf{A}(\boldsymbol{\alpha}^{(R)}) \, \text{diag}(\boldsymbol{\beta})\mathbf{A}(\boldsymbol{\alpha}^{(T)})^T, \tag{7.3}$$

and for simplicity we consider all paths with the same power, $\Omega_i = 1/d$. The latter assumption is considered for mathematical convenience and appears to be realistic in a frequency-flat scenario, where all the paths are likely to experience the same path loss and shadowing. For later use, we remark that the correlation matrix of the channel is

$$E[\mathbf{H}^H\mathbf{H}] = \frac{1}{d}\mathbf{A}(\boldsymbol{\alpha}^{(T)})^*\mathbf{A}(\boldsymbol{\alpha}^{(T)})^T. \tag{7.4}$$

7.4 MMSE-based precoding and equalization

At the receiver side, the received signal is linearly processed by the $N \times N$ matrix $\mathbf{G}$, passed through the feedback loop defined by the $N \times N$ strictly upper triangular matrix $\mathbf{D}$ and modulo reduced into $\mathcal{R}$ if $\mathbf{B} \neq \mathbf{0}$ (not shown in fig. 7.1). Here we optimize the general transceiver scheme of fig. 7.1 by minimizing the MSE between the variables at the input of the decision device and the effective data symbols $\mathbf{v} = \mathbf{a} + \mathbf{d}$ [6]. As previously discussed, we constrain the design of the operators at the transmitter side, i.e., of the matrices $\mathbf{F}$ and $\mathbf{B}$, to be based only on LT-CSI, represented by the second order statistics of channel and noise. In particular, the transmitter is given only the correlation matrix $E[\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}]$ [3] [5]. On the other hand, the operators $\mathbf{G}$ and $\mathbf{D}$ at the receiver side are allowed to depend directly on the I-CSI, i.e., on the channel matrix $\mathbf{H}$. Furthermore, we will assume perfect error recovery at the output of the decision device as it is usually done in the literature on decision feedback equalization [14]. The effect of error propagation will be investigated in Sec. 7.5 through simulations.

We now proceed with the derivation of the optimum precoding and equalization matrices. Since the vector at the input of the decision device can be written as $\mathbf{G}\mathbf{y} - \mathbf{D}\mathbf{v}$ (recall the assumption of no error propagation made above), the design problem can be stated as

$$\{\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}\} = \arg \min_{\{\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}\}} E[||\mathbf{G}\mathbf{y} - \mathbf{D}\mathbf{v} - \mathbf{v}||^2] \tag{7.5}$$
$$\text{s.t. } E[||\mathbf{F}\mathbf{z}||^2] \leq P,$$

we will show below how we take into account the different types of CSI's at the transmitter and the receiver. From fig. 7.1 one can easily show that $\mathbf{v} = \mathbf{C}\mathbf{z}$ where $\mathbf{C} = \mathbf{I} + \mathbf{B}$, so that

$$MSE(\mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{D}) = E[||\mathbf{G}\mathbf{y} - \mathbf{E}\mathbf{C}\mathbf{z}||^2] = E[||\mathbf{G}\mathbf{y} - \mathbf{K}\mathbf{z}||^2] \tag{7.6}$$

with $\mathbf{E} = \mathbf{I} + \mathbf{D}$. The upper triangular (with unit diagonal) feedback matrices $\mathbf{C}$ and $\mathbf{E}$ (or equivalently $\mathbf{B}$ and $\mathbf{D}$) can not be independently identified using the MMSE criterion. In the following we will thus set $\mathbf{K} = \mathbf{E}\mathbf{C}$ and restate (7.6) as $MSE(\mathbf{F}, \mathbf{G}, \mathbf{K})$. We remark that $\mathbf{K}$ is still an upper triangular matrix with unit diagonal.

From the standard theory of Wiener linear filtering, we get $\mathbf{G} = \mathbf{K}E[\mathbf{z}\mathbf{y}^H]E[\mathbf{y}\mathbf{y}^H]^{-1}$ and after algebraic manipulations

$$\begin{aligned}\mathbf{G} &= \mathbf{K}\mathbf{F}^H\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\mathbf{F}\mathbf{F}\tilde{\mathbf{H}}^H + \frac{N}{P}\mathbf{I}_N)^{-1}\cdot\mathbf{R}_n^{-H/2} = \\ \mathbf{G} &= \mathbf{K}\mathbf{Q}^{-1}\mathbf{F}^H\tilde{\mathbf{H}}^H\cdot\mathbf{R}_n^{-H/2} \end{aligned} \tag{7.7}$$

where $\mathbf{Q} = \mathbf{F}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{F} + N/P \mathbf{I}_N$ and $\tilde{\mathbf{H}} = \mathbf{R}_n^{-H/2} \mathbf{H}$. The result (7.7) states that the optimum linear filter at the front-end of the receiver performs the whitening of the received signal and then applies a linear operator that has the classical Wiener structure. Substituting (7.7) into the expression of $MSE(\mathbf{F}, \mathbf{G}, \mathbf{K})$ we obtain

$$MSE(\mathbf{F}, \mathbf{K}) = \text{tr}\{\mathbf{K} \mathbf{Q}^{-1} \mathbf{K}^H\}. \tag{7.8}$$

In our framework, minimizing (7.8) with respect to $\mathbf{K}$ leads to different results depending on the way the matrix $\mathbf{K}$ is factorized into the transmitter ($\mathbf{C}$) and receiver ($\mathbf{E}$) part. Here, we consider two cases:

    1) *non-linear equalization* ($\mathbf{C} = \mathbf{I} \Rightarrow \mathbf{K} = \mathbf{E}$): since the receiver has access to the I-CSI $\mathbf{H}$, the feedback matrix $\mathbf{K} = \mathbf{E}$ is obtained as

$$\mathbf{E} = \mathbf{V} \mathbf{Q}^{1/2}, \tag{7.9}$$

where $\mathbf{V}$ is a diagonal matrix that scales to unity the elements on the main diagonal of $\mathbf{K}$;

    2) *non-linear precoding* [4] ($\mathbf{E} = \mathbf{I} \Rightarrow \mathbf{K} = \mathbf{C}$): since the transmitter is given only the LT-CSI

$$\mathbf{R}_H = E[\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}] = E[\mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}] \tag{7.10}$$

we can not minimize (7.8). Instead, it is reasonable to consider $E[\text{tr}\{\mathbf{K} \mathbf{Q}^{-1} \mathbf{K}^H\}]$ as the loss function. It can be shown that

$$E[\text{tr}\{\mathbf{K} \mathbf{Q}^{-1} \mathbf{K}^H\}] \geq \text{tr}\{\mathbf{K} \bar{\mathbf{Q}}^{-1} \mathbf{K}^H\}, \tag{7.11}$$

where $\bar{\mathbf{Q}} = \mathbf{F}^H E[\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}] \mathbf{F} + N/P \mathbf{I}_N$ (see Appendix). Therefore, similarly to the approach of [3] and [4], we minimize the lower bound $\overline{MSE}(\mathbf{F}, \mathbf{K}) = \text{tr}\{\mathbf{K} \bar{\mathbf{Q}}^{-1} \mathbf{K}^H\}$ obtaining

$$\mathbf{C} = \mathbf{V} \bar{\mathbf{Q}}^{1/2}, \tag{7.12}$$

where $\mathbf{V}$ is the scaling matrix as in (7.9).

    After substitution of (7.9) if $\mathbf{C} = \mathbf{I}$ or (7.12) if $\mathbf{E} = \mathbf{I}$ according to the two cases discussed above, we should in principle minimize $\overline{MSE}(\mathbf{F}, \mathbf{K}) = \text{tr}\{\mathbf{K} \bar{\mathbf{Q}}^{-1} \mathbf{K}^H\}$ with respect to the transmit precoding matrix $\mathbf{F}$. To make the problem tractable and obtain a solution independent on $\mathbf{K}$, we minimize $\text{tr}\{\bar{\mathbf{Q}}^{-1}\}$ instead. In other words, the matrix $\mathbf{F}$ is designed by assuming that neither non-linear precoding nor non-linear equalization is included in the transceiver ($\mathbf{K} = \mathbf{I}$). Nonetheless, simulation results show that the so obtained linear precoder performs satisfactorily even for $\mathbf{K} \neq \mathbf{I}$ (see Sec. 7.5). The precoder $\mathbf{F}$ can thus be obtained following the steps outlined in [9]. It is

$$\mathbf{F} = \mathbf{U} \mathbf{\Phi} \tag{7.13}$$

where $\mathbf{U}$ is obtained from the eigenvalue decomposition of the LT-CSI $E[\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}] = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$ and $\mathbf{\Phi}$ is a diagonal matrix such that

$$|[\mathbf{\Phi}]_{ii}|^2 = \left(\frac{N + \sum_{n=1}^{\bar{N}} \lambda_{nn}^{-1}}{P/N \sum_{n=1}^{\bar{N}} \lambda_{nn}^{-1/2}} \lambda_{ii}^{-1/2} - \frac{N}{\lambda_{ii} P} \right)^+, \tag{7.14}$$

where $(x)^+ = \max(x, 0)$ and $\bar{N} \leq N$ is such that $|[\boldsymbol{\Phi}]_{nn}|^2 > 0$ for $n \in [1, \bar{N}]$ and $|[\boldsymbol{\Phi}]_{nn}|^2 = 0$ for all other n. Therefore, the optimal precoder modulates each signal in the spatial domain by an eigenvector of the LT-CSI matrix $\mathbf{R}_H$ of the channel. It is worth emphasizing that the correlation matrix $\mathbf{R}_H$, or its eigenvalue decomposition $\{\mathbf{U}, \boldsymbol{\Lambda}\}$, has to be updated only occasionally at the transmitter, e.g., by a low rate feedback channel in a FDD link, since it is assumed to be invariant over a large time scale. Recall from Chapter 2 that temporal variations of $\mathbf{R}_H$ (or equivalently $\mathbf{R}^{(T)}$) are likely to be caused by the movements of the transmitting array and these can be assumed to be small enough (compared to the geometry of environment and arrays) to guarantee the invariance of LT-CSI across multiple symbols.

Some remarks on the results of the optimization (7.7), (7.9)-(7.12) and (7.13) are in order. *i*) In case we relax the assumption of LT-CSI at the transmitter, i.e., we allow the matrices $\mathbf{C}$ and $\mathbf{F}$ to depend on the instantaneous CSI, we get $\mathbf{K} = \mathbf{I}$ and $\mathbf{F}$ and $\mathbf{G}$ coincide with the results derived in [9]. In other words, if both sides of the link have access to the channel matrix $\mathbf{H}$, the setting that minimizes the MSE (7.5) results in linear precoding and equalization. In this case, $\mathbf{F}$ and $\mathbf{G}$ are obtained from the singular value decomposition of the channel matrix $\tilde{\mathbf{H}}$, as it can be inferred from (7.7) and (7.13). *ii*) Setting $\mathbf{F} = \mathbf{I}$ and $\mathbf{E} = \mathbf{I}$ leads to the THP followed by a MMSE residual linear equalizer derived in [4]. 3) Following the approach of [9], the linear precoder $\mathbf{F}$ can be obtained alternatively by minimizing $\mathrm{tr}\{\bar{\mathbf{Q}}^{-1}\}$ subject to a peak power constraint or by maximizing the information rate, i.e., minimizing $\det\{\bar{\mathbf{Q}}^{-1}\}$. These alternatives will be further pursued in Sec. 7.4.2.

7.4.1 Long-term linear precoding as a whitening operation

Here we would like to further investigate the properties of the optimal linear precoding in (7.13). Notice that precoding (7.13) has been shown to maximize the capacity in [3], minimize the probability of error for orthogonal space-time codes in [5] and, as explained above, to minimize the MSE between the transmitted vector $\mathbf{z}$ and the decision variables for linear equalizers, under the assumption of LT-CSI at the transmitter. To start with, it is easy to show that the LT-CSI matrix $\mathbf{R}_H$ reads

$$\mathbf{R}_H = \mathbf{R}^{(T)} \cdot \mathrm{tr}\{\mathbf{R}_n^{-1}\mathbf{R}^{(R)}\}. \tag{7.15}$$

In particular, if we assume spatially white noise it simplifies as $\mathbf{R}_H = N/\sigma_n^2 \mathbf{R}^{(T)}$.

According to (7.15), the linear precoder performs the beamforming of the transmitted vector $\mathbf{a}$ along the *spatial modes* $\mathbf{U}$ of the channel (recall the discussion in Chapter 4). It is then clear that the role of the linear pre-equalizer $\mathbf{F}$ is that of cancelling the channel correlation at the transmitter side. Decorrelating the transfer matrix $\mathbf{H}$ is known to guarantee enhanced link performance [11]. Fig. 7.2-(a) shows the signal model (transmitter, channel and noise whitening) by explicitly including the channel correlation matrices. The cascade of $\mathbf{F}$ and $\mathbf{R}^{(T)1/2}$ is
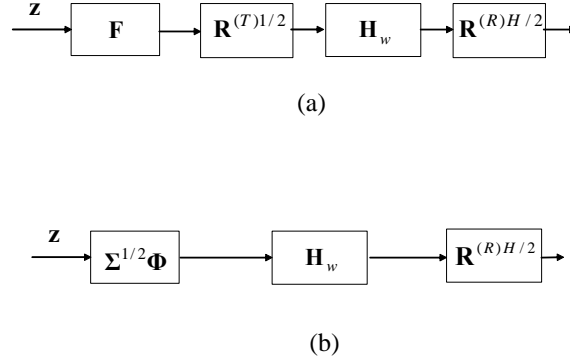
Figure 7.2: (a) Block diagram of the signal model (transmitter, channel and noise whitener). (b) Equivalent signal model that shows the role of linear precoding.

a diagonal matrix since $\mathbf{R}^{(T)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H$ (with $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}/\operatorname{tr}\{\mathbf{R}_n^{-1}\mathbf{R}^{(R)}\}$) and

$$\mathbf{R}^{(T)1/2}\mathbf{F} = \boldsymbol{\Sigma}^{1/2}\mathbf{U}^H\mathbf{U}\boldsymbol{\Phi} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Phi}. \tag{7.16}$$

Therefore, the signal model can be simplified as in fig. 7.2-(b), proving by this simple reasoning that the correlation matrix at the transmitter side is diagonalized by the precoder $\mathbf{F}$. The channel correlation at the receiver side, $\mathbf{R}^{(R)H/2}$ has to be dealt with at the receiver.

7.4.2 Different power allocation over the channel modes

In this Section, we consider two alternative power allocation schemes over the spatial modes that constraint the total radiated power as $E[||\mathbf{F}\mathbf{z}||^2] \leq P$.

1. *Waterfilling*: maximizing the capacity [3] or minimizing the probability of error for orthogonal space-time codes [5] we get the classical waterfilling solution:

$$|[\boldsymbol{\Phi}]_{ii}|^2 = \left(\frac{P/N + \sum_{n=1}^{\bar{N}} \lambda_{nn}^{-1}}{P/N\bar{N}} - \frac{1}{\lambda_{ii}P/N} \right)^+, \tag{7.17}$$

where $(x)^+ = \max(x, 0)$ and $\bar{N} \leq N$ is such that $|[\boldsymbol{\Phi}]_{nn}|^2 > 0$ for $n \in [1, \bar{N}]$ and $|[\boldsymbol{\Phi}]_{nn}|^2 = 0$ for all other n.

2. *Uniform power allocation*: assuming that the same power is transmitted from each eigenmode

$$|[\boldsymbol{\Phi}]_{ii}|^2 = P/N. \tag{7.18}$$

Recall that in case the MMSE is minimized, we get the power allocation in (7.14), referred to in the following as *MMSE waterfilling*.

7.5   Simulation results

The performance of the precoder/decoder structure of fig. 7.1 is first evaluated in terms of un-coded SER for a 16-QAM constellation ($M = 16$), $N = 8$ antennas and the diversity model. We compare the performance of the general setting with $\mathbf{K} = \mathbf{C}$ or $\mathbf{K} = \mathbf{E}$, referred to as NP-LE (non-linear precoding, linear equalization) and LP-NE (linear precoding, non-linear equalization) respectively, with the following special cases: 1) $\mathbf{B} = \mathbf{0}$, $\mathbf{F} = \mathbf{I}$: MMSE-V-BLAST receiver (or MMSE-DFE) with no stream ordering; 2) $\mathbf{B} = \mathbf{0}$ and $\mathbf{D} = \mathbf{0}$: linear precoding-linear equalization (LP-LE); 3) $\mathbf{F} = \mathbf{I}$ and $\mathbf{D} = \mathbf{0}$: THP with MMSE residual equalization [4]. We further limit the study to the spatially white noise case. Notice that appropriate scaling of the transmitted vector $\mathbf{a}$ is performed to compensate for the power amplification due to non-linear precoding [12] [13] so that the performance comparison is based on equal total average transmitted power P (see Sec. 7.2). It is worth emphasizing again that that all the schemes taken into account perform precoding based on LT-CSI at the transmitter, except DFE that does not entail any processing at the transmitter side. Moreover, where not stated otherwise, we consider linear precoding with power allocation according to MMSE waterfilling (7.14).

Non-linear precoding (or equalization) causes the N transmitted data streams to have different error rates. This problem can be tackled by, e.g., coding across the different streams or using more powerful codes on weaker streams. Here we limit the analysis to uncoded transmission, leaving the issues raised by the introduction of coding in the considered scheme (e.g., soft/hard equalization, horizontal/vertical layering) to further investigations. In the following, the SER is thus averaged over the N transmitted data streams.

In fig. 7.3 the SER is plotted as a function of SNR for $\rho_T = 0.4$ (left) and $\rho_T = 0.8$ (right) ($\rho_R = 0.4$). Apart from the expected performance degradation due to the decreased spatial diversity, it can be seen that the benefits (if any) of precoding based on LT-CSI compared to the DFE receiver are more relevant for increasing values of ρ_T. This is intuitively clear since for $\rho_T = 0$ the LT-CSI $E[\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}] = N/\sigma_n^2 \mathbf{R}^{(T)} = N/\sigma_n^2 \mathbf{I}_N$ does not bring any side information that can be exploited by the transmitter to improve the performance of the link. In this case, it is $\mathbf{F} = \mathbf{I}_N$ and $\mathbf{B} = \mathbf{0}$ from (7.13) and (7.12) respectively. Furthermore, it can be concluded that the LP-NE gives the best performance in terms of uncoded SER. Simulations show that similar gain can be obtained even for $SNR > 30dB$ (not shown in the figure). To study the effect of error propagation, the performance of genie-aided (i.e., perfect past decisions) DFE and LP-NE are shown as dashed lines. It is important to remark that when comparing the performance of the schemes of interest, other considerations, apart from the SER, should be taken into account. For instance, it is well-known that non-linear precoding at the transmitter causes a relevant increase of the dynamic range at the input of the decision device that can limit its feasibility [15].

To have a clearer understanding of the role of the spatial correlation at the transmitter side on the performance of different schemes, fig. 7.4 plots the uncoded SER as a function of ρ_T

for $SNR = 20dB$, $N = 8$ and $\rho_R = 0.4$. In accordance with the previous discussion, all precoding schemes outperform the DFE for ρ_T large enough. Moreover, the LP-NE structure shows the lowest SER except for very high values of ρ_T, where it is slightly outperformed by the THP scheme.

We now want to assess the effects of an imperfect I-CSI at the receiver. To this end, we assume that for the design of **G** (7.7) and **D** (7.9) only a noisy version of the channel matrix is available. A conventional UML estimate of the channel is carried out at the receiver with $L_P = N$ and power used for training equal to the power used for the data part of the burst. The SER is plotted as a function of SNR in fig. 7.5 for $\rho_T = 0.4$ (left) and $\rho_T = 0.8$ (right) respectively ($\rho_R = 0.4$, $N = 8$). The same considerations discussed for perfect I-CSI apply also to the case in which the channel estimation error is taken into account, except for the performance degradation due to the imperfect I-CSI. In particular, LP-NE still gives the best performance uniformly with respect to the SNR.

Let us now consider the beamforming model. According to (7.4) the spatial correlation at the transmitter side is mainly related to the number of resolvable angle of departure r_T (i.e., the number of spatial modes, see Chapter 4). To complement the analysis carried out in fig. 7.4 for the diversity model, fig. 7.6 shows the uncoded SER against r_T for $SNR = 20dB$, $N = 8$, $r_R = 8$ and the beamforming model. Increasing r_T produces two simultaneous effects: increasing the rank of the channel matrix ($\text{rank}(\mathbf{H}) = \min(r_T, r_R) = r_T$) and decreasing the spatial correlation at the transmitter side (i.e., increasing the spatial diversity). The first effect tends to reduce the SER (multiplexing gain, see, e.g., [2]) whereas the latter tends to lessen the benefits of precoding. Accordingly, the SER of DFE is monotonically decreasing while the SER of the different precoding schemes, with the only exception of THP, presents a U-shape. For a wide range of values of r_T, LP-NE results in the lowest SER as for the diversity case. Nevertheless, for $r_T = 1$ (high spatial correlation) and $r_T \geq 7$, THP presents the best performance.

In summary, precoding with long term channel state information appears to be advantageous in dense multipath channels (as for diversity model) with relatively large correlation at the transmitter ($\rho_T \geq 0.2$) or in sparse multipath channels (as for beamforming model). Moreover, the experimental results show that the most promising scheme is LP-NE, also considering the practical limitations of non-linear precoding [15]. The preferred scheme essentially adds a linear precoder to a modified BLAST receiver, where the feedforward filter **G** and the feedback filter **D** are designed by taking into account the precoder **F** according to (7.7) and (7.9).

The preferred scheme (LP-NE) is now evaluated for the three different power allocation strategies discussed in Sec. 7.4.2. Fig. 7.7 shows the uncoded SER versus SNR for the diversity model and $\rho_T = 0.8$ and $\rho_R = 0.4$ while fig. 7.8 plots the uncoded SER versus ρ_T for $SNR = 20dB$ and $\rho_R = 0.4$ (analogous behavior as a function of ρ_T is obtained lower SER as well). LP-NE guarantees better performance as compared to the DFE over the entire range
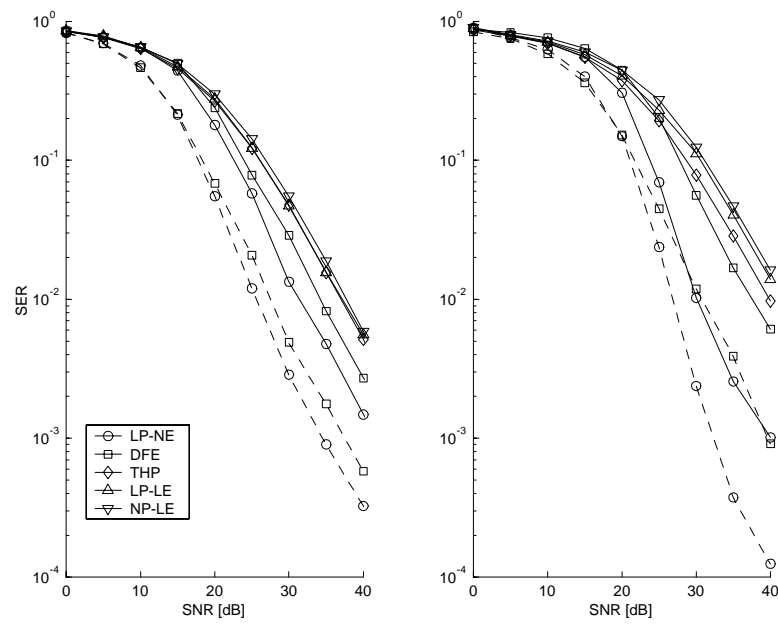
Figure 7.3: SER versus SNR for $\rho_T = 0.4$ (left) and $\rho_T = 0.8$ (right) ($\rho_R = 0.4$, $N = 8$). Dashed lines represent the case with genie-aided decision feedback.
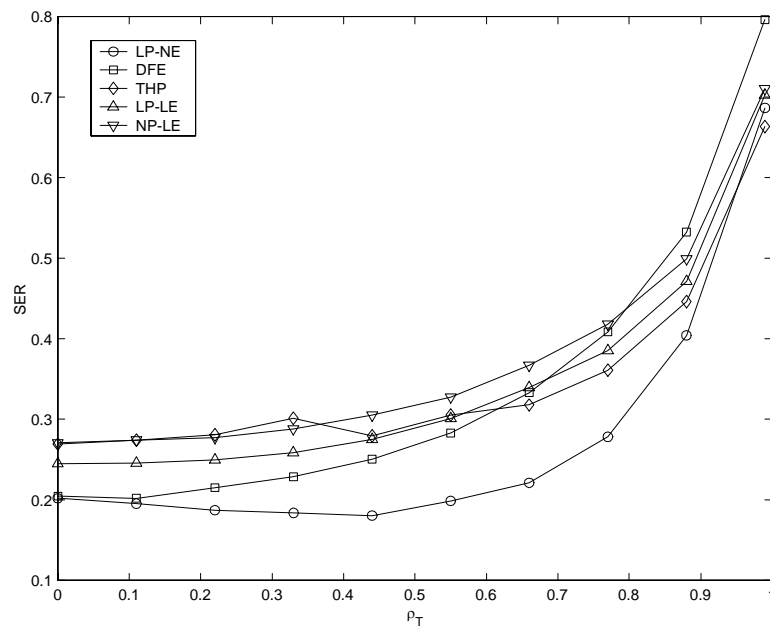

Figure 7.4: Effect of spatial correlation at the transmitter side on the uncoded SER ($SNR = 20dB$, $N = 8$, $\rho_R = 0.4$).
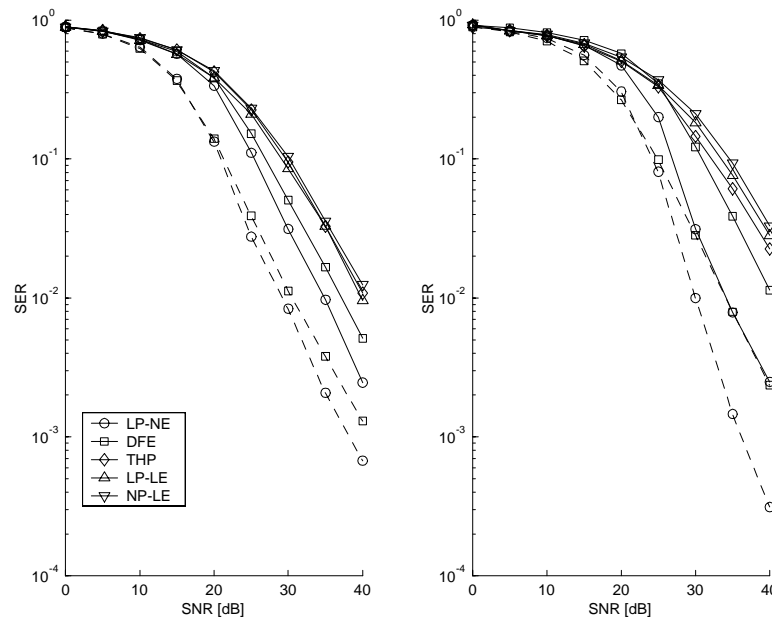
Figure 7.5: SER vs. SNR for $\rho_T = 0.4$ (left) and $\rho_T = 0.8$ (right) in case of imperfect I-CSI at the receiver ($\rho_R = 0.4$, $N = 8$). Dashed lines represent the case with genie-aided decision feedback.
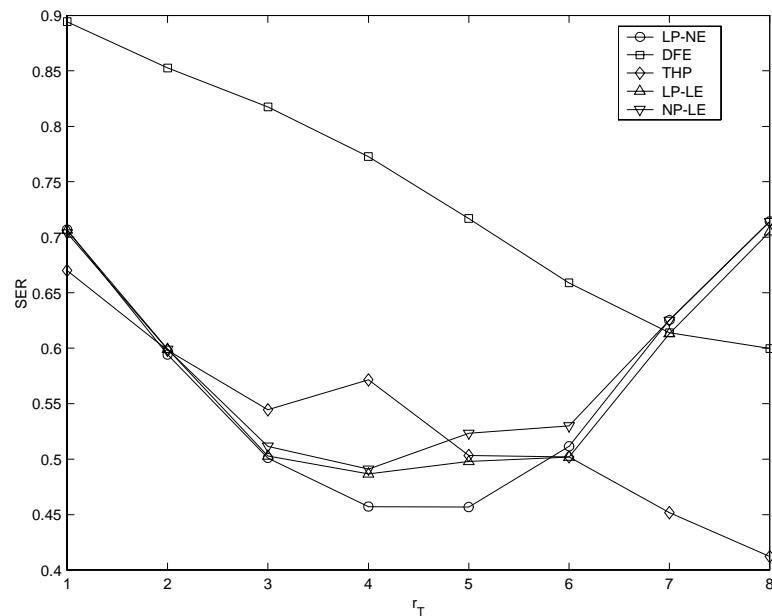

Figure 7.6: SER versus r_T for the beamforming model ($SNR = 20dB$, $N = 8$, $r_R = 8$).
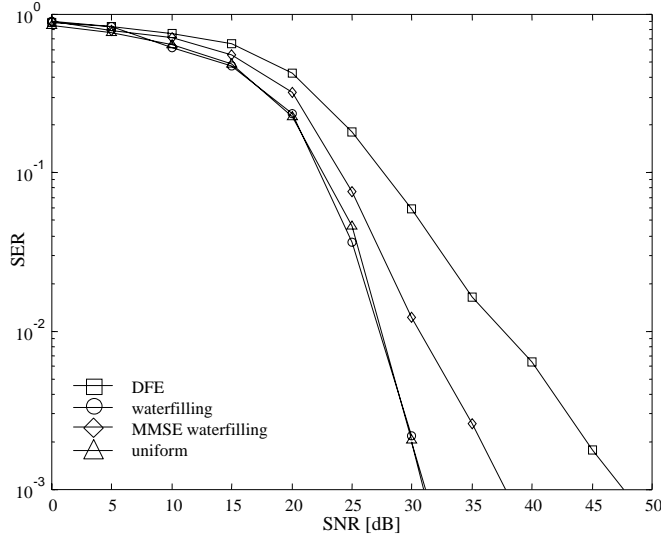
Figure 7.7: SER versus SNR for DFE and LP-NE with different power allocation schemes ($\rho_T = 0.8$, $\rho_R = 0.4$ and $N = 8$).

of correlation ρ_T and SNR considered. In particular, for $SER = 10^{-3}$, pre-equalization with waterfilling provides around $15dB$ gain in SNR. Moreover, waterfilling power allocation only slightly outperforms the uniform power allocation, that has the advantages of simplicity and reduced feedback (only the eigenvectors have to be transmitted by the receiver). Note that if the constellation size was allowed to vary with allocated power, then waterfilling solution could potentially increase its SNR margin as compared to the uniform power scheme. The degradation of the MMSE waterfilling scheme (around $5dB$ for $SER = 10^{-3}$) can be explained by recalling from Sec. 7.4 that the corresponding design assumes a linear equalizer (not a DFE receiver).

For $\rho_T = 0$, the transmitter can not capitalize on the LT-CSI $\mathbf{R}_H = N/\sigma_n^2 \mathbf{I}$ to improve the performance of the link ($\mathbf{F} = \mathbf{I}$) so that DFE and LP-NE have the same SER. For large ρ_T the performance of all the methods degrade because of the decreased spatial diversity and the performance gap among the methods shrinks as the channel matrix $\mathbf{H}$ becomes rank-deficient.

7.6   Conclusion

A transceiver structure for frequency flat MIMO channels that includes linear/non-linear pre-coding/equalization has been studied under the assumption that the CSI available at the transmitter is limited to the second order statistics of channel and noise (long-term CSI). Simulations have shown that relevant benefits can be obtained by exploiting the long term channel state information at the transmitter in both dense multipath channels with relatively large correlation
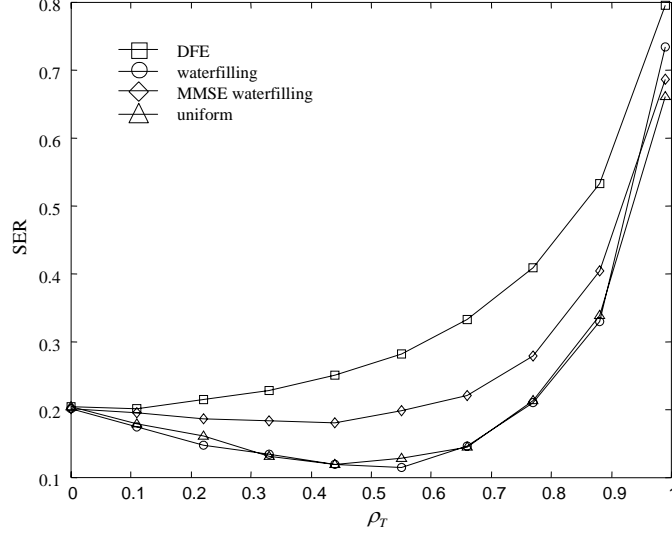
Figure 7.8: SER versus ρ_T for DFE and LP-NE ($SNR = 20dB$, $N = 8$, $\rho_R = 0.4$).

at the transmitter side and in sparse multipath channels. Moreover, the preferred scheme essentially adds a linear precoder to a modified BLAST receiver.

7.7 Appendix: proof of (7.11)

The inequality $E[\text{tr}\{\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H\}] \geq \text{tr}\{\mathbf{K}E[\mathbf{Q}]^{-1}\mathbf{K}^H\} = \text{tr}\{\mathbf{K}\bar{\mathbf{Q}}^{-1}\mathbf{K}^H\}$ directly follows from the Jensen's inequality once the function $\text{tr}\{\mathbf{K}\mathbf{Q}^{-1}\mathbf{K}^H\}$ is proved to be convex in the positive definite matrix $\mathbf{Q}$. To show the convexity of the function of interest, it is sufficient to demonstrate that $\mathbf{k}^H(\lambda\mathbf{Q}_1+(1-\lambda)\mathbf{Q}_2)^{-1}\mathbf{k} \leq \lambda\mathbf{k}^H\mathbf{Q}_1^{-1}\mathbf{k}+(1-\lambda)\mathbf{k}^H\mathbf{Q}_2^{-1}\mathbf{k}$, where $0 \leq \lambda \leq 1$ and $\mathbf{k}$ is any vector. The aforementioned condition can be stated as $(\lambda\mathbf{Q}_1 + (1 - \lambda)\mathbf{Q}_2)^{-1} \leq \lambda\mathbf{Q}_1^{-1} + (1 - \lambda)\mathbf{Q}_2^{-1}$ or equivalently as (th. 7.7.3 of [16])

$$\varrho((\lambda\mathbf{Q}_1 + (1 - \lambda)\mathbf{Q}_2)(\lambda\mathbf{Q}_1^{-1} + (1 - \lambda)\mathbf{Q}_2^{-1})) \geq 1, \tag{7.19}$$

where $\varrho(\cdot)$ denotes the spectral radius. After simple manipulations, we obtain

$$\varrho(\mathbf{Q}_1\mathbf{Q}_2^{-1} + \mathbf{Q}_2\mathbf{Q}_1^{-1}) \geq 2. \tag{7.20}$$

Since $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are hermitian matrices (in particular, they are positive definite), we can find a nonsingular matrix $\mathbf{Z}$ such that $\mathbf{Q}_1 = \mathbf{Z}\mathbf{Z}^H$ and $\mathbf{Q}_2 = \mathbf{Z}\mathbf{\Theta}\mathbf{Z}^H$, where $\mathbf{\Theta} = \text{diag}([\Theta_1 \cdots \Theta_N])$ is diagonal and Θ_i is real and positive (th. 7.6.3 and 7.6.5 of [16]). It follows that $\mathbf{Q}_1\mathbf{Q}_2^{-1} + \mathbf{Q}_2\mathbf{Q}_1^{-1} = \mathbf{Z}(\mathbf{\Theta} + \mathbf{\Theta}^{-1})\mathbf{Z}^{-1}$, which implies that (7.20) becomes

$$\Theta_i + 1/\Theta_i \geq 2 \quad \forall i = 1, 2, .., N, \tag{7.21}$$

that is clearly satisfied $\forall\Theta_i > 0$.

Bibliography

[1] V. Tarokh, N. Seshadri and A.R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998.

[2] G. D. Golden, C. J. Foschini, R. A. Valenzuela and P. W. Wolnianski, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics letters*, vol. 35, no. 1, pp. 14-16, Jan. 1999.

[3] M. T. Ivrlac, T. P. Kurpjuhn, C. Brunner and W. Utschick, "Efficient use of fading correlations in MIMO systems," *Proc. VTC 2001*, vol.4, pp. 2763-2767, 2001.

[4] R. F. H. Fischer, C. Windpassinger, A. Lampe and J. B. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," *Proc. International Zurich Seminar on Broadband Communications*, pp. 7_1-7_6, 2002.

[5] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Comm. Letters*, vol. 6, pp. 239-241, June 2002.

[6] R. F. H. Fischer, C. Windpassinger, A. Lampe and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," *Proc. of 4 ITG Conference on Source and Channel Coding*, pp. 139-147, January 2002.

[7] B. M. Hochwald and T. L. Marzetta, "Adapting a downlink array from uplink measurements," *IEEE Trans. Signal Proc.*, vol. 49, no. 3, pp. 642-653, March 2001.

[8] G. Ginis and J. M. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Comm. Letters*, vol. 5, no. 9, pp. 364-366, Sept. 2001.

[9] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Proc.,* vol. 50, no. 5, pp. 1051-1064, May 2002.

[10] A. Scaglione, G. B. Giannakis and S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: unification and optimal design," *IEEE Trans. Signal Proc.,* vol. 47, no. 7, pp. 1987-2006, July 1999.

[11] Shiu D. G. J. Foschini and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502-513, March 2000.

[12] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774-780, Aug. 1972.

[13] M. Tomlinson, "New automatic equaliser employing modulo arithmetic", *Electronic Letters*, pp. 138–139, March 1971.

[14] C. Tidestav, A. Ahlen and M. Sternad, "Realizable MIMO decision feedback equalizers: structure and design," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 121-133, Jan. 2001.

[15] R. F. H. Fischer, R. Tzschoppe and J. B. Huber, "Signal shaping for peak-power and dynamics reduction in transmission schemes employing precoding," *IEEE Trans. Comm.,* vol. 50, no. 5, pp. 736-741, May 2002.

[16] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, 1996.

Chapter 8

Channel aware scheduling for downlink with orthogonal precoding and fairness constraints

8.1 Introduction

IN the downlink of a broadcast (downlink) fading channel, the base station can capitalize on multiuser diversity provided by independent fading realizations across different users. Channel aware scheduling is a technique that allows to achieve this goal by appropriately timing transmission to a subset of one or more users in each available time (code/frequency) slot. Scheduling is performed according to the knowledge of the CSI available at the scheduler with the general goal of granting transmission to the users that have instantaneous channel near the peak [1].

If base station and users are equipped with a single antenna, it has been shown that transmission to the user with the strongest channel is a strategy that achieves channel capacity [2]. However, if the base station is equipped with an antenna array, more users can be served simultaneously in the same time slot. In particular, if the base station has N_T antennas, up to N_T users can be allocated in the same time slot with controlled interference. In this case, transmission to a single user is not the optimal solution and the design of the scheduler becomes more complicated depending on the beamforming and power allocation strategy [3] [4].

In MIMO systems (i.e., antenna array at both base station and terminals), the scheduler can leverage on another degree of freedom since each user can be assigned to multiple spatial channels [5]. In fact, if each user has n_R receiving antennas ($n_R \leq N_T$), the base station can grant up to n_R spatial channels (out of the available N_T) to any user [6]. The design of the scheduler has to take into account the processing performed at the transmitter (e.g., linear precoding and power allocation) and the receivers (e.g., linear equalizer).

Scheduling with linear processing at the base station that simply associates each spatial
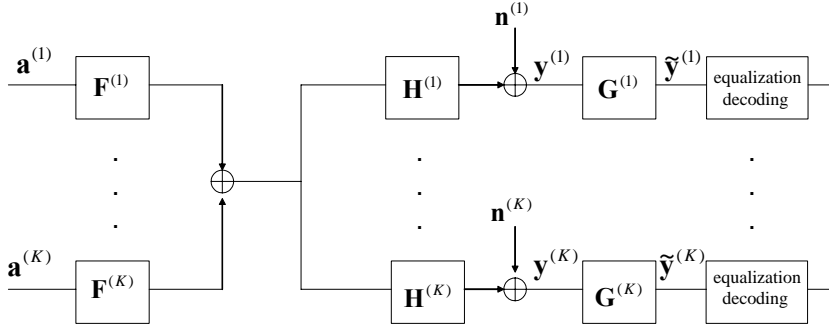
Figure 8.1: Block diagram of a broadcast channel with linear interfaces at the transmitter (base station) and receivers (users).

channel with a transmitting antenna and linear interfaces at the receivers has been considered in [5] (zero forcing equalizer) and [7] (MMSE equalizer). In this Chapter, we consider channel aware scheduling for orthogonal linear precoding at the base station (see Chapter 6) and linear zero-forcing equalizers at the receivers [8]. Using this transmitting/receiving strategy, the spatial channels intended for a given user do not interfere neither with the signal destined to other users nor among themselves. Here we consider perfect CSI at both transmitter and receiver.

Channel aware scheduling for orthogonal precoding has been first studied in [8]. Here we set the problem in a novel mathematical framework and propose a scheduling algorithm that is shown by simulation to guarantee superior performance in terms of sum capacity. Moreover, fairness constraints inspired by the proportional fair criterion [9] [10] are introduced in the scheduling process in order to guarantee the desired long term fairness properties.

8.2 Review of the signal model and problem formulation

The broadcast channel with linear interfaces at the transmitter and receivers is depicted in fig. 8.1. Let $\breve{\mathcal{K}}$ be the set of $\breve{K}$ available users. The ith user is equipped with an antenna array of $n_R^{(i)}$ elements and the base station with N_T antennas. The subset of K_ℓ users that are served by the base station within the ℓth time slot is denoted as $\mathcal{K}_\ell \subseteq \breve{\mathcal{K}}$ and its element indexed by $k = 1, 2, ..., K_\ell$.

The scheduler allocates $p_\ell^{(k)} \leq n_R^{(k)}$ spatial channels to the kth user so that all the available N_T spatial channels are used:

$$\sum_{\bar{k}=1}^{K_\ell} p_\ell^{(\breve{k})} = N_T. \tag{8.1}$$

The signal intended for the kth user, collected in the $p_\ell^{(k)} \times 1$ vector $\mathbf{a}_\ell^{(k)}$ is linearly precoded by the $N_T \times p_\ell^{(k)}$ matrix $\mathbf{F}_\ell^{(k)}$. Following the conventional notation (see, e.g., [6]) and referring to fig. 8.1, the signal received by the kth user across its $n_R^{(k)}$ receiving antennas within the ℓth

time slot can be written as the $n_R^{(k)} \times 1$ vector $\mathbf{y}_\ell^{(k)}$

$$\mathbf{y}_\ell^{(k)} = \mathbf{H}_\ell^{(k)} \mathbf{F}_\ell^{(k)} \mathbf{a}_\ell^{(k)} + \mathbf{H}_\ell^{(k)} \sum_{\substack{i \neq k \\ i \in \mathcal{K}(t)}} \mathbf{F}_\ell^{(i)} \mathbf{a}_\ell^{(i)} + \mathbf{n}_\ell^{(k)} \tag{8.2}$$

where $\mathbf{H}_\ell^{(k)}$ is the $n_R^{(k)} \times N_T$ channel matrix of the kth user and $\mathbf{n}_\ell^{(k)}$ is the zero mean additive Gaussian noise with $E[\mathbf{n}_\ell^{(k)} \mathbf{n}_\ell^{(k)H}] = \sigma_n^2 \mathbf{I}_{n_R^{(k)}}$ (we assume spatially white noise).

The received signal $\mathbf{y}_\ell^{(k)}$ lies in a $n_R^{(k)}$-dimensional linear space. However, only $p_\ell^{(k)} \leq n_R^{(k)}$ spatial channels are assigned to the kth user. Therefore, the useful part of the received signal spans a $p_\ell^{(k)}$-dimensional subspace that we refer to as *receiving subspace*. In order to account for this, at the receiver, the $n_R^{(k)} \times 1$ received signal $\mathbf{y}_\ell^{(k)}$ is pre-filtered by the $p_\ell^{(k)} \times n_R^{(k)}$ matrix $\mathbf{G}_\ell^{(k)}$

$$\tilde{\mathbf{y}}_\ell^{(k)} = \mathbf{G}_\ell^{(k)} \mathbf{y}_\ell^{(k)} = \tilde{\mathbf{H}}_\ell^{(k)} \mathbf{F}_\ell^{(k)} \mathbf{a}_\ell^{(k)} + \tilde{\mathbf{H}}_\ell^{(k)} \sum_{\substack{i \neq k \\ i \in \mathcal{K}(t)}} \mathbf{F}_\ell^{(i)} \mathbf{a}_\ell^{(i)} + \tilde{\mathbf{n}}_\ell^{(k)}, \tag{8.3}$$

where we have defined the $p_\ell^{(k)} \times N_T$ equivalent channel $\tilde{\mathbf{H}}_\ell^{(k)} = \mathbf{G}_\ell^{(k)} \mathbf{H}_\ell^{(k)}$ and $\tilde{\mathbf{n}}_\ell^{(k)} = \mathbf{G}_\ell^{(k)} \mathbf{n}_\ell^{(k)}$. In order to simplify the analysis and without limiting the generality of the approach, we assume

$$\mathbf{G}_\ell^{(k)} \mathbf{G}_\ell^{(k)H} = \mathbf{I}_{p_\ell^{(k)}}, \tag{8.4}$$

so that $E[\tilde{\mathbf{n}}_k(t) \tilde{\mathbf{n}}_k(t)^H] = \sigma_n^2 \mathbf{I}_{p_\ell^{(k)}}$. The range space of $\mathbf{G}_\ell^{(k)H}$ corresponds to the receiving subspace for the kth terminal. As a last step, equalization and detection is performed on $\tilde{\mathbf{y}}_\ell^{(k)}$.

In this work, we assume that the channel matrices $\mathbf{H}_\ell^{(k)}$ are known to the transmitter and receivers, e.g., by transmission of training sequences and feedback of the channel state information from the receivers to the base station. An analysis of the effect of imperfect channel state information and feedback delays is proposed in Chapter 6.

In order to simplify the notation, in the following the temporal dependence on ℓ is omitted.

8.2.1 Problem formulation

In principle, we would like to find the subset of users $\mathcal{K}$, the set of precoding matrices $\mathcal{F} = \{\mathbf{F}^{(i)}\}_{i=1}^{\check{K}}$ and pre-filtering matrices $\mathcal{G} = \{\mathbf{G}^{(i)}\}_{i=1}^{\check{K}}$ so that the sum capacity is maximized under a total power constraint (recall also that the constraints (8.1) and (8.4) have to be fullfilled):

$$\{\mathcal{G}, \mathcal{F}\} = \underset{\mathcal{G}, \mathcal{F}}{\operatorname{argmax}} \sum_{i=1}^{\check{K}} C^{(i)}(\mathcal{G}, \mathcal{F}) \tag{8.5a}$$

$$\text{s.t.} \sum_{i=1}^{\check{K}} \operatorname{tr}(\mathbf{F}^{(i)} \mathbf{F}^{(i)H}) \leq P, \tag{8.5b}$$

where $C^{(i)}(\mathcal{G}, \mathcal{F})$ is the link capacity for the ith user [6]

$$C^{(i)}(\mathcal{G}, \mathcal{F}) = \log_2 |\mathbf{I}_{p^{(i)}} + \mathbf{R}^{(i)-1}(\tilde{\mathbf{H}}^{(i)} \mathbf{F}^{(i)H} \mathbf{F}^{(i)H} \tilde{\mathbf{H}}^{(i)H})| \tag{8.6}$$

with

$$\mathbf{R}_n^{(i)} = \sigma_n^2 \mathbf{I}_{p^{(i)}} + \mathbf{H}^{(i)} \sum_{\substack{k \neq i \\ k \in \mathcal{K}}} \mathbf{F}^{(k)} \mathbf{F}^{(k)H} \mathbf{H}^{(i)H}. \tag{8.7}$$

In (8.6)-(8.7) the assumption of Gaussian codebooks with $E[\mathbf{a}^{(i)}\mathbf{a}^{(i)H}] = \mathbf{I}_{p^{(i)}}$ is implied. Moreover, as a result of the optimization problem (8.5) the ith user belongs to the set of active users $\mathcal{K}$ if $p^{(i)} > 0$ or equivalently $(\mathbf{F}^{(i)}, \mathbf{G}^{(i)})$ are not empty matrices.

Solution of the optimization problem (8.5) is not known, even for the case of given sets $\mathcal{K}$ and $\mathcal{G}$. In [8], an algorithm is proposed for obtaining an approximate solution based on the additional constraint of zero inter-user interference and the separate computation of precoding and scheduling as explained in Sec. 8.3. We refer to this algorithm as the Largest Singular Value (LSV) algorithm. The treatment is aimed at setting the results of [8] in the discussed mathematical framework and review the main concepts. A novel approximate solution of (8.5) is then proposed in Sec. 8.4. The algorithm, referred to as Successive Vector Selection (SVS) is still based on the inclusion of the zero inter-user interference constraint but, differently from [8], it performs jointly precoding and scheduling.

8.3   Orthogonal precoding with LSV scheduling

According to the approximate solution of (8.5) proposed in [8], at first the scheduling step is performed. This amounts to select the subset $\mathcal{K}$ and the corresponding K matrices $\mathbf{G}^{(k)}$. Recall that the choice of $\mathbf{G}^{(k)}$ implies the allocation of $p^{(k)}$ spatial channel to the kth user and the corresponding receiving subspace. Then, the design of the precoding matrices $\mathbf{F}^{(k)}$ is carried out with the additional constraint of granting inter-user interference free transmission.

In [8], selection of the subset $\mathcal{K}$ and of the corresponding K matrices $\{\mathbf{G}^{(k)}\}_{k=1}^{K}$ is performed so as to set as active the spatial channels corresponding to the largest singular values (LSV) of matrices $\{\mathbf{H}^{(k)}\}_{k=1}^{\check{K}}$. To elaborate, let $\lambda_j^{(k)}$, $j = 1, \ldots, r^{(k)} = rank(\mathbf{H}^{(k)})$ be the non-zero singular values of channel matrix $\mathbf{H}^{(k)}$ gathered in the diagonal matrix $\mathbf{\Lambda}^{(k)}$ and $(\mathbf{u}_j^{(k)}, \mathbf{v}_j^{(k)})$ the corresponding left and right singular vectors collected by columns in matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ respectively: $\mathbf{H}^{(k)} = \mathbf{U}^{(k)} \mathbf{\Lambda}^{(k)} \mathbf{V}^{(k)}$. The LSV algorithm selects the N_T largest singular values of the set $\{\lambda_j^{(k)} | k = 1, ..., \check{K}, j = 1, \ldots, r^{(k)}\}$ and builds matrices $\mathbf{G}^{(k)H}$ with the corresponding right singular vectors $\mathbf{u}_j^{(k)}$.

This algorithm can equivalently be stated as the solution of the following optimization problem: find the set $\mathcal{G} = \{\mathbf{G}^{(k)}\}_{k=1}^{\check{K}}$ so that (recall also constraints (8.1) and (8.4)):

$$\mathcal{G} = \underset{\mathcal{G}}{\operatorname{argmax}} \sum_{k=1}^{\check{K}} \left\| \mathbf{G}^{(k)} \mathbf{H}^{(k)} \right\|^2, \tag{8.8}$$

Notice that an user belongs to $\mathcal{K}$ if the corresponding number of assigned channel $p^{(k)}$ is not zero, or equivalently $\mathbf{G}^{(k)}$ is not empty.

Given the output of the scheduling algorithm (i.e., the set $\mathcal{K}$ and matrices $\{\mathbf{G}^{(k)}\}_{k=1}^{K}$), the precoding matrices $\mathbf{F}^{(k)}$ are selected according to the orthogonal precoding described in Sec. 6.4.1. In particular, we recall that the precoding matrices are selected as

$$\mathbf{F}^{(k)} = \bar{\mathbf{V}}_0^{(k)} \mathbf{V}^{(k)} \boldsymbol{\Phi}^{(k)}, \tag{8.9}$$

where $\bar{\mathbf{V}}_0^{(k)}$ is the null space of matrix $\bar{\mathbf{H}}^{(k)H}$ with $\bar{\mathbf{H}}^{(k)} = [\tilde{\mathbf{H}}^{(1)H} \cdots \tilde{\mathbf{H}}^{(k-1)H}, \tilde{\mathbf{H}}^{(k+1)H} \cdots \tilde{\mathbf{H}}^{(K)H}]^H$ that guarantees the condition of orthogonality

$$\tilde{\mathbf{H}}^{(i)} \mathbf{F}^{(k)} = \mathbf{0} \ \ if \ \ i \neq k, \tag{8.10}$$

and $\mathbf{V}^{(k)}$ is the subspace spanned by the rows of the single user matrix $\tilde{\mathbf{H}}^{(k)} \bar{\mathbf{V}}_0^{(k)}$ according to (6.13). Finally, $\boldsymbol{\Phi}^{(k)}$ is the diagonal matrix that defines the waterfilling power allocation (6.15).

8.4 Orthogonal precoding with SVS scheduling

The approximate solution of the problem (6.8) proposed by [8] suffers from degraded performance (as it will be shown by numerical results in Sec. 8.6) mainly because the precoding matrices $\mathbf{F}^{(k)}$ and the prefiltering matrices $\mathbf{G}^{(k)}$ (and the associated set $\mathcal{K}$) are optimized separately. Here we propose a joint optimization that approximates problem (6.8) as follows. *i)* The zero interference constraint (8.10) is imposed, thus obtaining a orthogonal precoding as in [8]. As explained in Sec. 6.4, the resulting precoding matrices have the form (8.9). *ii)* The objective function $C^{(i)}(\mathcal{G}, \mathcal{F})$ is approximated by its first term of the Taylor expansion: $C^{(i)}(\mathcal{G}, \mathcal{F}) \simeq 1/\sigma_n^2 \left\| \mathbf{G}^{(i)} \mathbf{H}^{(i)} \mathbf{F}^{(i)} \right\|^2$. The latter approximation is expected to hold at sufficiently low signal-to-noise ratios. The resulting optimization problem reads (recall also constraints (8.1) and (8.4) and define $\bar{\mathcal{V}}_0 = \{\bar{\mathbf{V}}_0^{(i)}\}_{i=1}^{\check{K}}$):

$$\{\mathcal{G}, \bar{\mathcal{V}}_0\} = \underset{\mathcal{G}, \bar{\mathcal{V}}_0}{\operatorname{argmax}} \sum_{i=1}^{\check{K}} N^{(i)}(\mathcal{G}, \bar{\mathcal{V}}_0), \tag{8.11a}$$

$$\text{s.t. } \tilde{\mathbf{H}}^{(i)} \bar{\mathbf{V}}_0^{(j)} = \mathbf{0} \, , i \neq j, \tag{8.11b}$$

where we defined $N^{(i)}(\mathcal{G}, \bar{\mathcal{V}}_0) = \left\| \mathbf{G}^{(i)} \mathbf{H}^{(i)} \bar{\mathbf{V}}_0^{(i)} \right\|^2$. The objective function (8.11a) is amenable to an efficient numerical optimization and will be shown in Sec. 8.6 to yield relevant advantages as compared to the separate optimization proposed in [8]. Notice that in order to simplify the solution of the problem, the remaining term of the precoding matrices (8.9), $\mathbf{V}^{(i)}$ and $\boldsymbol{\Phi}^{(i)}$ are assumed to be computed according to Sec. 6.4.1, thus guaranteeing zero interference among different streams and the enforcement of the total power constraint.

8.4.1 Successive Vector Selection (SVS) channel aware scheduling

Problem (8.11) can be efficiently solved by a greedy approach as detailed in the following. The idea is to select at each step the spatial channel that yields the largest increase of the

objective function (8.11a). Let us denote with the argument (n) the quantities of interest as computed at the nth iteration. At each iteration a spatial channel (out of the N_T available) is allocated to a specific user so that a total number of N_T iterations are needed. We are interested in updating the receiving subspaces $\mathbf{G}^{(i)}$ (initialized as $\mathbf{G}^{(i)}(0)$ equal to an empty matrix) and the transmitting subspaces $\bar{\mathbf{V}}_0^{(i)}$, or equivalently its orthogonal complement $\bar{\mathbf{V}}^{(i)}$ (initialization: $\bar{\mathbf{V}}_0^{(i)}(0) = \mathbf{I}_{N_T}$). Let $\mathbf{u}^{(j)}$ be a possible candidate vector to be included in the receiving subspace $\mathbf{G}^{(j)}[n]$ of user j at the nth iteration ($j = 1, ..., \breve{K}$). As a result of the choice of $\mathbf{u}^{(j)}$ at the nth iteration, the objective function (8.11a) modifies as (dropping the functional dependence on $\mathcal{G}, \bar{\mathcal{V}}_0$ for simplicity of notation)

$$\sum_{i=1}^{\breve{K}} N^{(i)}(n, \mathbf{u}^{(j)}) = \sum_{i=1}^{\breve{K}} N^{(i)}(n) + \sum_{i=1}^{\breve{K}} \Delta N^{(i)}(n, \mathbf{u}^{(j)}). \tag{8.12}$$

Among all the possible vectors $\mathbf{u}^{(j)}$ for all users $j = 1, ..., \breve{K}$, the vector $\mathbf{u}^{(j)}$ is selected so as to maximize the increase of the objective function $\sum_{i=1}^{\breve{K}} \Delta N^{(i)}(n, \mathbf{u}^{(j)})$. In the following, the computation of $\Delta N^{(i)}(n, \mathbf{u}^{(j)})$ is carried out.

To elaborate, we need to define for each user a basis $\mathbf{U}^{(j)}(n)$ that spans the range space of the channel matrix $\mathbf{H}^{(j)}$ that at the nth iteration has not be assigned to any receiving subspace. Formally, it is: $range\{\mathbf{U}^{(j)}(n)\} = range\{\mathbf{U}^{(j)}\} \cap null\{\mathbf{G}^{(j)}(n)\}$. Therefore, the corresponding initialization is $\mathbf{U}^{(j)}(0) = \mathbf{U}^{(j)}$. At the nth iteration we have $\sum_{i=1}^{\breve{K}} p^{(i)}(n) = n$ and the possible candidate vectors to be included in the receiving subspace of the jth user are linear combinations of the columns of $\mathbf{U}^{(j)}(n)$:

$$\mathbf{u}^{(j)} = \mathbf{U}^{(j)}(n)\mathbf{a}^{(j)}, \text{ with } \left\|\mathbf{a}^{(j)}\right\|^2 = 1. \tag{8.13}$$

With the selection of (8.13), the corresponding receiving subspace is updated as $\mathbf{G}^{(j)}(n)^H = [\mathbf{G}^{(j)}(n-1)^H \ \mathbf{u}_j^{(j)}]$ while its transmit subspace remains unchanged, $\bar{\mathbf{V}}_0^{(j)}(n) = \bar{\mathbf{V}}_0^{(j)}(n-1)$, since no new constraint (8.11b) is imposed upon it. It follows that

$$\Delta N^{(i)}(n, \mathbf{u}^{(j)}) = \left\|\mathbf{u}^{(j)H}\mathbf{H}^{(j)}\bar{\mathbf{V}}_0^{(j)}(n)\right\|^2. \tag{8.14}$$

Then, let $\mathbf{v}^{(j)} = \mathbf{H}^{(j)H}\mathbf{u}^{(j)}$ be the vector corresponding to $\mathbf{u}^{(j)}$ on the transmitter side. The choice of $\mathbf{u}^{(j)}$ for user j results in an additional zero-interference constraint for any user $i \neq j$ (see (8.11)), that leads to

$$\bar{\mathbf{V}}^{(i)}(n) = \left[\bar{\mathbf{V}}^{(i)}(n-1) \ \mathbf{w}^{(i)}\right], \tag{8.15}$$

where $\mathbf{w}^{(i)}$ is the projection of $\mathbf{v}^{(j)}$ over $\bar{\mathbf{V}}_0^{(i)}(n-1)$, scaled to unit length

$$\mathbf{w}_i = (\bar{\mathbf{V}}_0^{(i)}(n-1)\bar{\mathbf{V}}_0^{(i)}(n-1)^H\mathbf{v}^{(j)})/||\bar{\mathbf{V}}_0^{(i)}(n-1)\bar{\mathbf{V}}_0^{(i)}(n-1)^H\mathbf{v}^{(j)}||. \tag{8.16}$$

$\bar{\mathbf{V}}_0^{(i)}(n)$ is updated as well, so that $range(\bar{\mathbf{V}}_0^{(i)}(n)) = null(\bar{\mathbf{V}}^{(i)}(n)^H)$. This step can be performed, e.g., by updating the QR decomposition of (8.15) [12]. On the other hand, nothing

changes at the receiver side of the ith user, $\mathbf{G}^{(i)}(n) = \mathbf{G}^{(i)}(n-1)$. It is easy to show that

$$\Delta N^{(i)}(n, \mathbf{u}^{(j)}) = - \left\| \mathbf{G}^{(i)}(n)\mathbf{H}^{(i)}\mathbf{w}^{(i)} \right\|^2 \quad i \neq j. \tag{8.17}$$

To sum up, from (8.14) and (8.17) the increase of objective function (8.11) due to the choice of vector $\mathbf{u}^{(j)}$ at the nth iteration is

$$\sum_{i=1}^{\check{K}} \Delta N^{(i)}(n, \mathbf{u}^{(j)}) = \left\| \mathbf{u}^{(j)H}\mathbf{H}^{(j)}\bar{\mathbf{V}}_0^{(j)}(n) \right\|^2 - \sum_{i \neq j} \left\| \mathbf{G}^{(i)}(n)\mathbf{H}^{(i)}\mathbf{w}^{(i)} \right\|^2. \tag{8.18}$$

The first term in (8.18) accounts for the increased useful power received by user j on the newly assigned spatial channel, whereas the other terms represent the power loss suffered from the other users from not being allowed to transmit over $\mathbf{w}^{(i)}$. Recalling (8.16) and (8.13), function (8.18) can be easily recognized to be a sum of Rayleigh quotients in terms of vector $\mathbf{a}^{(j)}$. While the maximization of a single Rayleigh quotient is analytically feasible since it corresponds to the solution of a generalized eigenvalue problem, maximizing a sum of Rayleigh quotients is much more difficult and costly. Here, we resort to a sub-optimal approach, by restricting $\mathbf{a}^{(j)}$ to be a column of an identity matrix, which translates to restricting our search of the optimal $\mathbf{u}^{(j)}$ to the columns of $\mathbf{U}^{(j)}(n)$. This approach has been proved by simulation to yield performance very close to the optimum solution.

8.5   SVS algorithm with proportional fairness constraints

The algorithms discussed so far aim to maximize the system throughput. If the users are unbalanced, with some of them experiencing strongly attenuated channels (recall Sec. 3.4), it is expected that the algorithms will result in an unfair sharing of system resources that might preclude communication to some users (see also Sec. 8.6). Similarly to the proportional fair criterion [10], a scheduling procedure that achieves over a long term an appropriate balance between sum capacity and fairness among users can be defined by modifying (8.11) as follows (here we explicit for convenience the time dependence):

$$\{\mathcal{G}_\ell, \bar{\mathcal{V}}_\ell^0\} = \underset{\mathcal{G}_\ell, \bar{\mathcal{V}}_\ell^0}{\operatorname{argmax}} \sum_{i=1}^{\check{K}} \log(E[N^{(i)}(\mathcal{G}, \bar{\mathcal{V}}^0)]), \tag{8.19a}$$

$$\text{s.t. } \tilde{\mathbf{H}}_\ell^{(i)} \bar{\mathbf{V}}_{0,\ell}^{(j)} = \mathbf{0} , \, i \neq j. \tag{8.19b}$$

where $E[\cdot]$ refers to the long term average over time. Following the analysis of [9] and the considerations in Sec. 8.4 , it can be shown that a procedure that (approximately) converges to the solution of (8.19) can be obtained by implementing the SVS algorithm on *normalized* channel matrices

$$\bar{\mathbf{H}}_\ell^{(i)} = \frac{\bar{\mathbf{H}}_\ell^{(i)}}{\hat{\alpha}_\ell^{(i)}} \tag{8.20}$$

with

$$\hat{\alpha}_\ell^{(i)} = \left(1 - \frac{1}{\ell_c}\right)\hat{\alpha}_{\ell-1}^{(i)} + \frac{1}{\ell_c}N^{(i)}(\mathcal{G}_{\ell-1}, \bar{\mathcal{V}}_{\ell-1}^0), \qquad (8.21)$$

where parameter ℓ_c rules the memory of the algorithm. Parameter $\hat{\alpha}_\ell^{(i)}$ measures the channel power that each user has been allowed to use within a window of ℓ_c time slots. The rationale of the algorithm is that if a given user has been ignored by the scheduling procedure in the considered time window, the matrix scaling (8.20) will force the SVS algorithm to allocate resources to it.

The criterion (8.19) is a fairness constraints on the channel norms: its implication on the channel rates is not obvious and will be investigated in the next Section by numerical simulations.

8.6   Numerical simulations

The performance of the proposed SVS algorithm is compared with the LSV algorithm [8] by Monte Carlo simulations. We consider $\breve{K} = 4$ users, where each user has the same number of receiving antennas $n_R^{(i)} = 2$ while the base station is equipped with $N_T = 4$ antennas. Where not stated otherwise, the channels are assumed to be subject to identically distributed Raleigh fading, $\text{vec}(\mathbf{H}^{(i)}) \sim \mathcal{CN}(0, \mathbf{I}_{N_T n_{R_i}^{(i)}})$. As reference performance, a random user selection algorithm is considered that chooses randomly a set $\mathcal{K}$ of users such that (8.1) is satisfied. On this subset, orthogonal precoding is applied as detailed in Sec. 6.4.1. Moreover, the performance of a $N_T \times N_T$ single user MIMO link is evaluated in order to set a reference level for the sum capacity of the multiuser system.

The ergodic sum capacity is plotted versus the signal to noise ratio in fig. 8.2. The proposed SVS algorithm yields a gain of about $4dB$ as compared to the LSV algorithm, whose performance are, in this case, similar to random users selection. As explained in Sec. 8.4, the advantage of SVS is due to the joint computation of the transmitting and receiving subspaces.

Fig. 8.3 shows sum capacity versus outage probability for $SNR = 10dB$. It can be seen that the slope of the outage probability for SVS is comparable to that of a single user channel, proving the ability of the SVS algorithm to appropriately exploit the diversity of the broadcast channel.

Let us now consider unbalanced users in order to validate the performance of the SVS algorithm with fairness constraints. To be specific, as explained in Sec. 3.4, the channels are assumed to be selected so that $\text{vec}(\mathbf{H}^{(i)}) \sim \mathcal{CN}(0, \alpha^{(i)}\mathbf{I}_{N_T n_{R_i}})$, where $\alpha^{(1)} = 0dB$, $\alpha^{(2)} = -5dB$, $\alpha^{(3)} = -10dB$ and $\alpha^{(4)} = -20dB$. The performance of the SVS algorithm is evaluated with and without fairness constraints ($\ell_c = 20$). The results are summarized in fig. 8.4 in terms of ergodic sum capacity and individual ergodic capacity versus time ℓ. The total throughput loss increases as the fairness constraints are imposed by the scaling algorithm discussed in Sec. 8.5 and converges to approximately $2bit/s/Hz$. This decrease of the sum
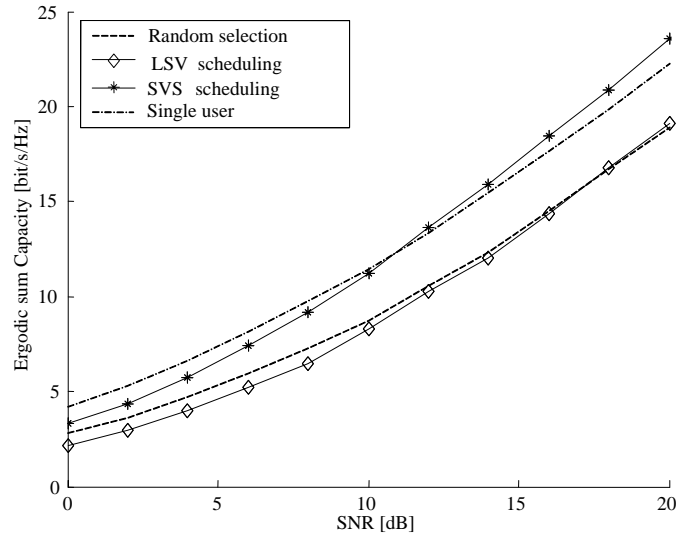
Figure 8.2: Ergodic sum capacity versus SNR ($\breve{K} = 4$, $N_T = 4$, $n_R^{(i)} = 2$).
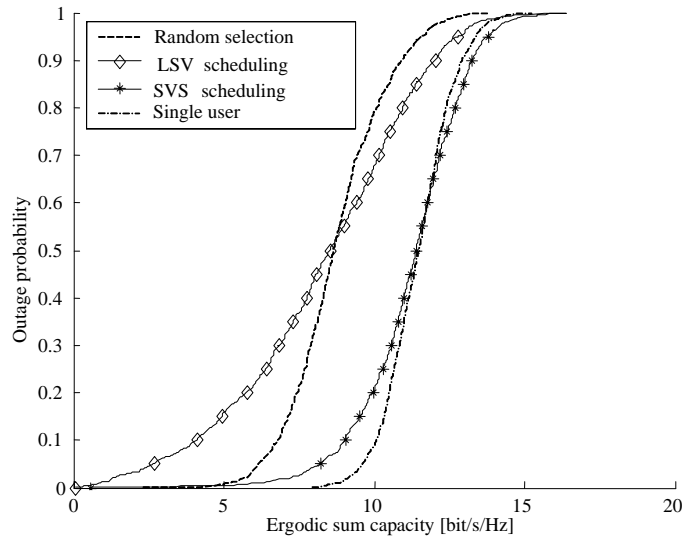

Figure 8.3: Sum capacity versus outage probability ($\breve{K} = 4$, $N_T = 4$, $n_R^{(i)} = 2$, $SNR = 10dB$).
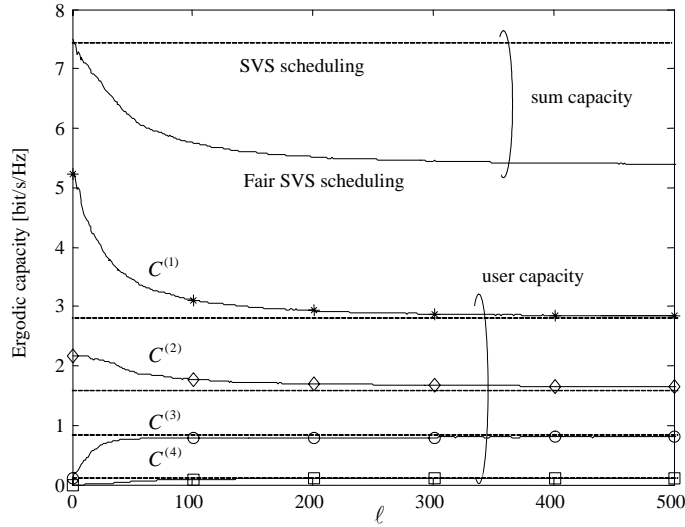
Figure 8.4: Ergodic sum capacity and individual capacities versus ℓ for the SVS algorithm with and without fairness constraints (unbalanced users, $\breve{K} = 4$, $N_T = 4$, $n_R^{(i)} = 2$, $SNR = 10dB$).

capacity translates in a more fair sharing of resources as proved by the individual channel capacities. Notice that the channel capacities for $\ell = 0$ correspond to the performance of the SVS algorithm with no fairness constraints. Even though the fairness constraint (8.19) was imposed on the channel norms, rather than on the individual rates as in [10], the simulation results indicate that the proportional fair criterion is very closely followed by the channel capacities as well. In fact, the user capacities approximately converge to the dashed lines in fig. 8.4 that denote the individual capacities as obtained by sharing the long term sum capacity according to the proportional fair criterion (i.e., in proportion to the single user capacities).

8.7 Conclusion

The problem of channel aware scheduling for broadcast MIMO channels with orthogonal linear precoding and linear interfaces at the receivers has been investigated. Starting from the work in [8], that introduced a scheduling algorithm based on the separate computation of precoding and equalization matrices (LSV algorithm), here we proposed a novel technique (SVS) that performs joint optimization of precoding and scheduling. This solution has been shown by simulation to be superior to known techniques in terms of sum capacity.

However, the sum-capacity is not the appropriate performance measure for practical systems where fairness constraints have to be guaranteed among the users. Therefore, a modification of the SVS algorithm has been proposed that is based on the enforcement of proportional fairness constraints. Simulation results have validated the modified SVS algorithm by showing

that it retains the desirable long term fairness properties.

Bibliography

[1] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," *Proc. IEEE Int. Computer Conf.*, pp. 331-335, June 1995.

[2] D. N. C. Tse, "Optimal power allocation over parallel Gaussian channels," *Proc. IEEE ISIT*, p. 27, 1997.

[3] H. Viswanathan and K. Kumaran, "Rate scheduling in multiple antenna downlink wireless systems," *Proc. Allerton Conference*, 2001.

[4] I. Koutsopoulos, T. Ren and L. Tassiulas, "The impact of space division multiplexing on resource allocation: a unified view," *Proc. IEEE Infocom,* pp. 533-543, 2003.

[5] R. W. Heath Jr., M. Airy, A. J. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," *Proc. Asilomar Conf. Signals, Systems and Computers*, pp. 1194-1199, Nov. 2001.

[6] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684-702, June 2003.

[7] O. Shin and K. Bok, "Antenna-assisted round robin scheduling for MIMO cellular systems," *IEEE Comm. Letters*, vol. 7, no. 3, pp. 109-111, March 2003.

[8] Q. H. Spencer, A. Lee Swindlehurst and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels", *IEEE Trans. Signal Processing*, vol. 52 , no. 2 , pp. 461-471, Feb. 2004.

[9] H. J. Kushner and P. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions", *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1250-1259, July 2004.

[10] P. Viswanath, D. N. C. Tse, R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277-1294, June 2002.

[11] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," submitted to *IEEE Trans. Signal Processing* [also available on http://mars.bell-labs.com/].

[12] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins, 1996, 3rd edition.

Chapter 9

Concluding remarks

THIS thesis has dealt with the analysis of the perfomance of single and multiuser MIMO systems over time-varying frequency-selective channels. Our investigation considered under the same framework both the time-domain and multicarrier transmission. In either case, the transmission is organized in blocks that contains both training (or pilot) symbols, that are used at the receiver side for CSI acquisition, and data symbols. The main contribution of this work has been the study of the performance of the CSI acquisition phase and of linear precoding and equalization under a realistic multipath channel model. In particular, the analysis has capitalized on the fundamental property of the time-varying MIMO channel. The time-variability of the propagation is due to two classes of parameters that have different varying rates. The first set encompasses the long-term features of the multipath, such as angles, delays and power-delay profile, whereas the second accounts for the fast varying fading process. Within the block-fading assumption, this implies that while the first set of parameters can be considered as approximately constant over multiple blocks, the latter varies from block to block.

Based on this property, the algebraic structure of the channel vector has been studied, revealing that long-term and fast varying multipath features can be analytically decoupled. In particular, it was shown that long-term parameters only determine the space-time channel correlation, that is generally rank-deficient. Therefore, this matrix is defined by the space-time channel modes, that are identified by its principal eigenvectors. On the other hand, the fading process translates into the time-varying amplitudes to be assigned on a block-by-block basis on each space-time mode. The implications of the results discussed above have been investigated for the CSI acquisition phase in the first part of this dissertation and for linear precoding and equalization in the second part.

In the first part of the thesis, the analysis of CSI acquisition is tackled at first from a theoretical point of view, by determining a lower bound on the performance of any unbiased channel estimator, through the computation of the hybrid CRB. This study not only allowed to assess the impact of channel and system characteristics on the performance of channel estimation, but

also discloses the properties of (asymptotically) optimal channel estimators. According to this result, the optimal strategy for channel estimation prescribes a separate computation of long-term and fast-varying multipath parameters. Linear estimators that are designed based on this idea have been then proposed. To be specific, these estimators perform a direct estimation of space-time channel modes through subspace tracking while computing the fading amplitudes by least squares techniques. Thorough performance analysis and simulation demonstrated that the proposed techniques are able to attain he theoretical bound under appropriate conditions.

The second part of the dissertation focused on the design of linear precoding and equalization based on different assumptions about the CSI available at the transmitter. At first, the impact of imperfect CSI at both ends of the link has been assessed from an information theoretic standpoint by evaluating a lower bound on the information rate. This study has been performed for both a single user MIMO system over a frequency-selective channel and the downlink of a multiuser MIMO system over a frequency-selective channel with orthogonal linear precoding. An information theoretic analysis of the downlink for more general linear or non-linear precoding algorithms is an open issue. Subsequently, design of linear/non-linear precoding and equalization based on long-term CSI at the transmitter for a single user MIMO system over a frequency-flat channel has been proposed. Finally, joint design of linear precoding, equalization and scheduling for the downlink of a multiuser MIMO system over a frequency-flat channel has been considered. The designed aimed at either maximizing the sum rate or ensuring some fairness constraints among the users. Therein, it was assumed that instantaneous CSI is available at both receiver and transmitter.

Throughout the thesis, the fraction of symbols dedicated in each block to the transmission of training has been held fixed. The benefits that can be gained from he exploitation of this additional degree of freedom by using long-term CSI at the transmitter are investigated in the Appendix.

Appendix: adaptive pilot pattern based on long term CSI at the transmitter for OFDM systems

10.1 Introduction

$\mathbf{I}$N order to achieve high spectral efficiency and reliability, wireless data transmission over fading channels requires techniques that are able to adaptively adjust to the channel state. Based on the CSI available at the transmitter, transmission parameters such as transmit power, constellation size and coding scheme can be adaptively chosen at the physical layer in order to satisfy some quality of service (QoS) criterion [1].

So far we have considered the ratio between the number of data (L_D) and pilot (L_P) subcarriers to be held fixed over all the transmitter blocks. In this Chapter, we propose to extend the set of parameters to be adapted based on CSI to the pilot (or training) arrangement, i.e., to the number of pilot symbols to be transmitted on each block. As with the other adaptive transmission techniques, adaptive pilot placement promises to increase the spectral efficiency of the system. This study will be performed on a single user SISO ($N_R = N_T = 1$) OFDM system since the extension to more complex system is conceptually straightforward.

Adaptive pilot placement can be described as follows. QoS requirements from higher layers determine a maximum channel estimation error that the system can tolerate. Based on the prediction of channel estimation error at the receiver, the transmitter can then allocate the pilot pattern (i.e., over a given number of OFDM symbols, say N_B) with the goal of guaranteeing the reliability of the channel estimate while minimizing the number of pilot subcarriers.

Prediction of the channel estimation error at the transmitter is herein performed by considering the performance of a Kalman channel estimator at the receiver. According to the practical considerations in Chapter 7, the transmitter is assumed to know the long term features of the channel, i.e., the second order statistics of the channel (long term CSI) and the average SNR.
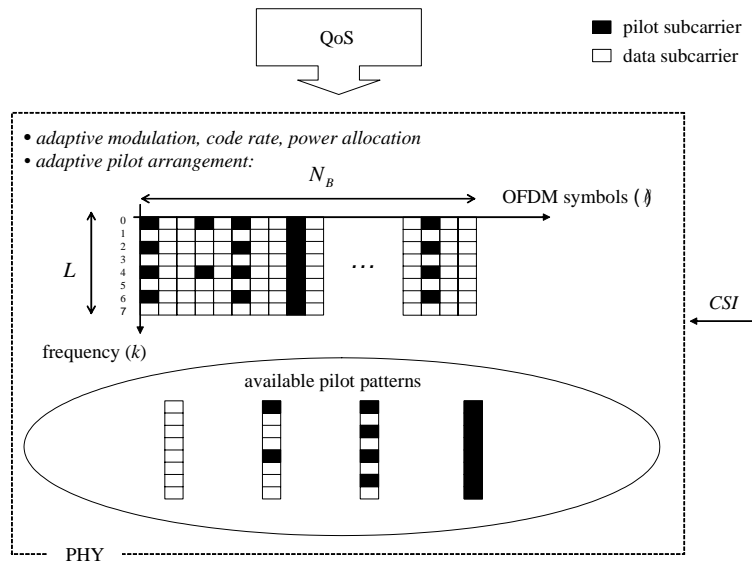
Figure 10.1: Description of an OFDM system with adaptive pilot placement,

This CSI can be either acquired directly by the transmitter in a TDD link or fedback by the receiver in a FDD link. Notice that in the latter case, the feedback link is only required to support a low rate since the considered CSI can be assumed to be slowly varying.

10.2 System description

10.2.1 Adaptive pilot pattern: motivation and fundamentals

The system is illustrated in fig. 10.1. The physical layer of the link employs OFDM modulation with training-based channel estimation. Higher layers set some constraint on QoS, such as frame/symbol error probability. The physical layer then adapts transmission parameters (modulation, code rate and power) in order to satisfy the QoS requirements. Furthermore, it computes a maximum SNR degradation η due to channel estimation (or equivalently a maximum channel estimation error) that leads to negligible system performance loss. A certain SNR degradation can be guaranteed by appropriately placing pilot subcarriers in each OFDM symbol as it will be explained in the following. The adaptive choice of the pilot pattern is made every N_B OFDM symbols, where N_B can be for instance the duration of the time-slot assigned to a given user in a TDMA system [5]. According to Chapter 3, the total number of subcarriers is L. Notice that L could be a fraction of the subcarriers available in the bandwidth of an OFDMA system [6]. Moreover, for simplicity, we assume that no guardband is allocated within the L available subcarrier ($L_G = 0$).

To ease the implementation of the OFDM modulator, the number of available subcarriers L is set to be a power of 2, $L = 2^l$. The subcarriers are indexed as $k = 0, 1, ..., L - 1$

while the OFDM symbols are denoted as $\ell = 0, 1, ..., N_B - 1$ (see fig. 10.1). Equi-spaced pilot subcarriers in the frequency domain are known to minimize the channel estimation error [7]; accordingly, the number of pilot arrangements that constitute possible choices for the transmitter in each OFDM symbol are $l + 1$ and precisely (fig. 10.1):

1. no pilot subcarriers ($L_P = 0$): the OFDM symbol contains only data subcarriers;

2. $L_P = 2^m$ equispaced pilot subcarriers (with $m = 1, ..., l$): the pilot subcarriers occupy the frequency bins $i \cdot 2^{l-m}$ with $i = 0, 1, ..., L_P - 1$.

10.2.2 Review of signal and channel model

From Chapter 3, to which we refer for details, we recall that the signal received on the kth subcarrier in the ℓth OFDM symbol can be written as (see (3.15))

$$y_\ell[k] = x_\ell[k] f_\ell[k] + n_\ell[k], \tag{10.1}$$

where $f_\ell[k]$ is the channel gain and $n_\ell[k]$ the additive Gaussian noise. By stacking the signal received over the used bandwidth ($\mathbf{y}_\ell = [y_\ell[0] \cdots y_\ell[L-1]]^T$), the measurement within the ℓth OFDM symbol can be written as

$$\mathbf{y}_\ell = \mathbf{X}_\ell \mathbf{f}_\ell + \mathbf{n}_\ell \tag{10.2}$$

where $\mathbf{X}_\ell$ is the $L \times L$ diagonal matrix (3.24), $\mathbf{X}_\ell = diag\{[x_\ell[0] \cdots x_\ell[L-1]]\}$, and $\mathbf{f}_\ell = [f_\ell[0] \cdots f_\ell[L-1]]^T$ is the $L \times 1$ channel vector in the frequency domain, related to the channel in the time domain through the $L \times W$ DFT matrix $\mathbf{\Theta}$

$$\mathbf{f}_\ell = \mathbf{\Theta} \cdot \mathbf{h}_\ell. \tag{10.3}$$

In order to simplify the analysis, we consider a channel characterized by sample spaced delays ($\mathbf{G}(\boldsymbol{\tau}) = \mathbf{I}_W$) so that the channel model described in Chapter 3 (for both beamforming and diversity scenarios) reads

$$\mathbf{h}_\ell = \mathbf{\Omega}^{1/2} \boldsymbol{\beta}_\ell. \tag{10.4}$$

Moreover, the Doppler spectrum of all the paths is assumed to be well approximated over the N_B OFDM symbols by an autoregressive model of order 1, i.e., $E[\boldsymbol{\beta}_\ell \boldsymbol{\beta}_{\ell-m}^H] = \rho^m \mathbf{I}_W$, where ρ can be written in terms of the Clarke's model as a function of the Doppler shift f_D as $\rho = J_o(2\pi f_D T_S)$ [11]. Notice that the second order statistics of the channel

$$E[\mathbf{h}_\ell \mathbf{h}_{\ell-m}^H] = \mathbf{\Omega} \rho^t \tag{10.5}$$

are assumed to be constant over the temporal horizon of interest (quasi static model of the long term channel features over N_B OFDM symbols as in Chapter 4).

From (10.3) and (10.4), the channel correlation in the frequency domain is $\mathbf{R}_F = E[\mathbf{f}_\ell \mathbf{f}_\ell^H] = \mathbf{\Theta}\mathbf{\Omega}\mathbf{\Theta}^H$ or equivalently, for the entries, $[\mathbf{R}_F]_{ij} = \sum_{n=0}^{W-1}[\mathbf{\Theta}]_{in}[\mathbf{\Theta}]_{jn}^*\Omega_n$, which implies that unless $L = W$ and the power delay profile is uniform, the channel is correlated over frequency.

10.2.3   Effective (average) signal to noise ratio SNR_{eff}

The (average) SNR over the kth subcarrier and the ℓth OFDM symbol in absence of channel estimation error can be written as (3.14). We recall that the SNR (3.14) would depend on the frequency bin k and on the OFDM symbol ℓ if adaptive power allocation is employed on a subcarrier (or group of subcarrier) basis and/or on a symbol by symbol basis (i.e., if $E[|x_\ell[k]|^2]$ is a function of ℓ and/or k). Here, we simplify the presentation by setting $E[|x_\ell[k]|^2] = LP$ (recall the discussion in Sec. 3.6) equal for all subcarrier and blocks (within the interval of interest). The extension to the most general case of adaptive power allocation is not covered here as it is straightforward and requires minor modifications.

The effect of channel estimation errors can be taken into account by defining an "effective" (average) SNR as follows. By introducing the channel estimate $\hat{f}_\ell[k]$, the received signal (10.1) can be restated as

$$y_\ell[k] = \hat{f}_\ell[k]x_\ell[k] + (f_\ell[k] - \hat{f}_\ell[k])x_\ell[k] + n_\ell[k] \tag{10.6}$$

The term related to the channel estimation error, $(f_\ell[k] - \hat{f}_\ell[k])x_\ell[k]$, can be then modelled as explained in Chapter 6 as an additive Gaussian noise with zero mean and power $E[|f_\ell[k] - \hat{f}_\ell[k]|^2]LP$, assuming that the error and the transmitted signal are independent [12]. For equi-spaced equi-powered pilot subcarriers, the channel estimate variance $E[|f_\ell[k] - \hat{f}_\ell[k]|^2] = MSE_\ell$ is independent on k so that the effective SNR can be defined as

$$SNR_\ell^{eff} = \frac{P}{(\sigma_n^2 + MSE_\ell \cdot LP)} = \frac{SNR}{1 + \overline{MSE}_\ell \cdot SNR}, \tag{10.7}$$

where $\overline{MSE}_\ell$ is the normalized (with respect to the channel norm $E[|f_\ell[k]|^2] = 1/L$) channel estimation error: $\overline{MSE}_\ell = MSE_\ell/(1/L)$.

10.2.4   Adaptive pilot pattern: problem formulation

As previously stated, the physical layer at the transmitter computes a maximum SNR loss η due to channel estimation. Equivalently, the physical layer sets a minimum effective SNR, denoted as $\overline{SNR}^{eff}$, that it is required to guarantee over the N_B OFDM symbols. Feasibility calls for $\overline{SNR}^{eff} < SNR$, or in other words $\eta = SNR/\overline{SNR}^{eff} > 1$.

The goal of the physical layer is to place pilot subcarriers over the N_B OFDM symbols in order to guarantee that

$$SNR_\ell^{eff} \geq \overline{SNR}^{eff}, \tag{10.8}$$

or in terms of channel estimation error

$$\overline{MSE}_\ell \leq \frac{1}{\overline{SNR}^{eff}} - \frac{1}{SNR}. \tag{10.9}$$

In the next Section, we will be able to write $\overline{MSE}_\ell$ as a function of the pilot pattern for a Kalman channel estimator (that we know from Chapter 4 to be the optimal channel estimation

strategy if we assume to know the multipath delays and the second order statistics of the channel). The physical layer at the transmitter side tries to *guarantee the minimum effective SNR* $(\overline{SNR}^{eff}$, *see (10.8) or (10.9)) for* $\ell = 0, 1, ..., N_B - 1$ *by using the minimum number of pilot subcarriers over the* N_B *OFDM symbols.*

Remark 9 *Decision directed or iterative techniques can improve the channel estimation accuracy at the cost of increased complexity. When these techniques are implemented, the requirement on* $\overline{MSE}_\ell$ *can be alleviated accordingly.*

Remark 10 *Similarly to all the adaptive transmission techniques, adaptive pilot pattern requires the transmitter to inform the receiver about the selected pilot arrangement (over the* $l + 1$ *available) for each* ℓ.

Remark 11 *If the adaptive modulation and code rate algorithm has selected different modes for different subcarriers or OFDM symbols, the maximum allowable SNR loss could be a function of* k *and* ℓ *itself. In this case, a possible choice in our framework is to define* η *with respect to the most demanding requirement, that is* $\eta = \min_{\ell,k} \eta(\ell, k)$.

10.3 Kalman channel estimation

From (10.4) and (10.5), the channel model and the received signal can be written in a state form equation

$$\boldsymbol{\beta}_\ell \;\; = \;\; \rho\boldsymbol{\beta}_{\ell-1} + \mathbf{v}_\ell \tag{10.10a}$$

$$\mathbf{y}_\ell \;\; = \;\; \mathbf{C}_\ell\boldsymbol{\beta}_\ell + \mathbf{w}_\ell, \tag{10.10b}$$

where $\mathbf{C}_\ell = \mathbf{X}_\ell\boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}$ and $\mathbf{v}_\ell$ is zero mean circularly symmetric Gaussian with $E[\mathbf{v}_\ell\mathbf{v}_{\ell-m}^H] = \sqrt{(1-\rho^2)}\mathbf{I}_W\delta_m$. Assuming that the second order statistics of the channel (long term CSI) ρ (or equivalently the Doppler shift f_D) and $\boldsymbol{\Omega}$ are known at the receiver (e.g., from long term measurements of the channel), the $L \times 1$ vector β_ℓ can be tracked by a Kalman filter [10]. Since we focus on training-based channel estimation, that is we do not consider the data symbols as useful information for channel estimation, the vector $\mathbf{x}_\ell$ will be thereafter redefined as $\tilde{\mathbf{x}}_\ell$, where:

$$[\tilde{\mathbf{x}}_\ell]_k \;\; = \;\; 0 \;\; \text{if the } k\text{th subcarrier contains a data symbol}$$

$$[\tilde{\mathbf{x}}_\ell]_k \;\; \neq \;\; 0 \;\; \text{if the } k\text{th subcarrier contains a pilot symbol.}$$

Matrices $\tilde{\mathbf{X}}_\ell$ and $\tilde{\mathbf{C}}_\ell = \tilde{\mathbf{X}}_\ell\boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}$ are defined accordingly.

The a posteriori error correlation matrix $\mathbf{R}_{\beta,\ell} = E[(\boldsymbol{\beta}_\ell - \hat{\boldsymbol{\beta}}_\ell)(\boldsymbol{\beta}_\ell - \hat{\boldsymbol{\beta}}_\ell)^H]$ of the Kalman filter can be written as a function of the pilot pattern $\tilde{\mathbf{x}}_\ell$ following the standard theory [10]. Since the estimate of the frequency domain channel vector is then obtained as

$$\hat{\mathbf{f}}_\ell = \boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}\hat{\boldsymbol{\beta}}_\ell, \tag{10.12}$$

the error correlation matrix for the channel estimate in the frequency domain follows as

$$\mathbf{R}_{e,\ell} = E[(\mathbf{f}_\ell - \hat{\mathbf{f}}_\ell)(\mathbf{f}_\ell - \hat{\mathbf{f}}_\ell)^H] = \boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}\mathbf{R}_{g,\ell}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Theta}^H. \tag{10.13}$$

Notice that $MSE_\ell = [\mathbf{R}_{e,\ell}]_{ii}$. It is easy to show that $\mathbf{R}_{e,\ell}$ can be computed recursively as a function of the pilot pattern $\tilde{\mathbf{x}}_\ell$ as (see Appendix-A)

$$\mathbf{R}_{e,\ell} = \mathbf{A}_\ell - \rho^2 \mathbf{A}_\ell \tilde{\mathbf{X}}_\ell^H (\tilde{\mathbf{X}}_\ell \mathbf{A}_\ell \tilde{\mathbf{X}}_\ell^H + \sigma_n^2 \mathbf{I})^{-1} \tilde{\mathbf{X}}_\ell \mathbf{A}_\ell \tag{10.14a}$$

$$\mathbf{A}_\ell = \rho^2 \mathbf{R}_{e,\ell-1} + (1-\rho^2)\mathbf{R}_F. \tag{10.14b}$$

Equations (10.14a) and (10.14b) can be implemented at the transmitter as long as the long term CSI (i.e., ρ and $\boldsymbol{\Omega}$) is known at the transmitter either by direct measurement (TDD link) or by (low rate) feedback by the receiver (FDD link). The initial estimation ($\ell = 0$) of the channel is assumed to be obtained by a traditional UML estimator

$$\boldsymbol{\beta}_0 = (\tilde{\mathbf{C}}_0^H \tilde{\mathbf{C}}_0)^{-1} \tilde{\mathbf{C}}_0^H \mathbf{y}_0, \tag{10.15}$$

so that $\mathbf{R}_{g,0} = \sigma_n^2 (\tilde{\mathbf{C}}_0^H \tilde{\mathbf{C}}_0)^{-1}$ and

$$\mathbf{R}_{e,0} = \sigma_n^2 \boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}(\tilde{\mathbf{C}}_0^H \tilde{\mathbf{C}}_0)^{-1}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Theta}^H = \sigma_n^2 \boldsymbol{\Theta}(\boldsymbol{\Theta}^H\boldsymbol{\Omega}_{x,0}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^H, \tag{10.16}$$

where the pilot pattern power profile $\boldsymbol{\Omega}_{x,\ell} = \tilde{\mathbf{X}}_\ell^H \tilde{\mathbf{X}}_\ell$ has been defined. Notice that in order to make the initial UML estimate (10.15) feasible, the initial training pattern $\tilde{\mathbf{x}}_0$ has to contain at least $L_P \geq W$ pilot subcarriers.

Restating the problem illustrated in Sec. 10.2.4, the transmitter has to guarantee that $MSE_\ell = [\mathbf{R}_{e,\ell}]_{ii}$ satisfies (10.9) for $\ell = 0, 1, ..., N_B - 1$ while minimizing the number of pilot subcarriers.

10.3.1 Training sequence design

From (10.16) it is clear that the initial channel estimation error only depends on the pilot pattern power profile $\boldsymbol{\Omega}_{x,\ell}$: the training sequence $\tilde{\mathbf{x}}_\ell$ can be drawn from any constellation as long as the power profile is the same. The same applies to the channel estimation error for any ℓ as it is shown in Appendix-B. Moreover, here we are interested as for the derivation of (10.7) in a uniform channel estimation error over the subcarriers so that we opt for a uniform power allocation over the pilot subcarriers.

10.4 "Greedy" pilot pattern

The choice of the optimum pilot pattern $\tilde{\mathbf{x}}_\ell$ $\ell = 0, 1, ..., N_B - 1$ would require to explore a large number of possible solutions, on the order of $(l+1)^{N_B}$. A suboptimum solution that proved to perform satisfactorily is the "greedy" algorithm described below:

for $\ell = 0$: select the minimum $m = 1, ..., l$ such that: *a*) $L_P = 2^m \geq W$ and *b*) the constraint (10.9) is satisfied;

for $\ell > 0$: select the minimum $L_P = 0$ or $L_P = 2^m$ ($m = 1, ..., l$) such that (10.9) is satisfied.

This algorithm is referred to as greedy since it selects for each OFDM symbol the best immediate solution, i.e., the minimum number of pilot subcarriers that guarantees (10.9).

10.5 Simulation results

Let us consider an OFDM link with $L = 64$ subcarriers and a block of $N_B = 10$ OFDM symbols. Moreover, $f_D T_S = 0.1$ ($\rho = 0.9975$), $W = 16$ and the power delay profile is uniform, $\mathbf{\Omega} = 1/W \mathbf{I}_W$. We start with a simple example in order to clarify the system under study and to show the effectiveness of the greedy solution. The SNR is known to be $SNR = 20dB$ and the physical layer sets as a requirement a SNR loss $\eta = 5dB$ or equivalently $\overline{SNR}^{eff} = 15dB$. It follows from (10.9) that $\overline{MSE}_\ell \leq 0.0216$. The dashed line in fig. 10.3 represents this upper bound on the normalized error $\overline{MSE}_\ell$. A first choice for the pilot pattern could be to place in each OFDM symbol the same number of pilot L_P that allows the initial least squares estimate to guarantee $\overline{MSE}_0 \leq 0.0216$. Fig. 10.2 shows this pattern in the first row ("fixed pattern") with $L_P = W = 16$. In this case, the total number of pilot subcarriers employed over $N_B = 10$ OFDM symbols turns out to be $16 \times 10 = 160$. If we used the initial least squares estimate for the entire block of N_B symbols ("only initial training"), we would end up with only 16 pilot subcarriers employed but we would not be able to satisfy (10.9) for $\ell > 4$. A practical and widely used choice is to re-train the system whenever we need it ("periodic re-training"). In this case, the use of $16 \times 4 = 64$ pilot subcarriers is enough to satisfy (10.9) for every ℓ. As shown in fig. 10.2 and 10.3, the greedy pilot pattern not only satisfies the SNR requirement but also allows to reduce the number of pilot subcarriers to 52, thus improving the spectral efficiency of the system. Notice that letting N_B grow, we could notice that the greedy pilot pattern reaches a steady state periodic pilot arrangement with period of 6 OFDM symbols such that L_P over each period varies as $\{0, 0, 2, 8, 2, 16\}$. By contrast, the "periodic re-training" algorithm (optimized to guarantee the same SNR loss $\eta = 5dB$) has a period of 3 OFDM symbols with L_P varying as $\{0, 0, 16\}$.

We now consider the performance of the greedy adaptive pilot pattern compared to the traditional periodic re-training approach for varying channel correlation (fig. 10.4) and SNR (fig. 10.5) for two different SNR loss $\eta = 3dB$ (dashed lines) and $\eta = 5dB$ (solid lines). As in the previous example, both schemes are implemented so as to satisfy (10.8) for $\ell = 0, 1, ..., 9$. The performance is shown in terms of the fraction of total subcarriers ($N_B \times L = 10 \times 64$) used as pilot subcarriers. As expected, increasing the channel variability (i.e., $f_D T_S$) decreasing the SNR loss η or increasing SNR cause the number of needed pilot subcarriers to increase. Moreover, the greedy algorithm (circles) uniformly outperforms the periodic training scheme
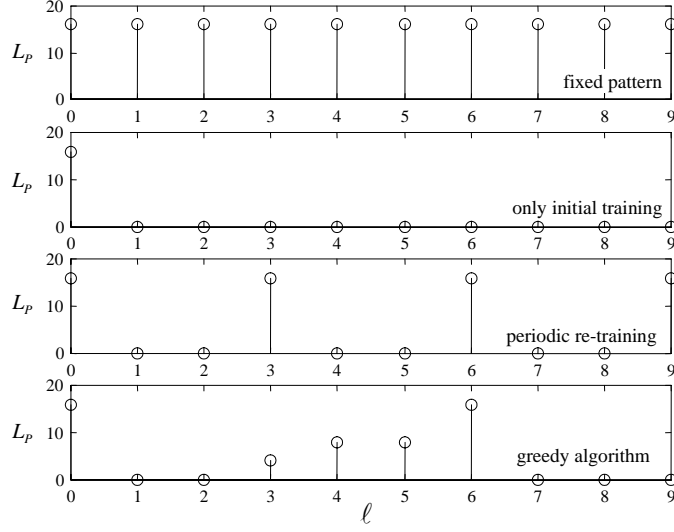
Figure 10.2: Number of pilot subcarriers L_P versus ℓ for different pilot arrangements.

(triangles) leading to a more efficient use of the available bandwidth.

10.6   Conclusion

In this Chapter, the extension of link adaptability to the pilot pattern has been proposed. A suboptimum solution based on the greedy principle has been considered and its effectiveness evaluated through simulation with respect to known (non-adaptive) pilot arrangements. Our findings show that adaptive pilot arrangement can greatly improve the spectral efficiency of the system when compared to conventional strategies.

10.7   Appendix-A: proof of (10.14a)-(10.14b)

For Kalman filtering [10], the correlation matrix $\mathbf{R}_{g,\ell}$ can be computed recursively as

$$\mathbf{R}_{g,\ell} = \breve{\mathbf{R}}_{g,\ell} - \rho \mathbf{G}_\ell \tilde{\mathbf{C}}_\ell \breve{\mathbf{R}}_{g,\ell}, \tag{10.17}$$

where

$$\breve{\mathbf{R}}_{g,\ell} = \rho^2 \mathbf{R}_{g,\ell-1} + (1-\rho^2)\mathbf{I} \tag{10.18}$$

is the correlation matrix of the one-step prediction error and

$$\mathbf{G}_\ell = \rho \breve{\mathbf{R}}_{g,\ell} \tilde{\mathbf{C}}_\ell^H (\tilde{\mathbf{C}}_\ell \breve{\mathbf{R}}_{g,\ell} \tilde{\mathbf{C}}_\ell^H + \sigma_n^2 \mathbf{I})^{-1} \tag{10.19}$$

is the Kalman gain. Recalling (10.13) and defining

$$\mathbf{A}_\ell = \mathbf{\Theta} \mathbf{\Omega}^{1/2} \breve{\mathbf{R}}_{g,\ell} \mathbf{\Omega}^{1/2} \mathbf{\Theta}^H = \rho^2 \mathbf{R}_{e,\ell-1} + (1-\rho^2)\mathbf{R}_F, \tag{10.20}$$
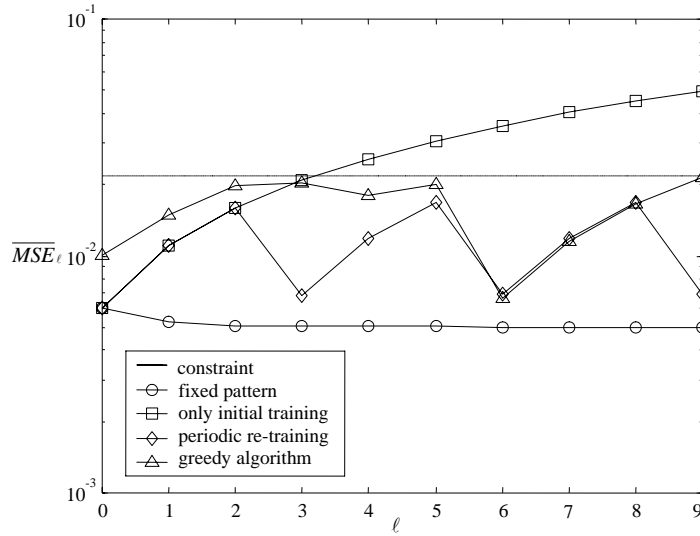
Figure 10.3: Normalized channel estimation error $\overline{MSE}_\ell$ as a function of the OFDM symbol ℓ for different pilot arrangements. Dashed line represents the constraint (10.9).
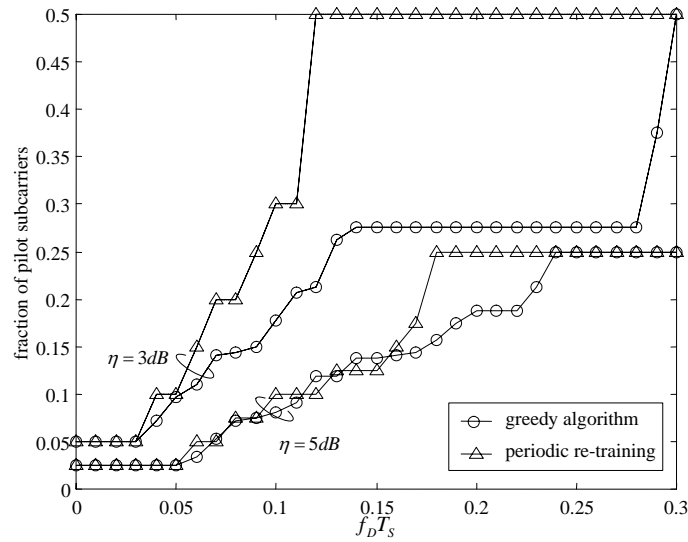

Figure 10.4: Fraction of total subcarriers ($N_B \times L = 10 \times 64$) used as pilot subcarriers for the periodic re-training and greedy algorithm versus the Doppler spread $f_D T_S$ ($SNR = 20dB$).
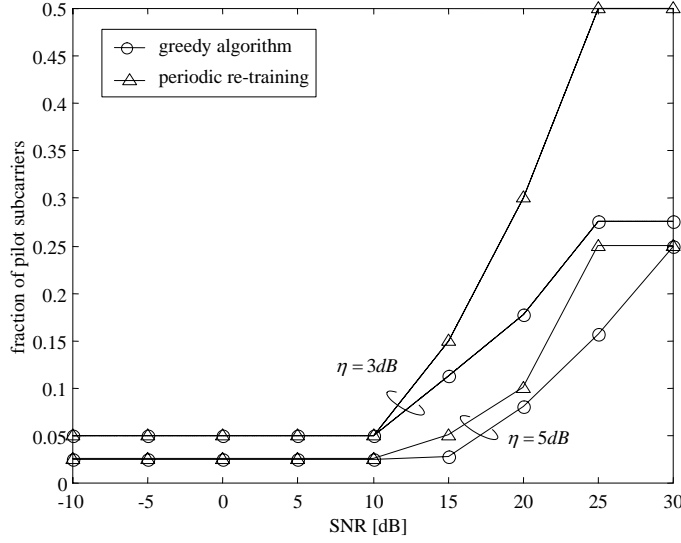
Figure 10.5: Fraction of total subcarriers ($N_B \times L = 10 \times 64$) used as pilot subcarriers for the period re-training and greedy algorithm versus SNR ($f_D T_S = 0.1$).

that coincides with (10.14b), we have from (10.17)

$$
\begin{aligned}
\mathbf{R}_{e,\ell} &= \mathbf{A}_\ell - \rho\boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}\mathbf{G}_\ell\tilde{\mathbf{C}}_\ell\breve{\mathbf{R}}_{g,\ell}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Theta}^H = \\
&= \mathbf{A}_\ell - \rho\boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}\mathbf{G}_\ell\tilde{\mathbf{X}}_\ell\mathbf{A}_\ell,
\end{aligned}
\tag{10.21}
$$

for the second equality we have used the definition $\tilde{\mathbf{C}}_\ell = \tilde{\mathbf{X}}_\ell\boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}$. Substituting (10.19) in (10.21) we finally get (10.14a).

10.8 Appendix-B: error correlation matrix as a function of the power allocation

From (10.17), (10.18) and (10.19), the error correlation matrix can be written as (we drop the argument ℓ for brevity)

$$
\mathbf{R}_g = \breve{\mathbf{R}}_g - \rho^2\breve{\mathbf{R}}_g\tilde{\mathbf{C}}^H(\tilde{\mathbf{C}}\breve{\mathbf{R}}_g\tilde{\mathbf{C}}^H + \sigma_n^2\mathbf{I})^{-1}\tilde{\mathbf{C}}\breve{\mathbf{R}}_g.
\tag{10.22}
$$

Applying the matrix inversion lemma, we get

$$
\begin{aligned}
\mathbf{R}_g &= \breve{\mathbf{R}}_g - \frac{\rho^2}{\sigma_n^2}[\breve{\mathbf{R}}_g\tilde{\mathbf{C}}^H\tilde{\mathbf{C}}\breve{\mathbf{R}}_g - \\
&\quad - \breve{\mathbf{R}}_g\tilde{\mathbf{C}}^H\tilde{\mathbf{C}}(\sigma_n^2\breve{\mathbf{R}}_g^{-1} + \tilde{\mathbf{C}}^H\tilde{\mathbf{C}})^{-1}\tilde{\mathbf{C}}^H\tilde{\mathbf{C}}\breve{\mathbf{R}}_g)],
\end{aligned}
\tag{10.23}
$$

that does not depend directly on $\tilde{\mathbf{X}}$ but only on the power allocation matrix $\boldsymbol{\Omega}_x$, since $\tilde{\mathbf{C}}^H\tilde{\mathbf{C}} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\Theta}^H\boldsymbol{\Omega}_x\boldsymbol{\Theta}\boldsymbol{\Omega}^{1/2}$. Since $\mathbf{R}_{g,\ell}$ and $\mathbf{R}_{e,\ell}$ are related through (10.13), the same applies to error correlation matrix $\mathbf{R}_{e,\ell}$.

Bibliography

[1] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986-1992, Nov. 1997.

[2] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: a convenient framework for time-frequency processing in wireless communications," *IEEE Proc. of the IEEE*, vol. 88, pp. 611-640, May 2000.

[3] A J. Goldsmith and S.-G. Chua, "Variable-rate, variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218-1230, Oct. 1997.

[4] S. Ye, R. S. Blum, L. J. Cimini Jr., "Adaptive modulation for variable-rate OFDM systems with imperfect channel state information," *Proc. VTC 2002,* pp. 768-771.

[5] ETSI TS 101 475, "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer," ETSI, Dec. 2001.

[6] IEEE Std 802.16c-2002 , "IEEE Standard for Local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems-Amendment 1: Detailed System Profiles for 10-66 GHz", 2002.

[7] I. Barhumi, G. Leus and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1615-1624, June 2003.

[8] A. Scaglione, G.B. Giannakis and S. Barbarossa, "Redundant filterbank precoders and equalizers. I. Unification and optimal designs," *IEEE Trans. on Signal Processing*, vol. 47, pp. 1988-2006, July 1999.

[9] B. Lu, X. Wang and K. R. Narayanan, "LDPC-based space-time coded OFDM systems over correlated fading channels: performance analysis and receiver design," *IEEE Trans. Comm.*, vol. 50, no.1, pp. 74-88, Jan. 2002.

[10] S. Haykin, *Adaptive filter theory*, Prentice Hall, fourth edition, 2002.

[11] C. Komninakis, C. Fragouli, A. H. Sayed and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing,* vol. 50, pp. 1065-1076, May 2002.

[12] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-963, April 2003.